Applicability of multifractal features as global characteristics of WorldView-2 panchromatic satellite images

Wojciech Drzewiecki1*, Anna Wawrzaszek2, Michał Krupiński2, Sebastian Aleksandrowicz2 and Katarzyna Bernat1

1AGH University of Science and Technology, al. Mickiewicza 30, 30 - 059 Kraków, Poland
2Space Research Centre of the Polish Academy of Sciences, ul. Bartycka 18A, 00 - 716 Warsaw, Poland
*Corresponding author, e-mail address: drzewiec@agh.edu.pl

Abstract
In this work we analyse fractal and multifractal characteristics for description and extraction of information from VHR satellite images. We propose the degree of multifractality as a global descriptor of satellite image content and investigate its usefulness for classification of WorldView-2 image chips into main land cover types. The research confirmed that it is possible to use the textural features as efficient global descriptors of WorldView-2 panchromatic image content. Results show that the degree of multifractality is related to land cover type prevailing in the imaged area. It was also proved that multifractal parameters should be considered as valuable textural features in the context of land cover classification.

Keywords: Multifractals, texture, classification, panchromatic VHR satellite images, WorldView-2.

Introduction
Image texture is considered as a primary factor of visual perception and one of the most crucial features used commonly in computer vision and content-based image retrieval (CBIR) [Shyu et al., 1998]. It is usually easy to recognize texture, but it is more difficult to define it, because texture, in contrast to colour, is not determined by a single point, but involves neighbouring area and can be related to a direction or a scale. Large spectrum of parameters has been created due to a lot of possible textural descriptors, to help to extract information about texture, also in the context of satellite images [Howarth and Ruger, 2007]. Textural characteristic can be calculated based on the entire image (global features), fragments of the image delineated by segmentation results or small groups of pixels formed by fixed or moving windows [Smeulders et al., 2000; Emerson et al., 2004; Datta et al., 2008].

The textural analysis becomes also an important component of the process of information extraction from satellite images. Different texture analysis techniques have been successfully used to describe the content of the images [Smeulders et al., 2000; Datta et al.,
2008; Samal et al., 2009]. They can be divided into four groups [Li et al., 2014]: structural
e.g. mathematical morphology) [Haralick, 1979; Haralick et al., 1987], statistical (e.g. image
first-order and second-order statistics) [Haralick, 1979], model-based (e.g. fractals
or autoregressive models) [De Souza, 1982; Mandelbrot, 1983] and transform texture
extraction techniques (e.g. wavelets) [Mallat, 1989].

When single-band panchromatic images are considered, textural analysis is necessary for
the automatic classification of their content [Erner and Düzgün, 2009]. However, as Cerra
and Datcu [2010] noticed, the major increase in spatial resolution of satellite sensors caused
the substantial changes in the information content of satellite imagery. They argued for
multiresolution (multi-scale) approach for textural characterization of Very High Resolution
(VHR) satellite images and proposed the approach based on Gibbs-Markov Random Fields.
The complexity and inhomogeneity of the satellite image content considered in different
scales is also reported by Sun et al. [2006] as a reason for sometimes poor performance of
fractals as image texture descriptors. In order to more adequately describe the inhomogeneous
datasets, the multifractal formalism has been proposed [Halsey et al., 1986; Lam and De
Cola, 1993; Sun et al., 2006; Wawrzaszek and Macek, 2010]. Multifractals have been
successfully used for analysis of medical images [Stojić et al., 2006], but the approach has
not been fully explored for remote sensing applications. It has been mainly used for SAR
(synthetic aperture radar) images and image segmentation [Du and Yeo, 2002; Blaschke
et al., 2004; Tso and Mather, 2009]. However, some works suggest that multifractals can
be also successfully applied to optical VHR satellite images. Voorons et al. [2003] used
multifractals for unsupervised classification of Ikonos image. Our initial study has shown
the potential of multifractal parameters as global descriptors of panchromatic VHR satellite
images [Drzewiecki et al., 2013; Wawrzaszek et al., 2013].

The aim of the research presented here is to investigate more thoroughly if the
multifractal formalism, as a generalization of the fractal geometry, may be used for better
characterization of panchromatic VHR image texture. In particular we want to compare
the fractal dimension and the degree of multifractality as global descriptors of satellite
image content and verify if any of these features may be useful for discrimination of basic
land cover types for CBIR applications. For this purpose we want to i) compare the values
of fractal and multifractal characteristics calculated for over 1000 image chips (512×512
pixels) cut from panchromatic WorldView-2 images; ii) check if it is possible to adequately
classify these image chips into categories related to the land cover type prevailing inside the
imaged area based on fractal and multifractal features; iii) compare the informative value
and classification performance of fractal and multifractal parameters with other kinds of
global textural descriptors.

