# Classification of the real remotely sensed image covered with clouds

## Lijana Stabingienė

*Department of Statistics, Klaipeda University*
H. Manto 84, LT-92294 Klaipėda
E-mail: lijana.stabingiene@gmail.com

**Abstract.** In this paper supervised classification method is proposed. It is based on Bayes discriminant functions (BDF) and it deals with the problem of optimal classification for images, which are corrupted by natural phenomenon such as cloud, smoke or fog. Solving such a problem is very important when we have remotely sensed information, which very often is corrupted by clouds. For example, the remotely sensed images from the territory of Lithuania are very often corrupted by clouds. The idea of classification, using BDF with incorporated spatial dependency between the observation to be classified and the training sample is presented in earlier works of the author. The novelty of this paper is the method how to use these methods for the real situation, i.e. for the remotely sensed image which is naturally covered by clouds. Visual and numerical results are presented in this paper, which show the advantage of this method against BDF ignoring spatial dependency between training sample and observation to be classified and against the method using grey level cooccurrence matrices.

**Keywords:** image classification, Gaussian random fields, Bayes discriminant function, supervised classification, semivariance.

## Introduction

Spatial classification is a problem of labeling pixels based on feature information and information about spatial adjacency relationships with training sample. Usually assumed that feature observations conditional on labels are independent (conditional independence) and normally distributed and the labels follow the Random Field model. This approach is widely used in image classification [7]. Atkinson and Naser proposed the geostatistically weighted classifier for remote sensed imagery and applied it to the remotely sensed image from IKONOS satellite [1]. Remotely sensed image classification is the process, when the thematic map is created from the image [11].

The incorporation of the spatial information (image texture, direction, closeness and other) into remotely sensed image classification is highly potential [5]. Dučinskas [2] was the one of the first authors who retracted the assumption of the independency, that observations to be classified are independent on training sample. In authors work [2] the classification problem is solved, when the observation must be separated into two classes according to the class label points with some feature information and according to the spatial dependency connections to the training sample. When more accurate calculations of spatial correlation are incorporated, the smaller classification errors are produced [3, 4].

According to the Tobler's first law of geography: everything is related to everything else, but near things are more related than distant things [10]. The formal property that describes this is spatial autocorrelation. Spatial autocorrelation represents the degree to which that correlation changes with distance [6]. The natural phenomena like clouds, fog and smoke have the form of correlated random filed, the dependence of intensity decreases when the distance between two points grows. Such fields can be modeled by the Gaussian random filed model.

In the present paper the method for image per pixel classification is proposed and it is applied for the real remotely sensed image, which is naturally covered by the clouds. In the previous papers [3, 8, 9] this methods was used only for artificially generated images, when the existence of the spatial correlation and spatial correlation range parameter were already known.

## 1 The main concepts and definitions

In this paper features are modeled by the Gaussian random field $\{Z(s)\colon s \in D\}$, where $s$ means pixel from the image and $D \subset \mathbb{R}^2$. The marginal model of observation $Z(s)$ in class $\Omega_l$ is $Z(s) = \mu_l + \varepsilon(s)$, where $\mu_l$ is the mean, and the error term is generated by the stationary Gaussian random field $\{\varepsilon(s)\colon s \in D\}$ with covariance function defined by model $\mathrm{cov}\{\varepsilon(s), \varepsilon(u)\} = \sigma^2 r(s-u)$ for all $s, u \in D$, where $r(s-u)$ is the spatial correlation function and $\sigma^2$ is the variance.

Let $L = \{1, 2\}$ be the label set, and let $S_n = \{s_i \in D;\ i = 1, \ldots, n\}$ be the set of training pixels (STP). Denote by $Y = (Y(s_1), \ldots, Y(s_n))'$, $Z = (Z(s_1), \ldots, Z(s_n))'$ the labels and features vector, and denote by $T' = (Z', Y')$ the training sample.

The model of $Z$ for given $Y = y$ is $Z = X_y \mu + E$, where $X_y$ is a design matrix, $\mu' = (\mu_1, \mu_2)$ and $E$ is the n-vector of random errors that has multivariate Gaussian distribution $N_n(0, \sigma^2 R)$.

