

Stochastic Approaches of Minimum Distance Method for Region Based Classification

Rogério G. Negri, Luciano V. Dutra, and Sidnei J.S. Sant'Anna

Instituto Nacional de Pesquisas Espaciais – INPE, São José dos Campos - SP, Brasil
Divisão de Processamento de Imagens – DPI
{rogerio,dutra,sidnei}@dpi.inpe.br

Abstract. Normally remote sensing image classification is performed pixelwise which produces a noisy classification. One way of improving such results is dividing the classification process in two steps. First, uniform regions by some criterion are detected and afterwards each unlabeled region is assigned to class of the “nearest” class using a so-called stochastic distance. The statistics are estimated by taking in account all the reference pixels. Three variations are investigated. The first variation is to assign to the unlabeled region a class that has the minimum average distance between this region and each one of reference samples of that class. The second is to assign the class of the closest reference sample. The third is to assign the most frequent class of the k closest reference regions. A simulation study is done to assess the performances. The simulations suggested that the most robust and simple approach is the second variation.

Keywords: region based classification, stochastic distances, image simulation, remote sensing.

1 Introduction

Traditionally land cover classification, using remote sensing data, is performed pixelwise, also known as pixel based classification. The main drawback of this approach is producing a noisy classification map.

One way of improving such results is using what is called a *contextual classifier*, where the pixel classification depends somehow on the state of its neighbours [1]. Another traditional approach is dividing the image classification process in two steps. First, uniform regions by some criterion are detected in the imagery and afterwards each region is assigned to just one class. This unknown region is assigned to the class closest to it, in terms of a so-called stochastic distance [2].

Stochastic distance is a generic name to a class of distances that measures the *difference* between two distributions, it is zero when two distributions are equal. The reference statistics is calculated using all data of the several training samples of a particular class.

In this paper we will investigate three variations of the aforementioned approach. A Monte Carlo study is done to assess the performance of the proposed modifications.

2 Stochastic Distances

The Jeffries-Matusita (JM), a well known stochastic distance between two distributions is defined as [3]:

$$JM(\mathbf{U}, \mathbf{V}) = \int_{\mathbf{x} \in \mathcal{X}} \left[\sqrt{f_{\mathbf{U}}(\mathbf{x}; \Theta_U)} - \sqrt{f_{\mathbf{V}}(\mathbf{x}; \Theta_V)} \right]^2 d\mathbf{x} \quad (1)$$

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} , defined on the same probability space \mathcal{X} .

From the equation (1) one can observe that if $f_{\mathbf{U}}$ and $f_{\mathbf{V}}$ are equal this distance is zero, and if these distributions are totally different this distance comes to two. For classes normally distributed the JM distance is written as [3]:

$$JM(\mathbf{U}, \mathbf{V}) = 2 \left(1 - e^{-B(\mathbf{U}, \mathbf{V})} \right) \quad (2)$$

where $B(\cdot, \cdot)$ is the Bhattacharyya distance, which for normal distributions is written as:

$$B(\mathbf{U}, \mathbf{V}) = \frac{1}{8} (\mu_{\mathbf{U}} - \mu_{\mathbf{V}})^T \left(\frac{\Sigma_{\mathbf{U}} + \Sigma_{\mathbf{V}}}{2} \right)^{-1} (\mu_{\mathbf{U}} - \mu_{\mathbf{V}}) + \frac{1}{2} \ln \left(\frac{0.5 |\Sigma_{\mathbf{U}} + \Sigma_{\mathbf{V}}|}{|\Sigma_{\mathbf{U}}| + |\Sigma_{\mathbf{V}}|} \right) \quad (3)$$

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

3 Region Based Classification Using Stochastic Distances

Formally, the pixel based classification process consists in associating the classes $\omega_j \in \Omega$, $j = 1, 2, \dots, c$; for each pattern (pixel) $\mathbf{x}_i \in \mathcal{I} \subset \mathcal{X}$, $i = 1, 2, \dots, n$, from a function $g : \mathcal{X} \rightarrow \Omega$, called classifier. In this context, \mathcal{I} represents an image composed by patterns \mathbf{x} defined on the *attribute space* \mathcal{X} . The different image classification methods can be understood as different ways of modeling the function g and applying on \mathcal{I} .

