CONTEXTUAL IMAGE CLASSIFICATION APPROACH FOR MONITORING OF AGRICULTURAL LAND COVER BY SUPPORT VECTOR MACHINES AND MARKOV RANDOM FIELDS

H. Vahidi [a, *], E. Monabbati [b]

[a] Department of Civil Engineering, Khavaran Institute of Higher Education, Mashhad, Iran - vahidi23800@itc.nl
[b] Department of Mathematics, Ferdowsi University of Mashhad, Mashhad, Iran - eh_mo236@stu-mail.um.ac.ir

KEY WORDS: Remote Sensing, Image Classification, Contextual, Support Vector Machines, Markov Random Fields

ABSTRACT:

The main idea of this paper is to integrate the non-contextual support vector machines (SVM) classifiers with Markov random fields (MRF) approach to develop a contextual framework for monitoring of agricultural land cover. To this end, the SVM and MRF approaches were integrated to exploit both spectral and spatial contextual information in the image for more accurate classification of remote sensing data from an agricultural region in Biddinghuizen, the Netherlands. Comparative analysis of this study clearly demonstrated that the proposed contextual method based on SVM-MRF models generates a higher average accuracy, overall accuracy and Kappa coefficient compared with non-contextual SVM method. Since the spatial information is considered in the proposed method, this study indicates that a neater, more homogonous and speckle-free results could be generated by the SVM-MRF approach.

1. INTRODUCTION

Many researches have been conducted to find a suitable method to classify the remote sensing data. The conventional pixel-based classification methods consider a pixel in isolation during an image processing approach, and do not utilize the contextual information within and between objects as an important source of knowledge for image classification.

The increasing dissatisfactions with classical pixel-based approach in the field of remote sensing have raised the new critiques on the necessity of changing of the conventional pixel based image processing method to more realistic modelling method for more accurate description of the real world in the late 1970s.

This fundamental paradigm in modelling of the existed relations or correlations within and between objects in real world can be observed in the early works of the researchers like Swain, et al. (1980; Swain, Stegel, & Smith, 1979), Gurney (1981), Swain, et al. (1981), Wharton (1982), Kittler and Föglein (1984), Di Zenzo, et al. (1987), Klein and James Press (1989) and Fung, et al. (1990). These researchers have discussed the shortcomings of the unsophisticated conventional image processing approaches as they neglect the valuable information that may be derived by considering the pixels by context.

Contextual image classification is an approach of classification based on contextual information which can be derived from the adjacent pixels that can be exploited in the images. In this framework, spatial contextual information can be expressed as the probability of presence of one object with respect to the presence of the neighbouring objects (Tso & Olsen, 2005).

In fact, the land cover types have been usually expanded on a large area compared with the pixel size of the adapted satellite remotely sensed image, so the associated land cover of any pixel in the image, i.e., the pixel label may dependent of the label of its neighbouring pixels. Furthermore, in practice, the satellite imagery sensors normally acquire particular portions of energy from neighboring pixels. To formulate these existing spatial relationships, the contextual image processing approach was proposed to consider the relation or correlation between neighboring pixels (Khedama & Belhadj-Aissaa, 2004; Solberg, Taxt, & Jain, 1996; Swain, et al., 1979). The contextual approach aims to improve the accuracy of results and efficiency of complex feature recognition and classification problems by consideration and modelling of the contextual knowledge on the incorporating regions in the real world (Ardila Lopez (2012); Blaschke, et al. (2010); Burnett & Blaschke (2003)).

In this paper the non-contextual support vector machines (SVM) classifiers was integrated with Markov random fields (MRF) model in a unique formulation to develop a spatial contextual classification approach for more accurate monitoring of agricultural land cover.

In the following sections, the generality of SVM and MRF methods will be discussed and the proposed model for the integration of these models will be introduced. Finally, the end of this paper is reserved for the evaluation of the results of proposed method and also the final comments and discussion.

2. METHODOLOGY

2.1 SVM Classification of Remotely Sensed Data

SVM is a robust supervised kernel-based non-contextual image classification method which derived from statistical learning theory. SVM classification approach can produce a sophisticated classification results in the complex and noisy images by minimizing the so-called structural risk, or classification errors, during solving of the classification problem (Chang & Lin, 2005; Tso & Olsen, 2005).

