# A Context-Sensitive Technique Based on Support Vector Machines for Image Classification

Francesca Bovolo and Lorenzo Bruzzone

Department of Information and Communication Technologies,
University of Trento,
Via Sommarive 14, 38050 Povo (TN), Italy
lorenzo.bruzzone@ing.unitn.it

**Abstract.** In this paper, a novel context-sensitive classification technique based on Support Vector Machines (CS-SVM) is proposed. This technique aims at exploiting the promising SVM method for classification of 2-D (or n-D) scenes by considering the spatial-context information of the pixel to be analyzed. The context-based architecture is defined by properly integrating SVMs with a Markov Random Field (MRF) approach. In the design of the resulting system, two main issues have been addressed: i) estimation of the observation term statistic (class-conditional densities) with a proper multiclass SVM architecture; ii) integration of the SVM approach in the framework of MRFs for modeling the prior model of images. Thanks to the effectiveness of the SVM machine learning strategy and to the capability of MRFs to properly model the spatial-contextual information of the scene, the resulting context-sensitive image classification procedure generates regularized classification maps characterized by a high accuracy. Experimental results obtained on Synthetic Aperture Radar (SAR) remote sensing images confirm the effectiveness of the proposed approach.

## 1 Introduction

Image classification plays a very important role in many application domains, like biomedical and remote sensing image analysis, industrial visual inspection, video surveillance, *etc*. Although often in real applications and commercial software packages image classification problems are addressed according to pixel-based (context-insensitive) classifiers, from a theoretical and practical point of view it is very important to develop classification techniques capable to exploit the spatial-context information present in the images. In this framework, it seems particularly relevant to develop context-sensitive classification methods capable to properly exploit the most promising pixel-based classification methodologies recently proposed in the literature. In this context, a promising machine learning approach to classification is that based on Support Vector Machines (SVMs). SVMs, originated from the statistical learning theory formulated by Vapnik and co-workers [1], are a distribution-free classification approach, which proven very effective in many context-insensitive classification and regression problems. SVM-based classifiers have three main advantages with respect to standard machine learning techniques based on neural networks: i) simple architecture design; ii) moderate computational complexity; iii) excellent generalization cability [1]. In particular, advantage iii) is very

S.K. Pal et al. (Eds.): PReMI 2005, LNCS 3776, pp. 260 – 265, 2005.

relevant with respect to image classification problems. In greater detail, designing a classifier characterized by good generalization properties requires assuming: i) a high number of training sample is available; ii) that the training samples are statistical independendent. In image classification problems this last assumption is frequently violated due to high dependency between neighboring pixels; thus the excellent generalization ability of SVMs and their robustness to the Hughes phenomenon [1] seem very suitable to the solution of image classification problems. However, at the present SVMs have been proposed as a context-insensitive classification procedure. Nonetheless, their effectiveness makes it attractive to extend them to address context-sensitive image classification problems.

One of the most promising approaches to context-sensitive classification is to exploit the spatial-context information in the decision rule [2]. In this framework, the correlation among labels can be modeled according to the analysis of either the prior model (regularization term) of the scene and/or the conditional density component (observation term) present in the acquired signal. In this context, in the literature there has been a wide emphasis on using statistical techniques based on Markov Random Fields (MRFs) [2].

In this paper, we propose to properly integrate the SVM technique within a MRFs framework, in order to obtain accurate and reliable classification maps taking into account context-spatial dependence between neighboring pixels. SVMs are used in the MRF for modeling the observation term, which results in a complex problem because the structural risk minimization principle driving SVM is not directly related to a Bayesian estimation of the conditional density functions of classes. This requires additional steps for a proper and effective modeling of the observation component.

Experimental results and comparisons, carried out on Synthetic Aperture Radar (SAR) remote sensing images, confirm the effectiveness of the proposed classification architecture.

## 2 The Proposed Context-Sensitive Technique Based on SVM

Let us consider a supervised image classification problem. Let $X$ be the image (of size $I \times J$ pixels) to be classified. The training set is composed of $L$ vectors $\mathbf{x}_i$ ($i = 1,2,\ldots,L$) from a $n$-dimensional feature space. Feature space definition is a data dependent task strictly related to user requirements and to the considered image. Here, we assume to have a feature set that properly describes the considered information classes. Each vector in the available training set is associated with a target which identifies the membership of the pattern $\mathbf{x}_i$ to one of the $K$ land-cover classes $\omega_j \in \mathbf{\Omega}$, $\mathbf{\Omega} = \{\omega_1, \omega_2, \ldots, \omega_K\}$, to be discriminated in the considered classification problem.