**Theoretical background**

**Fractal dimension**

The fractal dimension ($D_f$) measures the complexity of considered structures and is used
to describe a broad spectrum of data, including selected aspects of satellite remote sensing
[Mandelbrot, 1983; Sun et al., 2006]. The $D_f$ can be used to characterize both: local
behaviour around pixel and global when only single value for the entire image is computed.
In the frame of this work we treat the fractal dimension as a global descriptor and compute
its value by using Differential Box-Counting (DBC) method. This method proposed by Sarkar and Chaudhuri [1994] calculates fractal dimension $D_F$ by covering an object with boxes of varying size $r \to 0$ and analyzing the scaling relation:

$$N(r) \propto r^{-D_F} \quad [1]$$

where $N(r)$ denotes the number of boxes of size $r$ needed to cover the considered object. It is worth to underline that the same relation [1] is used in standard Box-Counting method. However, the major limitation of the box-counting method lies in the fact that the counting process of nonempty boxes $N(r)$ implies its use only for binary images. In DBC algorithm to approximate the quantity $N(r)$, a grayscale images are directly used (without image binarization). This is the significant advantage of DBC method, main steps of which are presented below. The image of size $m \times m$ is considered as a three-dimensional spatial surface $(x, y, z)$, where $(x, y)$ states pixel position and $z$ denotes the gray level of pixel. Then, the $(x, y)$ plane is divided into grids of size $s \times s$, where $\frac{m}{2} \geq s > 1$, and $s$ is an integer. In other words, the size of the grid is varied, starting from half the size of the image, and reduced to a double pixel. On each grid there is a column of cubical boxes of size $s \times s \times s'$ covering the grid. The $s'$ denotes height of each box and is calculated by using relation $[G/s']=[m/s]$, where $[.\.\.]$ indicates integer part operator and $G$ is the total number of gray levels. The boxes are indexed in the ascending order, starting from 1 at the bottom. If on the $(i, j)$th grid the minimum gray level of the image falls into box number and the maximum into box number, then the box-number covered image surface is $n_r(i, j) = l - k + 1$. The integration from all the grids at the specific grid ratio, $r = \frac{s}{m}$ allows to determine the number of boxes intersecting image intensity surface corresponding to each $r$:

$$N(r) = \sum_{i,j} n_r(i, j) \quad [2]$$

Finally, the fractal dimension $D_F$ is estimated from the least square linear fit of log $N(r)$ against log $1/r$. For more details, please refer to the work by Sarkar and Chaudhuri [1994].

**Degree of multifractality**

In our work we use multifractal formalism, as an extension of fractal theory, in order to globally describe the satellite image. Multifractal image analysis base on a statement that image can be treated as a multifractal which is a nontrivial combination of a number of fractals. Hence, description of multifractal inner structure, as the result of intertwined fractals, demands the consideration of the whole spectrum of the generalized dimensions $D_q$, unlike in the fractal case where only one dimension $D_F$ is studied by Wawrzaszek et al. [2013]. In general, multifractals can be described by a group of parameters, one of which is the degree of multifractality ($\Delta$)- a quantitative parameter we propose to use for the global description of satellite image. The degree of multifractality $\Delta$ confirmed its usefulness in the classification of the various 1D and 2D data, in particular in the analysis of the solar wind turbulence [Wawrzaszek and Macek, 2010; Macek et al., 2012].

In order to compute the degree of multifractality for real data we consider variability of the...
spectrum of dimensions as a nonincreasing function of real index $q$ (e.g., see Wawrzaszek and Macek [2010], Fig. 1a). The definition of $D_q$ was formulated in the context of analysis of natural measure of chaotic attractors by Grassberger [1983] and Hentschel and Procaccia [1983]. However, we emphasize that $D_q$ can also be applied to any measures [Ott, 1993, chapter 9], what was confirmed by numerous studies [Meneveau and Sreenivasan, 1991; Wawrzaszek and Macek, 2010; Chakraborty et al., 2014; Wawrzaszek et al., 2014].