Unlike conventional statistical maximum likelihood methods, there is no assumption about class distribution in the SVM classification method. This characteristic makes the SVM

---

* Corresponding author.

approach to show a substantial improvement over traditional methods(Tso & Olsen, 2005).

In the following the concepts and fundamentals of this classifier will be reviewed for the binary classification problems.

Suppose that there are two linearly non-separable classes of ω_1, ω_2 in an image (i.e., $L = \{\omega_1, \omega_2\}$). The training data set is denoted by pairs (x_i, y_i), $i = 1,...,n$, $y_i = \{1, -1\}$, $x_i \in \mathbb{R}^d$, $y_i \in L$ where x_i are the observed multispectral feature vectors and y_i is the label of the information class for training sample i. The SVM classifier establishes an optimal hyperplane that discriminate two classes in such a way that the distance (margin) from the hyperplane to the nearest training data points in each of the classes is as large as possible (Mather & Tso, 2010). The hyperplane can be formulated as the following discriminant function:

$$w^T x + b = 0 \tag{1}$$

where x is a point lying on the hyperplane, w denotes a normal to the hyperplane and the parameter is the bias.

It is obvious that all the training data should satisfy the following constraint:

$$y_i(w^T x + b) \geq 1 - \xi_i ; \xi_i \geq 0, \forall i \tag{2}$$

Where ξ_i denotes slack variables, which are proportional to some measure of cost, are introduced to relax the constraints (Cortes & Vapnik, 1995; Tso & Olsen, 2005; Vapnik, 1998). The optimal problem then leads to:

$$\min\{\frac{\|w\|^2}{2} + C\sum_{i=1}^{n} \xi_i\} \tag{3}$$

where C is a penalty parameter which can control the misclassification of training samples.

In the case where a linear hyperplane is unable to separate the information classes sophisticatedly, the nonlinear SVMs then can be adapted. To this end, this approach maps the training data set into a higher dimensional space called Hilbert space ($\mathcal{H}$) by a nonlinear vector mapping function called $\Phi : R^d \to \mathcal{H}$ and performs manipulation within this mapped feature space to improve the discrimination between classes(Tso & Olsen, 2005; Vapnik, 1998).

It can be proven that the SVM-based decision rule shown in Equation (1) can be equivalently expressed as following equation in a higher dimensional space $\mathcal{H}$:

$$f(x) = \text{sign}[\sum_{i=1}^{nsv} \alpha_i y_i \Phi(x).\Phi(x_i) + b] \tag{4}$$

where nsv denotes the number of support vectors. For more details see (Mather & Tso, 2010).

The computational loading of ($\Phi(x).\Phi(x_i)$) can be quite costly in a higher dimensional space, so a positive definite kernel function which represented as K(x,y), where K(x,y)= $\Phi(x).\Phi(y)$ was proposed to reduce the computational burden (Tso & Olsen, 2005; Vapnik, 2000).

Therefore the decision rule expressed in Equation (4) can be generalized into:

$$f(x) = \text{sign}[\sum_{i=1}^{nsv} \alpha_i y_i k(x, x_i) + b] \tag{5}$$

In practice, the role of a kernel function can be expressed that as mapping the training data into a higher dimensional space in order to increase the quality of discrimination of the training samples (Moser & Serpico, 2013; Tso & Olsen, 2005). For more details on different types of kernel function and their characteristics and methods for developing SVMs for multiclass classifications see (Mather & Tso, 2010).

2.2 MRF Modelling in Remotely Sensed Data Classification

MRFs are stochastic models that formulate the spatial context in Bayesian image analysis approaches by means of the minimization of suitable energy functions (Dubes & Jain, 1989). This method has been found as a very efficient tool for characterizing contextual information. In following, the concept and the fundamentals of MRF will be briefly reviewed.

All image analysis problems could be formulated as the labelling problem, that is, to assign suitable labels from a label (class) set $L = \{\omega_1, \omega_2, ..., \omega_n\}$ to all elements of the set of sites (pixels) $S = \{1, 2, ..., m\}$. The concept of Markov random fields (MRF) is widely used, when the contextual information is also taken into account in the computer vision problems. In these problems the main principle is that the label of a site depends on the labels of its neighbourhood.