The proposed technique integrates the SVM approach in the MRF framework according to the use of a proper classification architecture (see Fig. 1). The proposed architecture is composed of three main blocks: i) a multiclass SVM architecture based on the one-against-all (OAA) strategy; ii) a block devoted to the estimation of the conditional densities of classes; and iii) an MRF context-based classifier. As both the multiclass SVM classifier and the block for the estimation of class conditional density functions require a training phase for parameters estimation, the available training set is divided in two subsets $T_1$ and $T_2$, built up of $L_1$ and $L_2$ pattern (with $L_1$+$L_2$=$L$), respectively.

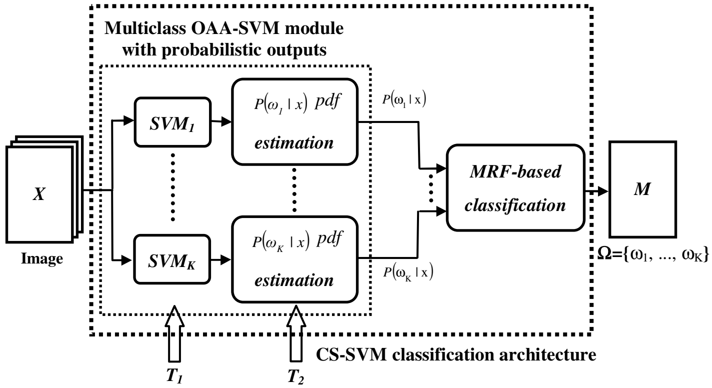


**Fig. 1.** Block scheme of the proposed CS-SVM based technique for image classification

## 2.1 Context-Insensitive Multiclass SVM Module

As can be seen in Fig. 1, the first block of the proposed architecture is based on a multiclass SVM classifier that defines a context-insensitive classification model by exploiting the concept of geometrical margin maximization rather than a *bayesian* statistical decision rule. SVMs have been originally designed to solve binary problems. To address multiclass problems, in the literature, among the others, two principal strategies have been proposed: the one-against-all (OAA) and the one-against-one (OAO). Here we focus our attention on the first one as it simplifies both the architecture and the phase of class conditional density estimation. The OAA-SVM architecture involves $K$ binary SVMs, one for each class. Each SVM is trained to discriminate between one information class $\omega_j$, $\omega_j \in \mathbf{\Omega}$, and all the others $\mathbf{\Omega} - \omega_j$. In the general case of nonlinear separable classes, the discriminant function defined by the *j-th* binary SVM can be expressed as:

$$f_j(\boldsymbol{x}) = \boldsymbol{w}_j \cdot \Phi_j(\boldsymbol{x}_i) + b_j \tag{1}$$

where $\Phi_j(\cdot)$ is a proper nonlinear kernel (which satisfies the Mercer's theorem) that maps the input feature space into a higher dimensional space where classes can be linearly separated by means of an optimal hyperplane defined by a normal vector $\boldsymbol{w}_j$ and a bias $b_j$. Once the kernel $\Phi_j(\cdot)$ is defined, training the SVM means estimating the $\Phi_j(\cdot)$ parameters and the values of $\boldsymbol{w}_j$ and $b_j$ so that the following cost function is minimized [1]:

$$\Psi_j(\boldsymbol{w}_j, \xi_j) = \frac{1}{2}\|\boldsymbol{w}_j\|^2 + C_j \sum_{i=1}^{L_1} \xi_{ji} \tag{2}$$

where $\xi_{ji}$ are the so-called *slack-variable* introduced to account for the non-separability of data, and the constant $C_j$ represents a regularization parameter that allows to control the penalty assigned to errors. For further details on the minimization of (2) and on SVM theory the reader is referred to [1]. The training of the $K$