In the frame of this work for the direct estimation of the function $D_q$ from the experimental data we use the Box-Counting Based Moment method [Halsey et al., 1986], presented together with schematic diagram in Wawrzaszek et al. [2014]. In the first step of this method the image of the size $m \times m$ is divided into $N(\delta) = (m/\delta)^2$ of square boxes of the size $\delta \times \delta$ (starting with a single pixel size $\delta = 1$ and ending with a box of image the size $\delta = m$, (see Wawrzaszek et al. [2014], Fig.1). For a given box the normalized measure is calculated according to the formula:

$$\mu(\delta) = \frac{p_i(\delta)}{\sum_{i=1}^{N(\delta)} p_i(\delta)} \quad [3]$$

where $i = 1, \ldots, N(\delta)$ labels the individual boxes sized $\delta$, and $p_i(\delta)$ denotes one of the three capacities: SUM, MAXimum or Biggest Calculated Deviation of grey levels (BCD) [Stojić et al., 2006]. More precisely, mentioned capacities are defined as follows:

$$p_i^{\text{SUM}}(\delta) = \sum_{(k, l) \in \Omega_i} g(k, l),$$

$$p_i^{\text{MAX}}(\delta) = \max_{(k, l) \in \Omega_i} g(k, l),$$

$$p_i^{\text{BCD}}(\delta) = \max_{(k, l) \in \Omega_i} |g(k, l) - \langle g \rangle|,$$

where $g(k, l)$ is a gray-scale intensity at point $(k, l)$, $\Omega_i$ is a set of all pixels $(k, l)$ in the box and $\langle g \rangle$ denotes the mean value of gray levels on the image. It is worth noting that $p_i(\delta)$ can be interpreted as the weight of the $i$th box and the denominator in Equation [3] is the total weight of the image. Therefore $\mu_i(\delta)$ denotes the relative weight of the $i$th box or the probability assigned to the $i$th box. As we said before in the frame of multifractal analysis different measures can be constructed to emphasize various effects and to describe different physical processes in considered data [Meneveau and Sreenivasan, 1991; Wawrzaszek and Macek, 2010]. In our opinion, the opportunity to define various measures (Eq. [4]) creates great possibilities for image texture analysis. With SUM measure (based on the sum of pixel values in considered boxes) one can investigate the self-similarity of the image (i.e. the pattern of brightness values), what is the case of fractal analysis, too. However, taking into consideration other capacities, it gives us the possibility of analysing the self-similarity of chosen image properties. Using MAX measure we are able to analyse the pattern of the highest image pixel values (i.e. the distribution of bright objects within the image). BCD measure, in turn, gives us the possibility of analysing the self-similarity of the pixel deviation from mean image value (i.e. pattern of objects contrasted with the image.
background). The approach may be extended further, as one can define more capacities in such a way that other image properties, important in particular application, are investigated (for example MIN measure may be designed to analyse the distribution of darkest pixels within an image).

In the next step of the Box-Counting Based Moment method the partition function $\chi(q, \delta)$ for various values of $\delta$ and $q$ is computed according to:

$$\chi(q, \delta) = \sum_{i=1}^{N(\delta)} (\mu_i(\delta))^q \quad [5]$$

As $q$ varies in Equation [5] different subsets associated with different measure density become dominant. For a multifractal measures, this partition function scales with the box size $\delta \rightarrow 0$ as:

$$\chi(q, \delta) \propto \delta^{D_q(q-1)} \quad [6]$$

Basing on the Equations [5] and [6] we obtain the spectrum of generalized dimensions $D_q$ defined by [Meneveau and Sreenivasan, 1991]:

$$D_q = \frac{1}{1-q} \lim_{\delta \rightarrow 0} \log \frac{\sum_{i=1}^{N(\delta)} (\mu_i(\delta))^q}{-\log \delta} \quad [7]$$

It is worth noting that the case when $q = 0$ corresponds to capacity dimension:

$$D_q = \lim_{\delta \rightarrow 0} \frac{\log N(\delta)}{-\log \delta} \quad [8]$$

which should be interpreted as the fractal dimension of the support of the measure, in other words, as the dimension of points set in which measure is concentrated [Halsey et al., 1986]. When $q = 1$ L’Hospital’s rule can be applied to give:

$$D_1 = \lim_{\delta \rightarrow 0} \frac{\sum_{i=1}^{N(\delta)} \mu_i(\delta) \log \mu_i(\delta)}{-\log \delta} \quad [9]$$

which is known as the information dimension (Eq. [28]). As $q \rightarrow +\infty$, the largest measure value $\mu_{\text{max}}$ dominates the sum in Equation [7] and we have $D_{+\infty} = \lim_{\delta \rightarrow 0} \frac{\log \mu_{\text{max}}}{\log \delta}$ [Meneveau and Sreenivasan, 1991]. In the case when $q \rightarrow -\infty$ the smallest measure value $\mu_{\text{min}}$ will dominate the sum and $D_{-\infty} = \lim_{\delta \rightarrow 0} \frac{\log \mu_{\text{min}}}{\log \delta}$.