A neighbourhood system for S is a set of subsets that define neighbours of sites. Formally it is defined as $N = \{N_i : i \in S\}$, where N_i is the set of neighbours of site i. Let $F = \{F_1, F_2, ..., F_m\}$ be a family of random variables on S where each F_i takes a value $f_i \in L$, where $f = \{f_1, f_2, ..., f_m\}$ is a configuration of F corresponding to a realization of the field. For simplicity, the probability that random variable Fi takes the value fi is denoted by $P(F_i = f_i)$, abbreviated $P(f_i)$ and the joint probability $P(F = f) = P(F_1 = f_1, ..., F_m = f_m)$ is denoted by $P(f)$ from now on. Now F is said to be MRF on S with respect to a neighborhood system N if and only if the following two conditions are satisfied:

$$P(f) > 0, \ \forall f \in F \tag{6}$$

$$P(f_i \mid f_{s-\{i\}}) = P(f_i \mid f_{Ni}) \tag{7}$$

The second property presents a local characteristic of F as an extension of Markovianity in Markov processes.

Image analysis problem with contextual information could be reduced to find a suitable configuration of F (i.e. f) with maximum a posteriori (MAP) when an observed data is available. Joint probability could be computed using the equivalence between MRF and the so called Gibbs random field (GRF) (For more details see (Li, 2009)). GRF is a set of random variables with Gibbs distribution which takes the following form:

$$P(f) = \frac{1}{Z} e^{\frac{-U(f)}{T}} \tag{8}$$

$$Z = \sum_{f \in F} e^{\frac{-U(f)}{T}} \tag{9}$$

Where Z is a normalizing factor, T is the temperature and $U(f)$ is the energy function. The energy function is defined as:

$$U(f) = \sum_{c \in C} V_c(f) \tag{10}$$

where summation is over the set of all possible cliques, C. Note that a clique is a simple subset of sites. $V_c(f)$ is called clique potential and depends on the local configuration on the clique c.

Given labelling f_i and observed multispectral feature vector x_i ($x_i \in R^d$), the posterior probability is:

$$P(f_i \mid x_i) = \frac{P(x_i \mid f_i)P(f_i)}{P(x_i)} \tag{11}$$

where $P(f_i)$ is the prior probability, $P(x_i \mid f_i)$ is the conditional p.d.f of the observation x_i (i.e. likelihood function) and $P(x_i)$ is a constant, independent from labelling f_i called the density of x_i. Thus:

$$P(f_i \mid x_i) \propto P(x_i \mid f_i)P(f_i) \tag{12}$$

Assume that the label field f is a MRF, then $P(x_i \mid f_i)$ is also a MRF (For more details see (Geman & Geman, 1984)). The equivalence of MRF and GRF implies that:

$$U(f_i \mid x_i) = U(x_i \mid f_i) + U(f_i) \tag{13}$$

where $U(f_i \mid x_i)$ is the posterior energy for a pixel, $U(x_i \mid f_i)$ denotes conditional energy and $U(f_i)$ (also can be

denoted as $U(f_i \mid f_{Ni})$) is the prior energy function for neighbourhood system.
The posterior energy for entire image is defined as:

$$U(f \mid x) = \sum_{i=1}^{n} U(f_i \mid x_i) \tag{14}$$

2.3 SVM-MRF-based Proposed Method for Classification of Remotely Sensed Data

This research adopted Radial Basis Function (RBF) as a kernel function for SVM method as it has been shown its ability in the classification of remotely sensed images.
To integrate contextual information based on MRF model to non-contextual SVM, the posterior energy is derived based on Equation (13).
According to Equation (4), SVM method assigns class labels to pixels based on decision function $f(x)$. To apply SVM model as conditional probability function (likelihood function) $P(x_i \mid f_i)$ for deriving the respective conditional energy, it is required to produce class probabilities instead of class labels. This is done by implementation of Plott's posteriori probabilities theory for SVM. For more details see (Lin, Lin, & Weng, 2007).
To control the contribution of prior and conditional energy function in Equation (13), an additional parameter λ is defined:

