FAST AND ACCURATE IMAGE CLASSIFICATION WITH HISTOGRAM BASED FEATURES AND ADDITIVE KERNEL SVM

Begüm Demir and Lorenzo Bruzzone

Dept. of Information Engineering and Computer Science, University of Trento,
Via Sommarive, 14, I-38123 Trento, Italy

ABSTRACT

Kernel-based image classification methods rely on the considered kernel functions that can be chosen with respect to prior information on the adopted features. In remote sensing, histogram features have recently gained an increasing interest due to their capability to address several critical classification problems (e.g., the problem of curse of dimensionality) when appropriate kernels and classifiers are selected. In view of that, in this paper we introduce in remote sensing additive kernels in the context of support vector machine classification (AK-SVM), which are suitable kernels for histogram based feature representations. In particular, we investigate the Histogram Intersection kernel and the chi-square kernel within the AK-SVM. Moreover, we present fast implementations of the AK-SVM to significantly speed up the classification phase of the SVM. Experimental results show the effectiveness of the AK-SVM in terms of classification accuracy and computational time when compared to SVMs with standard kernels.

Index Terms— Additive kernels, support vector machines, histogram intersection kernel, chi-square kernel, histogram based features, remote sensing.

1. INTRODUCTION

In the recent years kernel-based image classification techniques, in particular Support Vector Machines (SVMs), have become very popular in remote sensing (RS) due to their capability to effectively solve complex non-linear classification problems. Conventional kernel-based algorithms developed in RS mainly exploit general-purpose kernels, such as Gaussian Radial Basis Function (RBF) and polynomial kernels. This is due to the ability of these kernels to efficiently assess the similarities among low-level features in the kernel space.

In RS, a large number of low-level features (obtained by spectral, textural and geometrical/spatial feature-extraction techniques) are used in the framework of kernel classifiers by simply considering the general-purpose kernels. However, choosing a proper kernel is crucial for classification problems. A simple approach to select the kernel type can be defined on the basis of prior knowledge on the characteristics of the adopted features [1].

Recent developments in computer vision have shown that histogram features, which model marginal sample distributions to characterize the texture information, are very effective in many pattern recognition problems (such as image classification and visual appearance modeling) [2], [3]. This is due to their: i) simplicity; ii) high discrimination capability; and iii) sparse representation. Due to sparse representation of histograms, the generalization capability of kernel-based classifiers is high when: a) the features are high dimensional histograms (the dimension may vary from thousands to tens of thousands); and b) the proper kernels are selected [2]. Histogram features have recently gained an increasing interest in RS. A bag-of-visual-words representation of the local invariant features extracted by the scale invariant feature transform is used to characterize each image in [4]. Histogram features, which are modeled as marginal local distributions of several different filter responses at each pixel location, are found effective for RS image segmentation problems in [5]. Histogram based attribute profiles for classification of very high resolution RS images have been recently introduced in [6]. We expect that the use of histogram features will further attract attention in the RS community due to their above-mentioned effectiveness.

In this paper we focus our attention on histogram features and introduce additive kernels in the framework of SVMs (AK-SVMs) to classify RS images with histogram based features. Unlike SVMs with standard kernels, AK-SVMs are mainly suitable for histogram features. In particular, we present Histogram Intersection (HI) kernel and chi-square (χ^2) kernel within the AK-SVMs. Then, we introduce in RS fast implementations of AK-SVMs. Note that in recent years the AK-SVMs have been found very effective for image classification problems in the computer-vision community [3], whereas their use in RS for classification problems has not been explored yet. In this paper, we present and test the AK-SVMs and their fast implementations for RS image classification problems.