$r_0$ is vector of spatial correlations between $Z_0$ and $Z_n$ and $R$ is matrix of spatial correlations among components of $Z_n$. $Z_0$ is correlated with training sample, so we have to deal with conditional Gaussian distribution of $Z_0$ given $T = t$ ($Z = z, Y = y$) with means $\mu_{lt}^0$ and variance $\sigma_{0t}^2$.

**Proposition 1.** *The conditional distribution of $Y(s_0)$ given $T = t$ depends only on $Y = y$, i.e. $\pi_l(y) = P(Y(s_0) = l | T = t)$, $l = 1, 2$.*

Under the assumption of complete parametric certainty of populations, the Bayes discriminant function (BDF) minimizing the probability of misclassification (PMC) is formed by the logarithm of ratio of conditional densities described in Eq. (1)

$$W_t(Z_0; \Psi) = \left(Z_0 - \frac{1}{2}\left(\mu_{1t}^0 + \mu_{2t}^0\right)\right)\left(\mu_{1t}^0 - \mu_{2t}^0\right)/\sigma_0^2 + \gamma(y), \tag{1}$$

where $\gamma(y) = \ln(\pi_1(y)/\pi_2(y))$ and $\Psi = (\beta', \theta')'$. Denote the three component vector of parameter estimates by $\hat{\Psi}' = (\hat{\mu}', \hat{\sigma}^2)$. Then PBDF associated with BDF specified in Eq. (1) is

$$W_t(Z_0; \hat{\Psi}) = \left(Z_0 - \frac{1}{2}\left(\hat{\mu}_{1t}^0 + \hat{\mu}_{2t}^0\right)\right)\left(\hat{\mu}_{1t}^0 - \hat{\mu}_{2t}^0\right)/\hat{\sigma}_{2t}^0 + \gamma(y), \tag{2}$$

where $\hat{\mu}_{lt}^0 = E(Z_0|T = t; \ Y(s_0) = l) = \mu'_l + \alpha'_0(z_n - X_y\hat{\mu})$, $l = 1, 2$, $l = 1, 2$ and $\hat{\sigma}_{0t}^2 = V(Z_0|T = t; \ Y(s_0) = l) = \hat{\sigma}^2 R_{0n}$, where for $l = 1, 2$, $\hat{\mu}_{lt}^0 = \hat{\mu}_l + \alpha'_0(z_n - X_y\hat{\mu})$, and $\hat{\sigma}_{0t}^2 = \hat{\sigma}^2 R_{0n}$. Denote it by PBDFD. If $Z_0$ is assumed to be independent to $T$, then PBDF has the following form

$$W(Z_0; \hat{\Psi}) = \left(Z_0 - \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2)\right)(\hat{\mu}_1 - \hat{\mu}_2)/\hat{\sigma}^2 + \gamma(y), \tag{3}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are the estimates of $\mu$ and $\sigma^2$, based on $T = t$. Denote it by PBDFI.

## 2 Numerical example

In this example the described methodic Eq. (2) is applied on a real Lithuania territory image, which is naturally covered by clouds. This image is obtained with the Landsat7 satellite [12]. Just a part of the whole remotely sensed image is used for the experiment. The cropped image dimensions are $200 \times 200$ pixels and it contains two classes, i.e. first class is the forest and the second is non forest Fig. 1. The original image is naturally corrupted with clouds and such noise is modelled by Gaussian random field (GRF) with zero mean and exponential spatial correlation function given by $r(h) = \exp\{-|h|^2/\alpha\}$. Here $\alpha$ is a spatial correlation range parameter which must be estimated. This parameter is evaluated with function variofit from the geoR package in the environment of the $R$ system. For supervised classification methods Eqs. (2) and (3) the training sample with $n_1 = n_2 = 100$ is selected Fig. 1(b), (c) and 4, 8, 12 nearest neighbour schemes are used.

The remotely sensed image used for classification is naturally corrupted and then the exact correlation range parameter is unknown. This parameter is estimated according to the training sample points (pixels) using the geoR package of the statistical program $R$.