$$U(f_i \mid x_i) = (1-\lambda)U(x_i \mid f_i) + \lambda U(f_i) \tag{15}$$

Where $0 \leq \lambda \leq 1$. In the case that λ=0, the prior model (contextual information) is completely ignored. Contrary to this case, if $λ=1$, only the prior model is considered and non-contextual information is ignored. Since this study is concentrated on the integration of contextual characteristic to non-contextual SVM classifier based on MRF model, the value of λ should be defined in the range of $0 < \lambda < 1$.
According to Equation (8), to maximize $P(f)$ value, the energy function has to be minimized. Thus the problem of finding of optimal labelling is turn into energy function minimization problem.
The problem of finding optimal labelling is difficult if the energy function is not convex. Simulated annealing (S.A) is one of the stochastic-based global optimization methods which is widely used in the image analysis context. According to the abilities of this optimization technique in context of optimization in the field of image analysis, this method was adopted in this research. S.A tries to minimize the energy function iteratively. In the k th iteration of S.A a new solution, s^{k+1}, is generated from the previous iteration, s^k. Let $\Delta E = E(s^{k+1}) - E(s^k)$ be the difference between the energy of solutions of these two consecutive iterations. If $\Delta E \leq 0$ then s^{k+1} is accepted as the new solution, since it improves the objective function, otherwise, s^{k+1} is accepted with probability $\exp(-\Delta E / T^k)$, where T^k is the temperature at iteration k. Next, the current temperature is updated (decreased) and the process continues until some stopping criterion is hold, e.g. after 3 successive "temperature stages" with no acceptance. A "temperature stage" occurs when $12n$

perturbations are accepted or $100n$ perturbations are attempted, where n is the number of variables.

The temperature updating rule plays critical role in the convergence of the method to global optimum. A traditional choice is to use geometric formulation, that is, $T^{k+1} = \alpha T^k$, where α is a constant. Another updating rule, known as "adaptive" annealing, may provide better solutions, so it was adopted in this research. It is defined as $T^{k+1} = D^k T^k$ where:

$$D_k = \min\left\{ D_0, \frac{E_k}{\langle E_k \rangle} \right\} \qquad (16)$$

with D_0 =0.5 to 0.9, E_k : the minimal accepted energy during the iteration k and $\langle E_k \rangle$: the average energy accepted during the iteration k. Note that at high temperature since $E_k / \langle E_k \rangle$ is small then temperature is lowered quickly. For more details see (Li, 2009) and (Pétrowski & Taillard, 2006).

The proposed model was implemented in R programming language.

2.4 Study Area

The area of interest for this study is located in Biddinghuizen village in the province of Flevoland of the Netherlands (52°27′N 5°42′E). This area represents a typical modern agricultural region in the Netherlands. The agricultural fields in this study area are large and usually rectangular. The main crops of the study area are grass, potato, wheat, sugar-beet, bean, pea and onion. The elevation differences in the Biddinghuizen region are very small (Abkar, 1999).

2.5 Data

The image data that was used in this study is a Landsat Thematic Mapper (TM) image that was acquired on 7 July 1987 (Figure 1).

This data was selected as it was used extensively in the previous studies in the context of agricultural land cover classification by the different researchers like Janssen and Middelkoop (1992), Janssen and Molenaar (1995), Fatemi and Abkar (2003), Lesparre and Gorte (2006) and Vahidi (2010). Therefore, this characteristic could provide a good chance for further evaluation and comparison of the different proposed approaches for the agricultural land cover classification.
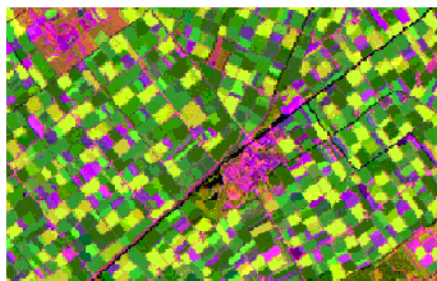

Figure 1. Landsat Thematic Mapper (TM) image (bands 3, 4, 5), Biddinghuizen, the Netherlands

The image was georeferenced to the Dutch national triangulation system using a first-degree affine transformation.