2. METHODOLOGY

2.1. Problem definition

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_P]$ be a RS image made up of P patterns, where $\mathbf{x}_i = [x_i^1, x_i^2, ..., x_i^L]$ is the i-th sample consisting L histogram features and x_i^l, $l = 1,...,L$ is the l-th histogram feature describing the pattern $\mathbf{x}_i$. In this work, the histogram features are characterized by the use of spectral histograms that model each pixel by considering marginal distributions of responses of a bank of filters [5]. However any histogram feature can be considered. The spectral histograms are computed in three steps [5]. In the first step a sequence of filters is applied to the RS image bands. Then, in the second step, a local histogram for each sample that represents local distributions of the filter responses within a spatial window centered at the corresponding pixel is computed. In the final step, local histograms (obtained at the same pixel location within the set of filtered images) are stacked to define a spectral histogram. By this way, the spectral histogram integrates responses of different filters and thus is an approximation of the underlying distribution of filter responses. The histogram features are given as input to a classifier to generate the corresponding land-cover map. It is worth noting that the spectral histograms can result to be very high-dimensional when: i) a wide range of filters (i.e., a dense sampling of the values of the parameters of the filters) is selected; ii) a high histogram bin number is chosen; and iii) the image consists of several spectral channels. This issue may lead to the Hughes phenomenon, and thus result in a poor generalization ability on classification systems.

To overcome this problem, in this paper, we propose to classify the histogram features by using appropriate kernels instead of using general purpose kernels. To this end, we introduce the AK-SVMs in RS that can generalize well on challenging classification problems (where the features are high dimensional histograms) through efficiently modeling the sparsity of data.

2.2. Additive kernel support vector machines

Let us first analyze the general concept of non-linear binary SVMs. SVMs divide the L-dimensional image feature space into two subspaces, one for each class, using a discriminant function. The decision rule $f(\mathbf{x}_i)$ of SVM in the kernel space is given by

$$f(\mathbf{x}_i) = \sum_{j=1}^{n} y_j \alpha_j K(\mathbf{x}_j, \mathbf{x}_i) + b \tag{1}$$

where $K(\cdot,\cdot)$ is the kernel function, n is the number of support vectors (SVs, which are a subset of the training samples) and $\alpha_j, j = 1,...,n$ are the Lagrange multipliers associated with SVs. A kernel is additive if it can be written as [3]:

$$K(\mathbf{x}_j, \mathbf{x}_i) = \sum_{l=1}^{L} K_l(x_j^l, x_i^l) \tag{2}$$

where $K_l(\cdot,\cdot)$ is the kernel function for the l-th feature only. Then, the resulting SVM decision function is also additive and can be written as:

$$f(\mathbf{x}_i) = \sum_{l=1}^{L} f_l(x_i^l) + b \tag{3}$$

where $f_l(x_i^l) = \sum_{j=1}^{n} y_j \alpha_j K_l(x_j^l, x_i^l)$. The most popular and effective additive kernels presented in the computer-vision community are:

1) histogram intersection (HI) kernel: $K(\mathbf{x}_j, \mathbf{x}_i) = \sum_{l=1}^{L} \min(x_j^l, x_i^l)$
2) chi-square (χ^2) kernel: $K(\mathbf{x}_j, \mathbf{x}_i) = \sum_{l=1}^{L} \frac{2x_j^l x_i^l}{x_j^l + x_i^l}$

Note that the HI and χ^2 kernels are positive definite parameter-free kernels for histogram based features [3]. The histogram based features, which model marginal sample distributions, represent the discrete densities and thus are sparse (i.e., several histogram bins have zero entries). Accordingly, even in the case of high dimensional histogram features, due to the high generalization capability of SVMs with HI and χ^2 kernels, the performance of the AK-SVMs is not degraded by the Hughes phenomenon. Moreover these kernels do not require the definition of any parameters.

In order to address multiclass problems, an ensemble of binary AK-SVMs can be considered. Standard One-Against-All (OAA) or One-Against-One (OAO) strategies can be used. In this work we adopt the OAA strategy, which involves an architecture consisting of one AK-SVM for each information class. Each AK-SVM solves a two-class problem defined by one information class against all the others. A sample is assigned to the class corresponding to the AK-SVM that yields the largest functional distance.

