SUPPORT VECTOR MACHINE AND BATHACHARRYA KERNEL FUNCTION FOR REGION BASED CLASSIFICATION

Rogério Galante Negri, Luciano Vieira Dutra, Sidnei João Siqueira Sant'Anna

Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos - SP, Brazil

ABSTRACT

Region based methods are indicated to classify image with strong heterogeneity, where only the spectral information is not enough. Different approaches have been proposed to perform this kind of classification. This study presents a new approach for region based classification that consists in use the Support Vector Machine (SVM) method with Bhattacharyya kernel function. A high resolution IKONOS image was classified. The classification results shows that SVM method using the Bhattacharyya kernel is better than Minimum Distance Classifier and conventional SVM.

***Index Terms*—** Region based classification, Support Vector Machine, stochastic distances, Bhattacharyya kernel function

1. INTRODUCTION

Image classification is a term referred to pattern recognition techniques which are applied on images aiming to recognize different elements of a scene. Remote sensing of the Earth is one of the areas where image classification techniques have been widely used. The main goal of classification techniques for images acquired by remote sensing satellites is to identify and to label land use and land cover classes.

Usually the classification is based on the information extracted from the image elements. For a long period, the information extraction was conducted using only the spectral characteristics of the pixels, by the so-called "pixel based methods". However, due to the high information variability (scene heterogeneity) introduced by the very high resolution images, the use of only spectral information becomes unsatisfactory [1].

Alternatively, the "region based classification methods" seem to be a good solution to this problem. Firstly, these methods aggregate pixels into homogeneous objects by a segmentation algorithm and then each object is classified individually [2]. Spectral, texture and spatial attributes can be extracted from previously defined object, and used in the classification procedure. The classification process is usually carried out by conventional techniques, such as Maximum Likelihood Classification [3], Support Vector Machine [2] or Minimum Distance Classifier based on stochastic distances [4].

Knowing the relevance of the image region based classification this study presents a new approach to use Support Vector Machine on region based classification, proposing a kernel function derived from a stochastic distance. An experiment with high resolution remote sensing image shows the superiority of this approach in comparison with the approaches presented in [2] and [4].

2. STOCHASTIC REGION BASED CLASSIFICATION

Stochastic distances come from the measures of Information and Entropy, present on Information Theory formalized in [5]. This theory is strictly based on statistical concepts and was designed to determine the information transfer capability of a communication channel.

Starting from concepts of Information and Entropy, other measures have been proposed to quantify information, resulting in the development of a general concept called Divergence (Relative Entropy). The Divergence can be used to measure the discriminability between two sets of information, and it can be treated as random variables, according to the distance between their probability distributions.

A formal treatment presented by [6] defines a general class of divergence measures called "$h - \phi$ divergences". Different stochastic distances are obtained from the $h - \phi$ divergences, for example, Kullback-Leibler, Jensen-Shannon and Bhattacharyya [7].

The Bhattacharyya distance is a classical stochastic distance commonly used in image processing applications, given by:

$$B(\mathbf{U},\mathbf{V}) = -\log\left(\int_{\mathbf{r}\in\mathcal{S}} \sqrt{f_{\mathbf{U}}\left(\mathbf{r};\Theta_{\mathbf{U}}\right) f_{\mathbf{V}}\left(\mathbf{r};\Theta_{\mathbf{V}}\right)} d\mathbf{r}\right) \tag{1}$$

where $f_{\mathbf{U}}$ and $f_{\mathbf{V}}$ are probability density functions with parameters $\Theta_{\mathbf{U}}$ and $\Theta_{\mathbf{V}}$, respectively related to the random variables $\mathbf{U}$ and $\mathbf{V}$, which are defined on the same probability space $\mathcal{S}$.