The pixels were resampled to the original size of 30 m by 30 m. The spectral bands 3, 4, and 5 of TM image were used in this research. Also the GIS-ready reference land cover map of the study area for 1987 crop year was used in the stages of training sample collection and validation of the results of this research (Figure 2).
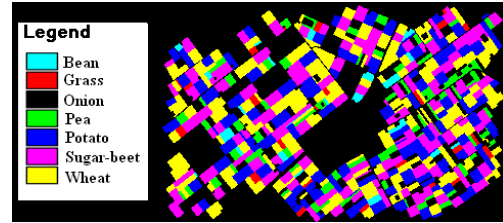

Figure 2. Land cover map of study area

3. RESULTS AND DISCUSSION

Figure 3 shows the classified image based on SVM method where C=10 and σ=1. Accuracy assessment clearly shows the potential of SVM method for classification of the major classes in the image (Table1).
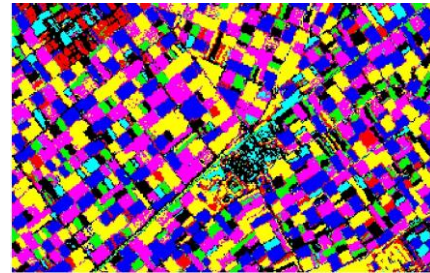

Figure 3. Classified image generated by SVM, where C=10 and σ=1

| | | Reference | | | | | | | |
		Grass	Wheat	Potato	Sugar-beet	Bean	Pea	Onion	Total
Classified	Grass	303	18	12	3	7	17	0	360
	Wheat	1	1561	4	43	0	4	0	1613
	Potato	8	0	1081	0	21	1	0	1111
	Sugar-beet	1	39	0	954	13	3	2	1012
	Bean	4	3	0	4	298	8	8	325
	Pea	9	0	0	0	6	564	0	579
	Onion	0	5	0	12	25	22	494	558
	Total	326	1626	1097	1016	370	619	504	5558
	Overall Accuracy = 94.5%								
	Kappa Coefficient = 93.3%								

Table 1. Accuracy assessment of SVM method

In the proposed model, the SVM-based classified image was adopted as a suitable configuration of F by the MRF model.

According to the contextual origin of the proposed SVM-MRF-based approach, this method is very effective against noisy small and limited regions. These regions are labelled based on the likelihood function and prior probability according to their neighbourhood existed state. However, beside the advantages of this approach in handling of this type of classification errors, there is a drawback to this method. In some realistic cases in the real world, it can be observed that a small dissimilar region (class) has been surrounded by a homogeneous area (class). So,

this configuration of the information classes may increase the chance of misclassification of the small regions, as the contextual filter may consider these correctly classified pixels as the noisy pixels, depending on the size of the region and the temperature state. Therefore, the classified image configuration with the minimized total posterior energy is not necessarily has the best classification result. To find the optimized solution among the S.A algorithm results which were generated at the consecutive iterations of this optimization approach, the averaged accuracy (A.A), overall accuracy (O.A) and kappa coefficient (k) criteria were also be considered in this study beside the total energy of the generated configurations.

Figure 4 demonstrates the temperature state and total energy state of an image according to consecutive iterations of S.A method. The implemented model was stopped by the controlling criterion of the adaptive S.A algorithm after 37 consecutive iterations. However, although the optimized configuration of classified image from the total energy content aspect was generated at iteration 37, but according to Figure 5 it can be observed that the best O.A, AA, and k of the classification practice occurred in the iteration 26. Therefore, the classification product of this iteration of S.A algorithm was adopted as the result of SVM-MRF classification method.

The classified image based on SVM-MRF method is illustrated in Figure 6, where C=10, σ=1, λ=0.95 and T=0.02. Table 2 shows the results of accuracy assessment for the classified image based on SVM-MRF method.
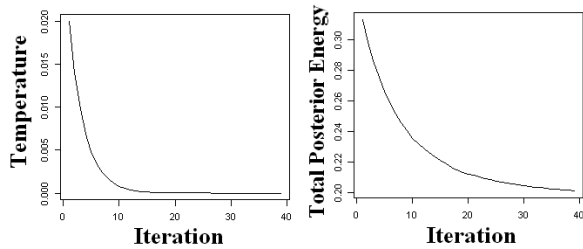

Figure 4. Temperature and total energy state of an image according to iterations of S.A method
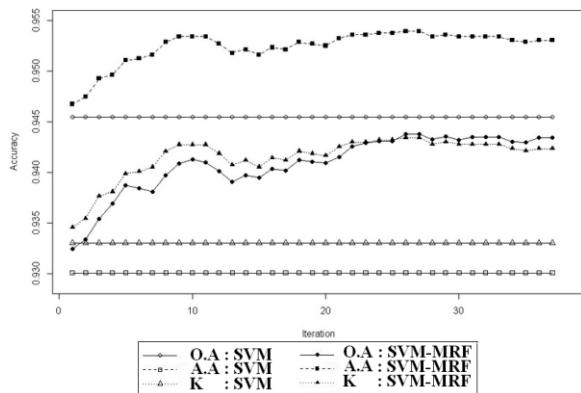