At the same time both classes training sample points are used in order to identify the correlation range parameter. In order to minimize the influence of different classes to the accuracy of the modeling, the means of corresponding classes are subtracted from the training sample points feature values. This way transformed points with their coordinates are then used to calculate empirical semivariogram. It is done with geoR packages command variog. The parametric model is then fitted to the empirical semivariogram points using variofit command, which uses least squares method for



(a)　　　　　　　　(b)　　　　　　　　(c)
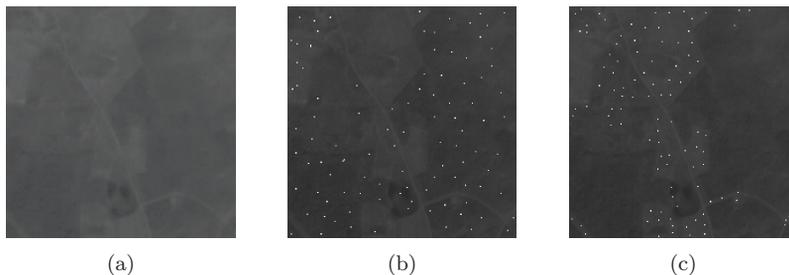
**Fig. 1.** (a) Image covered by clouds, used for classification, (b) training sample for the forest class, (c) training sample for the non forest class.
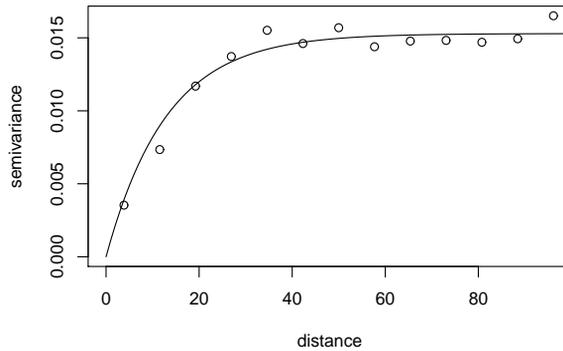
**Fig. 2.** Exponential semivariance model best fitted to the experiment data.

**Table 1.** The overall accuracy of classification.

| NN scheme | PBDFD | PBDFI | GLCM |
|-----------|--------|--------|--------|
| 4 | 0.9033 | 0.9020 | |
| 8 | 0.8817 | 0.8747 | 0.7722 |
| 12 | 0.8779 | 0.8623 | |

fitting. After several trials, several different types of models were fitted according to the shape of the empirical semivariogram, it was determined that for this concrete situation best model was of exponential type. Fitted model is presented in Fig. 2.

The value of the correlation range parameter $\alpha = 13.0305$ was estimated. This parameter is used for further classification using two different methods Eqs. (2), (3). The results of the classification are shown in Table 1 numerically and in Fig. 3 visually. The accuracy is calculated according to the classified image of the identically the same territory obtained by the same satellite but three months later and this image is not covered by any clouds. The presented errors are obtained by using three different nearest neighbourhood (NN) schemes. The point is classified according to the information obtained from the number of nearest points from the training sample. The classification results are also compared with unsupervised classification method based on grey level cooccurrence matrices (GLCM) [5], using average parameter and window of size 7.
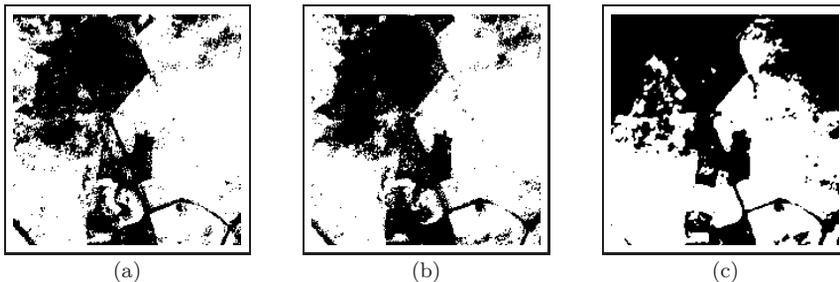


**Fig. 3.** (a) Classification results for PBDFD, (b) and PBDFI, (c) methods using 12 nearest neighbour scheme and method based on GLCM.