Depending on the application, different probability density functions can be used in (1). When $f_{\mathbf{U}}$ and $f_{\mathbf{V}}$ correspond

Thanks to CAPES, CNPq (Process: 307666/2011-5) and FAPESP (Process: 08/58112-0 and 08/57719-9) agencies for funding.

to the Multivariate Gaussian distribution, (1) is rewritten as:

$$B_G(\mathbf{U},\mathbf{V}) = \frac{1}{8}\left(\mu_\mathbf{U}-\mu_\mathbf{V}\right)^T\left(\frac{\Sigma_\mathbf{U}+\Sigma_\mathbf{V}}{2}\right)^{-1}\left(\mu_\mathbf{U}-\mu_\mathbf{V}\right)+ \\ +\frac{1}{2}\ln\left(\frac{0.5|\Sigma_\mathbf{U}+\Sigma_\mathbf{V}|}{|\Sigma_\mathbf{U}|+|\Sigma_\mathbf{V}|}\right) \tag{2}$$

where $\mu_\mathbf{Z}$ and $\Sigma_\mathbf{Z}$ are the average vector and covariance matrix estimated for modeling the random variable $\mathbf{Z}$, with $(\cdot)^T$, $|\cdot|$ and $(\cdot)^{-1}$ denoting respectively the transpose, determinant and inverse operations.

In [4] Minimum Distance Classifier based on stochastic distances is used on region classification problems. This approach starts with a training set $\mathcal{D}_l = \{(\mathbf{x}_i, \omega_j) \in \mathcal{X}\times\Omega;\ i = 1, 2, \ldots, m < n;\ j = 1, 2, \ldots, c\}$ composed by training regions defined on the attribute space $\mathcal{X}$ with classes on Ω. The trainig regions are treated as random variables. Regions $\mathbf{x}$ in $\mathcal{D}_l$ labeled as ω_j are used to estimate probability density functions f_{ω_j}. Based on the estimated functions for each class, regions $\mathbf{x}_i$, $i = 1, \ldots, n$; of a given image $\mathcal{I} \subset \mathcal{X}$ are associated to ω_j, and then denoted by $(\mathbf{x}_i, \omega_j)$, when the stochastic distance $M(f_{\mathbf{x}_i}, f_{\omega_j})$ is minimal, i.e.:

$$(\mathbf{x}_i, \omega_j) \Leftrightarrow \underset{j=1,\ldots,c}{\arg\min}\, M(f_{\mathbf{x}_i}, f_{\omega_j}) \tag{3}$$

where $f_{\mathbf{x}_i}$ represents a probability density function that models $\mathbf{x}_i$, which are considered as random variables.

3. SUPPORT VECTOR MACHINE AND BHATTACHARYYA KERNEL FUNCTION

Support Vector Machine (SVM) is a classification method that has received great attention in recent years. The SVM has excellent generalization capability, it is independent of data distribution and it presents robustness to the Huge's phenomena. For more details about this method the work of Theodoridis and Koutroumbas ([8]) is recommended.

An important and attractive characteristic on SVM is the possibility of use kernel functions. These functions provide conditions to apply SVM in different types of data and problems. The kernel functions must be symmetric and must satisfy the Mercer's conditions. However, such verification is not always trivial. Fortunately, there are alternative ways to develop such functions, for example, based on radial basis functions model, written as [9]:

$$k(\mathbf{x}_u, \mathbf{x}_v) = g(d(\mathbf{x}_u, \mathbf{x}_v)) \tag{4}$$

where $\mathbf{x}_u, \mathbf{x}_v$ are patterns in $\mathcal{X}$, $d : \mathcal{X}^2 \mapsto \mathbb{R}$ is a distance measure and $g : \mathbb{R} \mapsto \mathbb{R}$ is a strictly positive continuous real function, for instance, the exponential function $g(z) = e^{-z}$.

The patterns used in (4) are not restricted to numerical vectors. Different data types, such as strings, graphs and sets, can be manipulated when an appropriate distance measure is adopted. This condition allows the integration of stochastic distances to radial basis function (RBF) model, resulting in kernels able to treat regions delimited over the images as patterns in a classification problem. Several combinations of $g(\cdot,\cdot)$ and $d(\cdot)$ functions can be used to derive a kernel function. Adopting, respectively, the distance B_G and the negative exponential functions, the following kernel function is defined:

$$k(\mathbf{x}_u, \mathbf{x}_v) = e^{-B(\mathbf{x}_u, \mathbf{x}_v)} \tag{5}$$

where $\mathbf{x}_u$ e $\mathbf{x}_v$ represents regions over an image $\mathcal{I} \subset \mathcal{X}$.

The function (5) was initially proposed in [10], where it is called Bhattacharyya Kernel. In [11] is presented an application of this kernel function for classification of audio signals by Neural Networks. A kernel function based on Kulback-Leilber divergence is proposed in [12] for multimedia data classification. In this study is applied the Bhattacharyya Kernel on region based classification by SVM.