2.3. Fast implementations of additive kernel support vector machines

The computational complexity of the classification phase of the AK-SVM is linearly proportional to the number of SVs derived in the learning phase of the classifier, as in all non-linear SVMs. To classify a sample $\mathbf{x}_i$ with the standard implementation of AK-SVMs, n kernel computations are needed and all the n SVs must be stored in memory (see (1)). Let us assume that these kernels can be computed with a complexity given by $O(L)$, where L is the number of features. Then, the overall complexity of the standard implementation of AK-SVM classifier is $O(Ln)$. Thus, if the value of n (which often is a significant fraction of the training samples) increases, the computational complexity of AK-SVM classification increases. In order to reduce the computational time, here we introduce in RS fast implementations of the AK-SVM [3]. Due to the space constraints, we focus on the fast implementation of AK-SVM with HI kernel only. A similar approach can be used also for the χ^2 kernel. Considering the HI kernel, the decision function in (1) can be rewritten as [3]:

$$f(\mathbf{x}_i) = \sum_{j=1}^{n} y_j \alpha_j \left(\sum_{l=1}^{L} \min\left(x_j^l, x_i^l\right) \right) + b \tag{4}$$

If we change the order of summations in (4), we obtain:

$$f(\mathbf{x}_i) = \sum_{l=1}^{L} \left(\sum_{j=1}^{n} y_j \alpha_j \min\left(x_j^l, x_i^l\right) \right) + b = \sum_{l=1}^{L} f_l(x_i^l) + b \tag{5}$$

where $f_l(x_i^l) = \sum_{j=1}^{n} y_j \alpha_j \min\left(x_j^l, x_i^l\right)$. If the x_j^l values are sorted in increasing order (with corresponding α values and labels y_i) and largest r index is found such that $x_j^l < x_i^l$, $f_l(x_i^l)$ can be calculated using the equation of $f_l(x_i^l) = \sum_{j=1}^{r} y_j \alpha_j x_j^l + x_i^l \sum_{j=r+1}^{n} y_j \alpha_j$. Both first and second summation parts in this equation are independent from x_i^l, and depend only on the SVs and the Lagrange multipliers. Thus, these parts can be pre-computed only once, while r should be estimated with respect to each x_i^l. As a result, $f(\mathbf{x}_i)$ can be computed with a significantly reduced complexity $O(\log(n))$ and the AK-SVM with HI kernel can be implemented with a speed up of $O(n/\log(n))$ without any loss in the classification accuracy. This approach provides the fast exact implementation of AK-SVM with HI kernel.

It is worth noting that the decision function given above may allow approximating the final discrimination function by approximating each dimension independently. To this end piecewise linear polynomial approximation is found very efficient in [3]. As a result, an additional advantage of the AK-SVMs with HI kernel is that its approximate versions can be computed much faster. This results in a computational complexity similar to that of the linear SVM. Thus by keeping the advantages of non-linear SVMs, the AK-SVMs provide more accurate classification results than linear SVMs with a similar computational complexity.

3. EXPERIMENTAL RESULTS

Experiments were conducted on an image acquired by the Quickbird multispectral sensor on the city of Trento (Italy) in October 2005 (see Fig. 1). This image includes the four pan-sharpened multispectral bands and the panchromatic channel with a spatial resolution of 0.7 m. The available ground reference samples are representative of the five land cover classes (i.e., water, road, fields, forest, bare soil). They were randomly divided to derive a test set of 2936 samples and a training set of 2796 samples. Table I shows the land cover classes and the related number of samples used in the experiments.

Fig.1. True color composite of the Trento Quickbird data set

Table 1: Number of samples of each class in the training and the test sets

Land-Cover Class	Training Set	Test Set
Water	419	531
Asphalt	619	602
Field	635	766
Forest	880	820
Bare soil	243	217
Total	2796	2936

In the experiments, we evaluated the effectiveness of the HI and χ^2 kernels for the classification problems by using the spectral histograms. As suggested in [5], the spectral histogram features are obtained using three types of filters: i) the intensity filter; ii) the Laplacian of Gaussian filters; and iii) the Gabor filters. The intensity filter provides the original image itself. The Laplacian of Gaussian filters are obtained with sigma values of 0.2 and 1, while Gabor filters are obtained with orientations of 45° and 90° when variances on both directions are set to 2. In the experiments the number of histogram bins is set to 10 and the local window size is fixed to 5×5. Accordingly, the final size L

of each pattern characterized by spectral histograms is 250. Note that by increasing the histogram bin number and varying the filtering parameter values, the length of the spectral histogram feature vector can significantly increase. In the experiments, we compared the results of the AK-SVM implemented with HI and χ^2 kernels with those obtained by using SVM with the standard RBF and polynomial kernels. The parameters of RBF and polynomial kernels were chosen performing a grid-search model selection, whereas the HI and χ^2 kernels are parameter free. In the model selection, the regularization parameter of SVM is also estimated in all cases.