considered SVMs (which is carried out on the training set $T_1$) results in the definition of the set of discrimination functions $f_j$ (with $j = 1, ..., K$). For each pattern the sign of $f_j(\boldsymbol{x}_i)$ indicates whether the given pattern $\boldsymbol{x}_i$ belongs to the class $\omega_j$ or not, whereas its absolute value quantifies how far the pattern is from the *j-th* discriminating function. Since the output of each SVM defines the distance between the pattern and the hyperplane, to derive an estimation of the conditional class probabilities $p(x \mid \omega_j)$, $\omega_j \in \boldsymbol{\Omega}$, to be used in the MRF, it is necessary to exploit a proper estimation procedure. A way to compute the *j-th* posterior probability from *j-th* SVM's outputs is to model it as a sigmoid function of the distance from the *j-th* hyperplane as:

$$P(\omega_j \mid \boldsymbol{x}_i) = \frac{1}{1 + \exp[S_j f_j(\boldsymbol{x}_i) + R_j]} \tag{3}$$

where $S_j$ and $R_j$ are the sigmoid parameters computed on $T_2$ by minimizing the cross-entropy error function [3]. It is worth noting that, for each class, the observation term (proportional to the conditional class probabilities) is obtained dividing the posterior probabilities by the prior probabilities (computed as relative frequency of the patterns belonging to the considered class in $T_1$).

## 2.2 Context-Sensitive MRF-Based Module

This module integrates the SVM based estimates within the MRF framework according to the Bayesian decision theory. The MRF-based classification strategy exploits the assumption that a pixel belonging to a given class $\omega_j \in \boldsymbol{\Omega}$ is likely to be surrounded by pixel belonging to the same class. Therefore exploiting this dependence may yield to more reliable and accurate classification results.

Let $\mathbf{C} = \{\mathbf{C}_l, 1 \le l \le Z\}$ with $Z$=K$^{IJ}$ be composed of all the sets of labels in the image $\boldsymbol{X}$, where $\mathbf{C}_l = \{C_l(m,n),\ 1 \le m \le I,\ 1 \le n \le J\}$ (with $C_l(m,n) \in \boldsymbol{\Omega}$) is a generic set of label in $\boldsymbol{X}$. The formulation of the Bayes decision rule for minimum error in terms of Markovian approach to classification is equivalent to minimizing the following energy function [2],[4]:

$$U(\boldsymbol{X} \mid C_l) = \sum_{n=1}^{I} \sum_{m=1}^{J} \left[ U_{ci}(\boldsymbol{X}(n,m) \mid C_l(n,m)) + U_{cs}(C_l(n,m) \mid \{C_l(g,h), (g,h) \in N(n,m)\}) \right] \tag{4}$$

where $U_{ci}(.)$ is the context-insensitive energy term that represents the statistics of the considered image (observation term), and $U_{cs}(.)$ is the context-sensitive energy term computed over the local neighborhood system $N(n,m)$ of the generic pixel in spatial position $(n,m)$. The shape and the size of $N(.)$ depends on the considered image and application [2]. As we are modeling the random variable $\boldsymbol{X}$ as an MRF over the neighborhood system, the context-sensitive term should be a Gibbs energy function and shows an exponential dependency from the prior model for the class labels $P(C_l)$ [2]. In practice $U_{cs}(.)$ is defined as:

$$U_{cs}\left(C_l(n,m) \mid \{C_l(g,h),(g,h) \in N(n,m)\}\right) = \sum_{(g,h)\in N(n,m)} \beta\,\delta\left(C_l(n,m) \mid C_l(g,h)\right) \tag{5}$$

where $\delta$ is the Kronecker delta function [2] and $\beta$ is a constant that controls the influence of the spatial-contextual information in the classification process. Under the assumption of class conditional independence [2], $U_{ci}(.)$ can be expressed as the product of the conditional class probabilities $p(x \mid \omega_j)$. In this paper this term has been estimated by means of the probabilistic OAA-SVM classifier system introduced in the previous paragraph.

Concerning the optimization procedure for minimizing (4), in the proposed approach we analyzed the effectiveness of the three main strategies proposed in the literature: i) iterated conditional modes (ICM) [4]; ii) simulated annealing (SA) [2]; and iii) maximizer of posterior marginals (MPM) [5].