4. EXPERIMENTS AND RESULTS

In [2] the SVM method is applied to region classification using as input the mean and standard deviation information of regions in each spectral band. Under this consideration the regions are reduced to vectors, and then the classification may be conducted by the conventional SVM method. The RBF $k(\mathbf{x}_u, \mathbf{x}_v) = e^{-\gamma\|\mathbf{x}_u-\mathbf{x}_v\|^2}$ was used as kernel funtion and the "one-against-one" strategy was applied to address the multi-class problem. For comparison purposes this strategy is also used in this work.

In [4] the region classification problem was addressed by Minimum Distance Classifier method based on stochastic distances. In this study was also used the same approach.

The classification performance was assessed in the three mentioned approaches (SVM with Bhattacharyya kernel, SVM based on mean and standar deviation and Minimum Distance Classifier based on stochastic distance) by the overall accuracy measure. This measure consists on the percentage of correctly classified pixels with respect to validation samples.

The image used in this study corresponds to an urban area of São José dos Campos city, São Paulo State, Brazil, acquired by IKONOS Satellite on June 27, 2007. This image has size of 500×500 pixels and it is illustrated in Figure 1(a). The training and validation samples used in the classification process are described in Figure 1(b) and Table 1. A Region Growing algorithm [13] was used to segment the study image.

The classification results obtained by the different approaches are shown in Figure 2 and its confusion matrices in Tables 2, 3 and 4. The SVM method associated with the kernel function presents 94.26% of overall accuracy and it overcame the classification results presented in [2] and [4], which reached accuracies of 64.31% and 76.66%, respectively. The results show that the use of proposed kernel function allows the accurately classification of multi-modal classes, i.e., classes with different spectral responses, as "roof" and "soil" classes. This fact was not fully verified in the other two classification approaches (SVM based on mean and standar deviation and Minimum Distance Classifier based on stochastic distance). From the confusion matrices it can be seen that misclassification is strongly diminished by using SVM with Bhattacharyya kernel.

Table 2. Classification confusion matrix of SVM with Bhattacharyya kernel function (%).

Ref.\Class.	Roof	Street	L.Veg	River	Tree	Soil
Roof	89.3	5.5	0	0	3.2	2
Street	0.5	93.2	6.3	0	0	0
L.Veg	0	0	100	0	0	0
River	0	0	0	100	0	0
Tree	0	0	0	0.2	99.8	0
Soil	8	0	0	0	0	92

Table 3. Classification confusion matrix of SVM based on mean and standard deviation information (%).

Ref.\Class.	Roof	Street	L.Veg	River	Tree	Soil
Roof	68.3	15.6	1.6	0	0	14.5
Street	2.5	91.2	6.3	0	0	0
L.Veg	0	9.3	90.7	0	0	0
River	0	0	0	0	100	0
Tree	0	0	16.1	0	83.9	0
Soil	46.3	3.8	3.3	0	11.5	35

Table 4. Classification confusion matrix of Minimum Distance Classifier based on stochastic distance (%).

Ref.\Class.	Roof	Street	L.Veg	River	Tree	Soil
Roof	56.5	5.4	0	0	0	38.1
Street	0.5	93.2	6.3	0	0	0
L.Veg	0	0	100	0	0	0
River	0	0	0	100	0	0
Tree	0	0	0	0.2	99.8	0
Soil	23.8	0	14.2	0	11.6	50.4

5. CONCLUSIONS

According to the results, the kernel function proposed in this study is indicated when the SVM method is used in region based classification procedures. This classification procedure presented the best accuracy result. The classification based on stochastic distances also showed satisfactory results, however, its accuracy classifying multimodal classes is worse than SVM methods. The SVM method based only on region mean and standard deviation had the worst classification results. Also, employing a similar approach, different new kernel functions can be derived by using other stochastic distances. These kernel functions can be applied to other data types, such as SAR images.

6. REFERENCES

[1] X. Gigandet, M. B. Cuadra, A. Pointet, L. Cammoun, R. Caloz, and J. Thiran, "Region-based satellite image classification: method and validation," in *IEEE International Conference on Image Processing*, 2005, vol. 3, pp. III–832–5.