## 3  Conclusions

Proposed supervised method Eq. (2) perform better then the method ignoring dependency on class labels from the training sample Eq. (3).

The bigger number of the nearest neighbours used for classification not always increases the classification accuracy.

PBDFI method gives smoother visual result, but it misclassifies places which are hard to separate (more corrupted by clouds), when the PBDFD method classifies them more accurately.

Unsupervised classification method, based on GLCM performs rather well, but misclassifies places which are corrupted by clouds stronger.

The modeling of the semivariogram gave rather big correlation range parameter, so it can be stated, that clouds can be modeled with Gaussian random fields.

## References

[1] P.M. Atkinson and D.K. Naser. A geostatistically weighted k-NN classifier for remotely sensed imagery. *Geogr. Anal.*, **42**:204–225, 2010.

[2] K. Ducinskas. Approximation of the expected error rate in classificattion of the Gaussian random field observations. *Stat. Prob. Lett.*, **79**:138–144, 2009.

[3] K. Ducinskas and L. Stabingienė. Expected Bayes error rate in supervised classification of spatial Gaussian data. *Informatica*, **22**(3):371–381, 2011.

[4] K. Ducinskas, L. Stabingienė and G. Stabingis. Image classification based on Bayes discriminant functions. *Proc. Env. Sci.*, **7**:218–223, 2011.

[5] R.M. Haralick. Statistical and structural approaches to texture. *Proc. IEEE*, **67**:786–804, 1979.

[6] J.G. Liu and P.J. Mason. *Essential Image Processing and GIS for Remote Sensing.* Wiley-Blackwell, UK, 9600 Garsington Road, Oxford, OX4 2DQ, UK 111 River Street, Hoboken, NJ 07030-5774, USA, 2009.

[7] R. Nishii and S. Eguchi. Image classification based on Markov random field model with Jeffreys divergence. *J. Mult. Anal.*, **97**(9):1997–2008, 2006.

[8] L. Stabingienė, G. Stabingis and K. Dučinskas. Comparison of linear discriminant functions in image classification. *Liet. mat. rink. LMD darbai*, **51**, 2010.

[9] L. Stabingienė, G. Stabingis and K. Dučinskas. Comparison of the classification methods for the images modeled by Gaussian random fields. *Liet. mat. rink. LMD darbai*, **52**:200–204, 2011.

[10] W. Tobler. A computer movie simulating urban growth in Detroit region. *Econ. Geogr.*, **46**:234–240, 1970.

[11] B. Tso and P. M. Mather. *Classification methods for remotely sensed data.* Taylor and Francis, New York, 2001.

[12] USGS Earth Explorer. EarthExplorer. Available from Internet: http://edcsns17.cr.usgs.gov/NewEarthExplorer.

REZIUMĖ

## Klasifikavimas realaus nuotolinio stebėjimo vaizdo, padengto debesimis

*L. Stabingienė*

Darbe pasiūlyta klasifikavimo su mokymu metodika, paremta Bajeso diskriminantinėmis funkcijomis (BDF), skirta optimaliai klasifikuoti vaizdams, kurie yra sugadinti tokių reiškinių kaip debesys, dūmai ar rūkas. Problemos išsprendimas yra ypač aktualus tada, kai turime palydovinių nuotraukų informaciją, kur gana dažnai vaizdas yra padengtas debesimis. Vaizdo klasifikavimo idėja, naudojant BDF su inkorporuota erdvine priklausomybe tarp klasifikuojamo stebinio ir mokymo imties, yra aprašyta ankstesniuose autorės darbuose. Šio darbo naujumas yra tas, jog siūloma metodika yra pritaikyta realiai situacijai, t. y., nuotolinio stebėjimo vaizdui natūraliai sugadintam debesimis. Darbe pateikti vaizdiniai ir skaitiniai rezultatai rodo pasiūlytos metodikos pranašumą prieš BDF, ignoruojančias erdvinę priklausomybę ir prieš GLCM metodus.

*Raktiniai žodžiai*: vaizdų klasifikavimas, GRF, BDF, klasifikavimas su mokymu, semivariograma.