Table 2 shows the classification results obtained by the SVM classifier when the HI, χ^2, RBF and polynomial kernels are used. From the results one can observe that both HI and χ^2 kernels provide much higher accuracies than RBF and polynomial kernels, while the best performance is obtained by the χ^2 kernel. As an example, the SVM with χ^2 kernel provides almost 4% higher accuracy than those obtained by the SVMs with RBF and polynomial kernels. From these results, we can observe that, the use of HI and χ^2 kernels significantly increases the classification accuracy with respect to general kernels when using histogram features.

Table 2: Overall accuracy obtained by the SVM classifier when HI, χ^2, RBF and polynomial kernels are used

Selected Kernel	Classification Accuracy
HI	**93.87**
χ^2	**95.77**
RBF	91.27
Polynomial	91.12

In the experiments we also compared different implementations of AK-SVMs with the HI kernel: i) standard implementation; ii) fast exact implementation; and iii) fast approximate implementation (see Table 3). From the results one can observe that the computational time of the standard implementation of the AK-SVM with HI kernel is significantly reduced by both the fast exact and fast approximate implementations, while obtaining the same accuracies. For example, in order to reach a classification accuracy of 93.87%, the fast exact implementation of the AK-SVM with HI kernel is about 23 times faster than the standard implementation. Moreover, the fast approximate implementation is 34 times faster than the standard AK-SVM implementation without any loss in the classification accuracy. Note that the number of SVs is found as 75 in these experiments and the gain in time is expected to increase if the number of SVs increases.

Table 3: Overall accuracy and speed up factor obtained by different implementations of AK-SVMs with the HI kernel

Implementation of the AK-SVM	Classification Accuracy	Speed Up
Standard	**93.87**	**-**
Exact	**93.87**	23×
Approximate	93.84	34×

4. CONCLUSION

In this paper we have presented the SVMs with additive kernels (AK-SVM) in the context of RS image classification problems when the histogram features are considered. In particular we introduced in RS two different additive kernels: i) histogram intersection kernel; and ii) chi-square kernel. Moreover, we also presented fast implementations of AK-SVMs with HI kernels. These implementations allow computing the SVM decision function in a relatively small number of calculations, thus significantly speeding up the classification process without any loss in the classification accuracy. As a final remark, we would like to point out that the joint use of histogram features with appropriate kernels and their fast implementations are very promising for RS classification problems due to the achieved: i) high classification accuracy; ii) high generalization capability; and iii) relatively small computational time.

REFERENCES

[1] F. Odone, A. Barla, A. Verri, “Building kernels from binary strings for image matching,” *IEEE Transactions on Image Processing*, vol. 14, no.2, pp. 169-180, 2015.

[2] O. Chapelle, S. P. Haffner, and V. Vapnik. “Support vector machines for histogram-based image classification”, *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1055–1064, May 1999.

[3] S. Maji, A.C. Berg, and J. Malik, “Efficient classification for additive kernel SVMs,” *IEEE Transactions on Pattern Analysis and Machine Intelligenc*e, vol. 35, no.1, pp. 66-77, 2013.

[4] B. Demir, L. Bruzzone, “A Novel active learning method in relevance feedback for content based remote sensing image retrieval,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, pp. 2323–2334, 2015.

[5] J. Yuan, D. Wang, and R. Li, “Remote sensing image segmentation by combining spectral and texture features”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp.16-24, Jan. 2014.

[6] B. Demir , L. Bruzzone, “ Histogram based attribute profiles for classification of very high resolution remote sensing images”, *International Conference on Geoscience and Remote Sensing Symposium*, Milan, Italy, 2015.