Figure 5. Overall accuracy (O.A), average accuracy (A.A) and kappa coefficient (K) for SVM method and SVM-MRF method according to iterations of S.A method
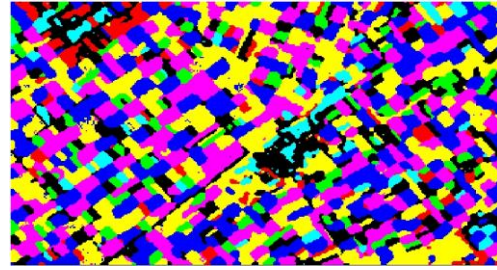

Figure 6. Classified image generated by SVM-MRF, where C=10, σ=1, λ=0.95 and T=0.02

		Reference						
		Grass	Wheat	Potato	Sugar-beet	Bean	Pea	Onion	Total
Classified	Grass	316	16	7	0	4	14	0	357
	Wheat	2	1541	7	18	0	4	0	1572
	Potato	5	12	1082	0	20	0	0	1119
	Sugar-beet	0	52	0	997	17	6	0	1072
	Bean	0	0	0	0	305	11	4	320
	Pea	0	2	0	0	2	563	2	569
	Onion	3	3	1	1	22	21	498	549
	Total	326	1626	1097	1016	370	619	504	5558
Overall Accuracy = 95.3%									
Kappa Coefficient = 94.3%									

Table 2. Accuracy assessment of SVM-MRF method

The proposed contextual method provides 1.3% increase in average accuracy, 0.8% in overall accuracy and 1% in Kappa coefficient compared to a non-contextual SVM (Table 3). However, as Table 3 was illustrated, this approach was considerably increased the producer's accuracy in the case of grass and sugar-beet classification while it was slightly decreased the producer's classification accuracy of wheat and onion crops as a result of the misclassification of the correctly classified small regions.

	Producer's Accuracy	
	SVM	SVM-MRF
Grass	92.9%	96.9%
Wheat	96.0%	94.7%
Potato	98.5%	98.6%
Sugar-beet	93.8%	98.1%
Bean	80.5%	82.4%
Pea	91.1%	90.9%
Onion	98.0%	98.8%
Average Accuracy	93.0%	94.3%
Overall Accuracy	94.5%	95.3%
Kappa Coefficient	93.3%	94.3%

Table 3. Accuracy comparison for SVM and SVM-MRF methods

A visual comparison of the classified images (Figure 3 and Figure 6) shows that the SVM-MRF model was produced a neater and speckle-free classification results and improved spatial regularity in the generated map compared to the SVM approach.Conclusions
Results of this study showed the distinctive abilities and advantages of contextual methods based on the integration of SVM and MRF methods over non-contextual SVM method.

Comparative analysis clearly demonstrated that substantially higher average accuracy, overall accuracy and Kappa coefficient could be gained by incorporating the spatial contextual information by SVM-MRF-based proposed method compared with SVM method. Since the spatial information is considered in the proposed method, the study indicates a neater, more homogonous and speckle-free results could be generated by the SVM-MRF approach.

4. REFERENCES

Abkar, A. 1999. Likelihood-based segmentation and classification of remotely sensed images. A Bayesian optimization approach for combining rs and gis. ITC, Enschede.

Ardila Lopez, J. P. 2012. Object - based methods for mapping and monitoring of urban trees with multitemporal image analysis. University of Twente, Faculty of Geo-Information and Earth Observation (ITC), Enschede.

Blaschke, T., Johansen, K., & Tiede, D. 2010. Object-Based Image Analysis for Vegetation Mapping and Monitoring. Advances in Environmental Remote Sensing: Sensors, Algorithms, and Applications, 7, 241.

Burnett, C., & Blaschke, T. 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. [doi: 10.1016/S0304-3800(03)00139-X]. Ecological Modelling, 168(3), 233-249.

Chang, C., & Lin, C. 2005. LIBSVM: a library for support vector machines (2001), Software.

Cortes, C., & Vapnik, V. 1995. Support-vector networks. Machine learning, 20(3), 273-297.