For region based classification the images are initially organized into different regions using *image segmentation* techniques. More details about these techniques can be found in [4].

In the region based approach the object is no longer pixels but regions. A region \mathbf{R}_l , $l = 1, \dots, L$ is a set of spatially connected pixels that attends a particular uniformity criterion defined by the segmentor, $\bigcup_{l=1}^L \mathbf{R}_l = \mathcal{I}$. In region based classification the whole region is assigned at once, to a particular class. All pixels in the \mathbf{R}_l region has the same class.

To train a supervised region based methodology it is required a set of labelled regions $\mathcal{D} = \{(\mathbf{R}_l, \omega_j) \in \mathcal{X} \times \Omega, l = 1, \dots, N; j = 1, \dots, c\}$, where N is the total number of training regions. The standard *Minimum Stochastic Distance Classifier* uses the information of pixels distribution inside the unlabelled region to

associate it to the class which has the closest distribution. The classes distributions are modeled using the information of \mathcal{D} . Being \mathbf{R}_l an unlabelled region, i.e. $l \notin \mathcal{D}$, $M(f_{\mathbf{R}_l}, f_{\omega_j})$ a stochastic distance between $f_{\mathbf{R}_l}$, the distribution of \mathbf{R}_l and f_{ω_j} , the distribution of the class ω_j , an assignment of ω_j to \mathbf{R}_l , denoted by (\mathbf{R}_l, ω_j) , is made if the following rule is satisfied:

$$(\mathbf{R}_l, \omega_j) \Leftrightarrow j = \arg \min_{j=1, \dots, c} M(f_{\mathbf{R}_l}, f_{\omega_j}) \quad (4)$$

The class distribution are estimated taking into account all the pixels that are present inside the many labelled regions taken as reference to a particular class. In this work we investigate other alternatives for the standard classification rule defined in (4).

The *Minimum Mean Stochastic Distance* rule (M_{mean}) is obtained by the arithmetic mean of the stochastic distances between a given unlabelled region and each one of training regions of the same class. This distance can be defined as:

$$M_{mean}(f_{\mathbf{R}_l}, f_{\omega_j}) = \frac{1}{t_j} \sum_{k=1}^{t_j} M(f_{\mathbf{R}_l}, f_{\omega_{j\mathbf{R}_k}}) \quad (5)$$

where $f_{\omega_{j\mathbf{R}_k}}$ is a probability density function that models the k -th training region of ω_j , which has t_j training regions in \mathcal{D} .

Other option is the M_{min} distance, which is the minimum distance between an unlabelled region \mathbf{R}_l and each one of the training regions of ω_j . This rule is written as:

$$M_{min}(f_{\mathbf{R}_l}, f_{\omega_j}) = \min \left\{ M(f_{\mathbf{R}_l}, f_{\omega_{j\mathbf{R}_k}}) : k = 1, \dots, t_j \right\} \quad (6)$$

Another option is suggested by (6). It would be the stochastic distance form of the k -Nearest-Neighbors classifier (M_{knn}), which is defined by associating a class to an unlabelled region as the most frequent class of the k -nearest training regions.

4 Experiments and Results

To evaluate the classification results a Monte Carlo study was performed. The study was based on simulated images having optical characteristics. The simulation process is divided into three main steps: creating phantom image, computing statistical parameters and simulating image patches. The phantom image is an idealized cartoon model, which contains the image regions. Mean vector and covariance matrix are statistics extracted from different targets (classes) in real image. The adopted image model is introduced in the image patch simulation step (image synthesis).

For this study its was created a phantom image that is formed by concatenating six identical blocks of 512×512 pixels, partitioned in 44 segments having

different forms and sizes. Each block will represent a distinct class. For classification purposes the upper quarter of each block was selected as training samples. Figure 1(a) illustrates the phantom image and the respective training regions.