[2] D. Liu and F. Xia, "Assessing object-based classification: advantages and limitations," *Remote Sensing Letters*, vol. 1, no. 4, pp. 187–194, 2010.

[3] V. Walter, "Object-based classification of remote sensing data for change detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, no. 3-4, pp. 225–238, 2004.

[4] W. B. Silva, L. O. Pereira, S. J. S. Sant'Anna, C. C. Freitas, R. J. P. S. Guimarães, and A. C. Frery, "Land cover discrimination at Brazilian Amazon using region based classifier and stochastic distance," in *2011 IEEE International Geoscience and Remote Sensing Symposium*, 2011, pp. 2900–2903.

[5] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, 1948.

[6] M. Salicru, D. Morales, M. L. Menendez, and L. Pardo, "On the applications of divergence type measures in testing statistical hypotheses," *Journal of Multivariate Analysis*, vol. 51, no. 2, pp. 372–391, 1994.

[7] A.D.C. Nascimento, R.J. Cintra, and A.C. Frery, "Hypothesis testing in speckled data with stochastic distances," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 1, pp. 373 –385, jan. 2010.

[8] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Fourth Edition*, Academic Press, 4th edition, 2008.

[9] B. Schölkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*, Adaptive computation and machine learning. MIT Press, 2002.

[10] R. Kondor and T. Jebara, "A kernel between sets of vectors," in *In International Conference on Machine Learning (ICML)*, 2003.

[11] J. Y. Kim and D. C. Park, "Application of Bhattacharyya kernel-based centroid neural network to the classification of audio signals," in *International Joint Conference on Neural Networks*, 2009, pp. 2948–2952.

[12] P. J. Moreno, P. P. Ho, and N. Vasconcelos, "A Kullback-Leibler divergence based kernel for svm classification in multimedia applications," in *Advances in Neural Information Processing Systems*. 2004, MIT Press.

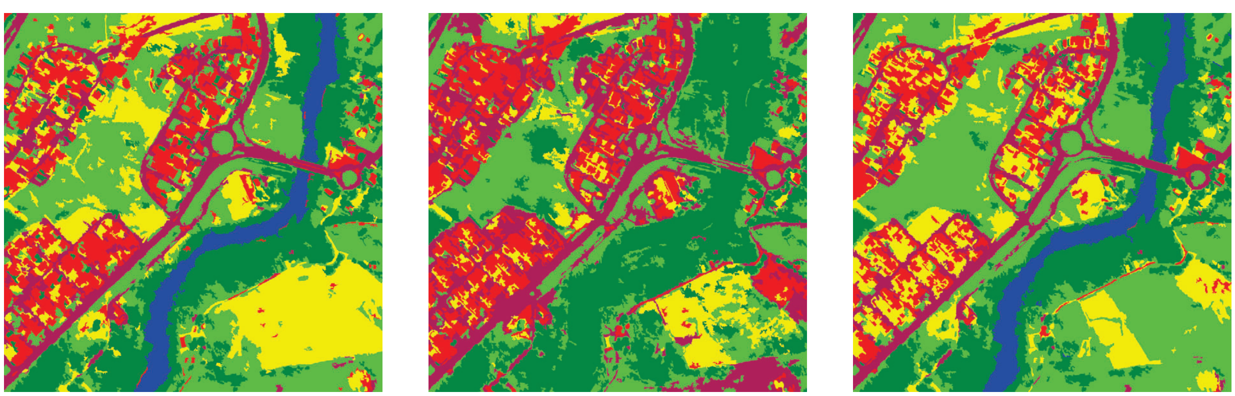
(a) IKONOS image of São José dos Campos, Brazil.

(b) Segmented image, representing the regions according to the spectral average, training samples (solid) and validation (hatched).

Fig. 1. Image and the classes samples.

Table 1. Training and validation samples size.

Class	**Training** # **of pixels**	**Training** # **of polygons**	**Validation** # **of pixels**	**Validation** # **of polygons**
Roof	2367	21	1231	12
Street	1513	15	557	7
Low Vegetation	976	11	464	6
River	620	5	372	3
Tree	1407	10	628	5
Soil	1358	13	702	8



(a) SVM classification based on Bhattacharyya kernel function.

(b) SVM classification based on region mean and standard deviation.

(c) Minimum Distance Classifier based on stochastic distance.

Fig. 2. Classification results obtained by the analyzed methods.

[13] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001.