Di Zenzo, S., Bernstein, R., Degloria, S., & Kolsky, H. 1987. Gaussian maximum likelihood and contextual classification algorithms for multicrop classification. Geoscience and Remote Sensing, IEEE Transactions on(6), 805-814.

Dubes, R. C., & Jain, A. K. 1989. Random field models in image analysis. Journal of Applied Statistics, 16, 131–163.

Fatemi, S. B., & Abkar, A. A. 2003. A Top-Down Approach for GIS and Remote Sensing Integration. Paper presented at the The Second Conference On Machine Vision, Image Processing & Applications (MVIP 2003), Tehran, Iran.

Fung, P., Grebbin, G., & Attikiouzel, Y. 1990. Contextual classification and segmentation of textured images. Paper presented at the Acoustics, Speech, and Signal Processing International Conference, ICASSP-90., 1990

Geman, S., & Geman, D. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. Pattern Analysis and Machine Intelligence, IEEE Transactions on(6), 721-741.

Gurney, C. M. 1981. The use of contextual information to improve land cover classification of digital remotely sensed data. [doi: 10.1080/01431168108948372]. International Journal of Remote Sensing, 2(4), 379-388.

Janssen, L., & Middelkoop, H. 1992. Knowledge-based crop classification of a Landsat Thematic Mapper image. International Journal of Remote Sensing, 13(15), 2827-2837.

Janssen, L. L., & Molenaar, M. 1995. Terrain objects, their dynamics and their monitoring by the integration of GIS and remote sensing. Geoscience and Remote Sensing, IEEE Transactions on, 33(3), 749-758.

Khedama, R., & Belhadj-Aissaa, A. 2004. Contextual classification of remotely sensed data using MAP approach and MRF. ISPRS Journal of Photogrammetry and Remote Sensing, 35(7), 11-16.

Kittler, J., & Föglein, J. 1984. Contextual classification of multispectral pixel data. Image and Vision Computing, 2(1), 13-29.

Klein, R., & James Press, S. 1989. Contextual Bayesian classification of remotely sensed data. Communications in statistics-theory and methods, 18(9), 3177-3202.

Lesparre, J., & Gorte, B. 2006. Using mixed pixels for the training of a maximum likelihood classification. Paper presented at the Proceedings of ISPRS Commission VII Mid-term Symposium, Enschede, the Netherlands.

Li, S. Z. 2009. Markov random field modeling in image analysis: Springer.

Lin, H.-T., Lin, C.-J., & Weng, R. C. 2007. A note on Platt's probabilistic outputs for support vector machines. Machine learning, 68(3), 267-276.

Mather, P., & Tso, B. 2010. Classification methods for remotely sensed data: CRC press.

Moser, G., & Serpico, S. B. 2013. Combining Support Vector Machines and Markov Random Fields in an Integrated Framework for Contextual Image Classification. Geoscience and Remote Sensing, IEEE Transactions on, 51(5), 2734-2752.

Pétrowski, A., & Taillard, E. 2006. Metaheuristics for hard optimization: Springer.

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

Swain, P. H., Siegel, H. J., & Smith, B. W. 1980. Contextual classification of multispectral remote sensing data using a multiprocessor system. Geoscience and Remote Sensing, IEEE Transactions on(2), 197-203.

Swain, P. H., Stegel, H. J., & Smith, B. W. 1979. A method for classifying multispectral remote sensing data using context. Paper presented at the LARS Symposia.

Swain, P. H., Vardeman, S. B., & Tilton, J. C. 1981. Contextual classification of multispectral image data. Pattern Recognition, 13(6), 429-441.

Tso, B., & Olsen, R. C. 2005. A contextual classification scheme based on MRF model with improved parameter estimation and multiscale fuzzy line process. Remote Sensing of Environment, 97(1), 127-136.

Vahidi, H. 2010. Evaluation of the Pixel Based and Object Based Classification Methods For Monitoring Of Agricultural Land Cover Case study: Biddinghuizen - the Netherlands. Paper presented at the Geomatics 89, Tehran, Iran.

Vapnik, V. 2000 The nature of statistical learning theory: springer.

Vapnik, V. N. 1998. Statistical learning theory.

Wharton, S. W. 1982. A contextual classification method for recognizing land use patterns in high resolution remotely sensed data. Pattern Recognition, 15(4), 317-324.