A four band (visible and near infrared) LANDSAT5/TM image was used to estimate statistical parameters (mean vector and covariance matrix) of six distinct classes. The TM image was acquired on September 26th 2010 and covers an area located along BR-163 highway, next to the Tapajós National Forest in Pará State, Brazil. The land cover classes used on the statistical estimation procedure were *primary forest*, *old regeneration* (secondary forest older than 20 years old), *new regeneration* (secondary forest newer than eight years old), *clean pasture*, *dirty pasture* and *agriculture fields*.

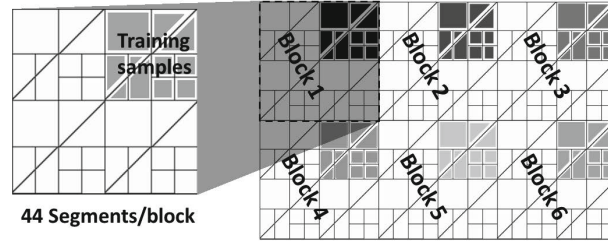
As mentioned, the optical characteristics are required for the simulated images, therefore a Multivariate Gaussian distribution is assumed for the data distribution. The image synthesis step consists of a joint Gaussian distribution and the expression (7) which defines the covariance structure of the simulated image patches. The expression for a single pixel $\tilde{\mathbf{p}}_{ab}$ of the segment b of block a is given by:

$$\tilde{\mathbf{p}}_{ab} = [(E_a \times L_a \times \nu) \times \zeta_b] + (\mu_a \times \psi_b) \quad (7)$$

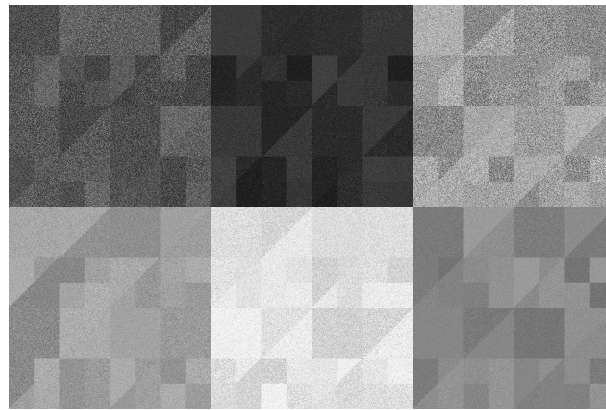
where μ_a represents a q -dimensional mean vector of the class a , E_a is the eigenvectors matrix derived from the covariance matrix of class a (Σ_a), L_a is the diagonal matrix defined by the square root of the eigenvalues of Σ_a and ν is a q -dimensional random vector generated by a standard Multivariate Gaussian distribution. The parameters ζ_b and ψ_b are random scalars uniformly distributed, with $b \in \{1, 2, \dots, 44\}$ indexing each segment. They were adopted to model the mean and variance intraclass fluctuations. Note that the index a , in this simulation case, ranges from 1 to 6, since it is used to distinguish each image block.

The columns of E_a are organized according to the descending order of their respective eigenvalues, as well as the diagonal elements of L_a . The variance information of each band is determined by L_a . It is important to notice that for this simulation experiment the values of ζ_b and ψ_b ranged from 0.55 to 1.45 and from 0.90 to 1.10, respectively, as they are commonly found in real case. The classification results were analyzed based on 100 simulated images having four spectral bands and six classes. In Figure 1(b) is illustrated one spectral band of a simulated image, where the fluctuations can be clearly observed.

The simulated images have been classified by the *Minimum Distance Classifier* defined in (4) and by its derivatives M_{mean} , M_{min} and M_{knn} , described, respectively by (5), (6), and the last phrase of section 3. To conduct the Monte Carlo study three different scenarios were employed for the classification analysis. The scenarios were characterized by the number of classes (six, four or two) present on the images. The images having four classes were designed by grouping the image blocks 1, 4 and 5 (*primary forest*, *clean pasture* and *new regeneration* classes) as a single class. On the other hand, grouping the image blocks 1, 5 and 6 (*primary forest*, *new regeneration* and *dirty pasture* classes) as a single class and the image blocks 2, 3 and 4 (*agriculture*, *old regeneration* and *clean pasture*



(a) Phantom image and location of the training regions



(b) A band of the simulated image, in gray scale

Fig. 1. The *primary forest*, *agriculture*, *old regeneration*, *clean pasture*, *new regeneration* and *dirty pasture* classes are represented by blocks 1 to 6, respectively

classes) in another single class generate the two classes images. These grouping process aims to verify the classifier capability of dealing with heterogeneous classes.