## 3 Experimental Results

Several experiments on a remote sensing application case were carried out in order to assess the effectiveness of the proposed CS-SVM classifier with respect to the context-insensitive SVM (CI-SVM) classifier. The considered data consists of SAR images acquired by the ERS-1 satellite over the city of Berna (Switzerland). The considered classification problem is characterized by four land-cover classes: forest, urban area, water and fields. The analysis was carried out on a data set made up of 8 temporally filtered images, acquired in the period between June 1995 and May 1996, plus a feature based on the average long-term coherence (the reader is referred to [6] for greater detail on the data set). In all experiments, SVMs with Gaussian radial basis kernel functions were used and context-based classification was carried out using a second order neighborhood system. Several trials were carried out to find the best values of all the binary SVMs classifiers and sigmoidal functions parameters; also the optimal context weight $\beta$ was derived for each of the three MRF optimization procedures.

Table 1 summarizes the best accuracies obtained with the different techniques on the test set. As expected, the use of the spatial-context information allows to properly model the autocorrelation function between pixels regularizing the classification map and increasing the overall accuracy. In greater detail, the proposed CS-SVM approach (with all the optimization) significantly increased the overall accuracy and the kappa

**Table 1.** Overall and Kappa accuracies obtained using a standard context-insensitive (CI-SVM) and the proposed context-sensitive technique based on SVM (CS-SVM) with different optimization algorithms

| **Classification strategy** | | **Kappa coefficient** | **Overall Accuracy (%)** | **Forest** | **Urban area** | **Water** | **Fields** |
|---|---|---|---|---|---|---|---|
| Standard CI-SVM | | 0.88 | 91.97 | 96.11 | 91.20 | 81.14 | 90.45 |
| Proposed CS-SVM | ICM | 0.91 | 94.38 | 98.54 | 93.48 | 83.56 | 92.90 |
| | MPM | 0.93 | 95.52 | 99.43 | 94.13 | 83.49 | 94.47 |
| | SA | 0.94 | 95.92 | 99.43 | 94.60 | 84.00 | 95.03 |

accuracy with respect to the CI-SVM technique. In particular, the highest accuracy was obtained with the simulated annealing optimization procedure, with a significant increase, i.e. 6%, with respect to the classical pixel-based classification strategy. It is worth noting that also class-by-class accuracies increased significantly when introducing the context information in the classification process. Finally, from an analysis of the obtained classification maps, it is possible to observe that the proposed approach properly regularizes the resulting classification maps.

## 4 Conclusions

In this paper, a novel context-sensitive image classification approach based on SVMs has been proposed. This approach exploits a properly defined multiclass SVM architecture for estimating the distribution of the observation term to be jointly optimized with the prior term in the global energy function used for modeling the classification problem under the MRF framework.

Experimental results carried out on remote sensing SAR images confirmed the effectiveness of the proposed context-sensitive approach, which significantly outperformed the standard CI-SVM classifier. This mainly depends on the capability of the proposed method to both properly take into account the contextual information (included in the model of the *a-priori* term) in the decision process and model the observation term with an effective and reliable machine learning methodology.

As future development of this work we plan to further extend the experimental investigation of the proposed image classification approach to other application domains. In addition, we will investigate the possibility to include the spatial-context information directly in the cost function optimized in the learning phase of SVMs.

## References

1. Vapnik, V. N.,: Statistical Learning Theory, Wiley, 1998.
2. Geman, S., Geman, D.: Stochastic Relaxation, Gibbs Distribution, and Bayesian Restoration of Images. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-6 (1984) 768-778.
3. Platt, J.C.: Probabilistic Outputs for Support Vector Machines and Comparison to Regularized Likelihood Methods. In: Smola, A., Bartlett, P., Scholkopf, B., Schuurmans, D. (eds.): Advances in Large Margin Classifiers. Cambridge, MA, MIT Press (1999).
4. Besag, J.: On the Statistical Analysis of Dirty Pictures. *J. Roy. Statist. Soc. B*. 48 (1986) 259-302.
5. Marroquin,J., Mitter, S., Poggio, T.: Probabilistic Solution of Ill-posed Problems in Computational Vision. *J. Amer. Stat. Assoc.*. 82 (1987) 76-89.
6. Bruzzone, L., Marconcini, M., Wegmuller, U., Wiesmann, A.: An advanced system for the automatic classification of multitemporal SAR images. *IEEE Trans. Geosci. Rem. Sens.* 42 (2004) 1321-1334