Classification is performed for all set of classes using the JM distance as M in the (2) and its derivatives. It was adopted $k = 3$ for M_{knn} . Overall accuracy is computed for each classification case considering the 33 segments not used as training regions.

Figures 2, 3 and 4 present the simulation results for images containing 6, 4 and 2 classes respectively. Table 1 brings the mean overall accuracy (OA) and its standard deviation for the simulated results. It is possible to observe that M_{min} and M_{knn} produce the highest OA values for all studied cases, indicating that they are quite robust to multi-modality, which is also reflected in Table 1. The average overall accuracy of M_{min} and M_{knn} have been found statistically indifferent at 95% of confidence level. Figure 5(a) is the reference map, Figures 5(b), (c) and (d) are typical examples of M_{knn} , M_{mean} and M classifications for two classes.

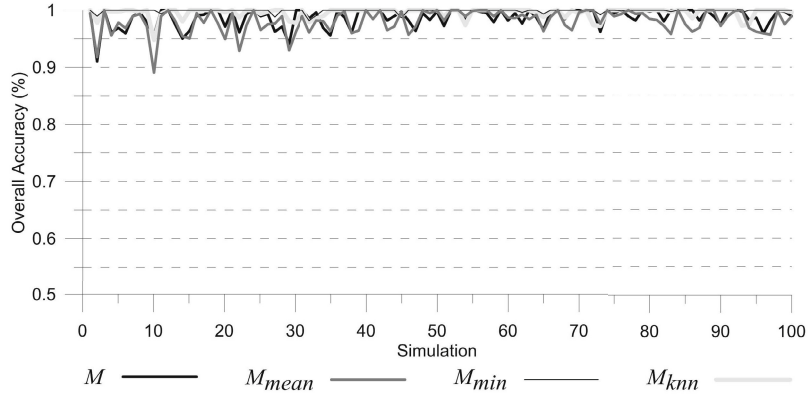


Fig. 2. Overall accuracy results for six classes image

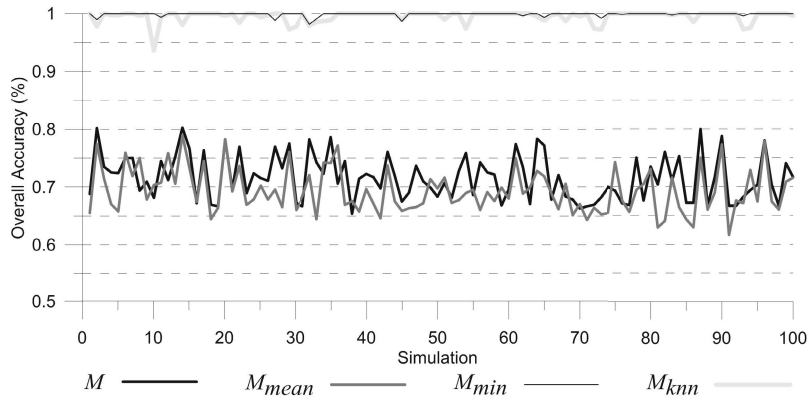


Fig. 3. Overall accuracy results for four classes image

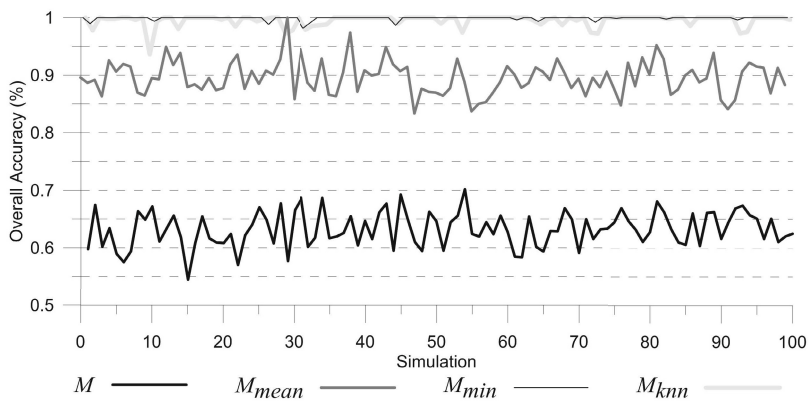
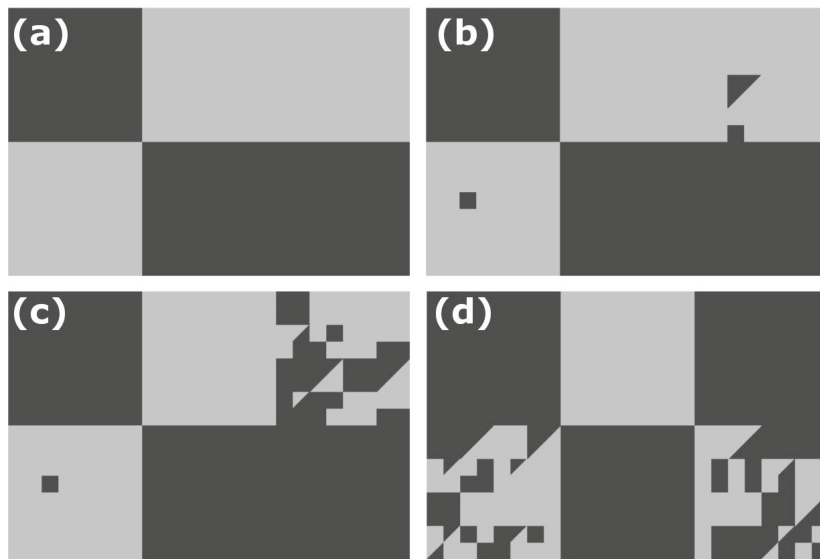


Fig. 4. Overall accuracy results for two classes image

Table 1. Overall accuracy results

# Classes		M	M_{mean}	M_{min}	M_{knn}
6	Average	0.987	0.981	0.999	0.996
	Std. Deviation	0.016	0.020	0.003	0.009
4	Average	0.716	0.694	0.999	0.995
	Std. Deviation	0.039	0.039	0.003	0.010
2	Average	0.632	0.896	0.999	0.995
	Std. Deviation	0.030	0.029	0.003	0.010

**Fig. 5.** Typical examples of classification results for two classes: (a) reference map, (b) M_{min} classification, (c) M_{mean} classification, (d) M classification

5 Conclusions

Unlike what is commonly used for region based classification which is the M function based on JM distance, the M_{min} or M_{knn} rule would be preferred because its robustness. Also as the M_{min} rule is not statistically different from the M_{knn} rule then M_{min} is preferred because it is simpler and quicker.

The conclusion implies that the user have to be careful to collect training samples, at least one representative for each aspect of the class to be mapped. One example is the use of region based classification for classifying roofs tops in high resolution imagery, which is a problem intrinsically multi-modal. Instead of training each type of roof tops separately, it is possible to add samples of each type of roof in a general roof top category, and then the classification is expected to work properly using the M_{min} version.

Acknowledgments. This work was supported by CAPES, CNPq (Grant # 307666/2011-5) and FAPESP (Grant # 08/58112-0 and 08/57719-9).

References

1. Kartikeyan, B., Gopalakrishna, B., Kalubarme, M.H., Majumder, K.L.: Contextual techniques for classification of high and low resolution remote sensing data. *International Journal of Remote Sensing* 15(5), 1037–1051 (1994)
2. Silva, W.B., Pereira, L.O., Sant'Anna, S.J.S., Freitas, C.C., Guimarães, R.J.P.S., Frery, A.C.: Land cover discrimination at brazilian Amazon using region based classifier and stochastic distance. In: 2011 IEEE International Geoscience and Remote Sensing Symposium, pp. 2900–2903 (2011)
3. Richards, J.A., Jia, X.: *Remote Sensing Digital Image Analysis: An Introduction*. Springer (2005)
4. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 3rd edn. Prentice-Hall, Inc., Upper Saddle River (2006)