Then, the difference of the maximum and minimum dimension $D_q$, associated with the
least dense and the most dense regions in the considered measure, defines the degree of multifractality and is given by Wawrzaszek and Macek [2010]:

\[ \Delta = D_{\infty} - D_{+\infty} \]  

[10]

In particular, for monofractal measures, where no additional information is gained by examining higher moments \( q, D_q \), is a constant function of \( q \) and then degree of multifractality equals zero. Alternatively, for multifractals the \( D_q \), as the sensitive to changes both strong and weak heterogeneities of measure, becomes a nonlinear function of \( q \) and \( \Delta > 0 \).

Theoretically, the generalized dimension function \( D_q \) is defined for all real values of \( q \). In practice, owing to the limited data set we can only determine values of \( D_q \) for the narrowed number of moments \( q \). In our analysis we performed analysis for \(-5 \leq q \leq 5\) and \( \Delta \) has been calculated as the difference between \( D_{-5} \) and \( D_{+5} \), while \( \Delta_p \) as the difference between \( D_0 \) and \( D_{+5} \). It is worth to underline that the separate analysis of the parameter \( \Delta_p \) have been performed for two reasons. Firstly, the consideration of the results for positive moments \( q \) allows to take into account only the high concentration of measure (regions of the image with high intensity values), dwarfing the smaller ones, what seems to be also useful in the texture description. Secondly, errors in the calculation of the generalized dimensions for positive \( q \) are smaller than for \( q < 0 \), therefore the the uncertainties of the determining parameter \( \Delta_p \) are also smaller than for \( \Delta \).

Finally, as a result of multifractal analysis performed for each image we obtain the following set of parameters: the degree of multifractality \( \Delta \) for SUM (\( \Delta_{SUM} \)), MAX (\( \Delta_{MAX} \)) and BCD (\( \Delta_{BCD} \)) capacities, as well as the degree of multifractality for positive values of index \( q (\Delta_{p_{SUM}}, \Delta_{p_{MAX}} \) and \( \Delta_{p_{BCD}} \)).

It is worth to underline that standard statistical analysis based on the second-order moments characterizes a measure around the mean value and can also be used as the information about the heterogeneity of the measure on a given scale \( \delta \). In multifractal analysis, considering the partition function (see Eq. [5]), we can control how the moments of measure scale with \( \delta \) and the information about this scaling can be summarized in the form of multifractal function \( D_q \) and in the calculated degree of multifractality. Therefore degree of multifractality compared with standard deviation gives rather global information about the heterogeneity of measure. Moreover for some multifractal measures where extreme values need to be taken into account, the first and second statistical moments may not be sufficient for characterizing heterogeneity. Therefore higher-order moments magnifying large (\( q > 0 \)) or small (\( q < 0 \)) concentrations of measure are taken into account in multifractal analysis. For more discussion about this issue see [Cheng, 1999].

**Methods**

**Data**

In our research we used Very High Resolution panchromatic WorldView-2 (WV2) images. Three scenes covering different areas of Poland were divided into subsets (image chips), 512x512 pixels each. We have selected this size of subsets to provide reliable scaling expressed by Equations [1] and [6], and then good accuracy of fractal and multifractal features calculation. For analyses we considered only four broadly defined land cover
classes: agriculture, forest, urban and water. Subsets where dominating land cover class covered less than 75% of the imaged area were excluded from analysis. Each of remained 1031 image chips was manually labelled based on the prevailing land cover class as assessed by experienced photo interpreter (164 as agriculture, 329 - forest, 244 - urban, 294 - water). The resulting sets of samples of each land cover class are still internally diverse (Fig. 1).

Figure 1 - Examples of WorldView-2 subsets for four land cover classes: agriculture (a, b), forest (c, d), urban (e, f) and water (g, h).
Comparison of fractal and multifractal characteristics

For each of 1031 WorldView-2 image subsets included into the analysis the fractal dimension $D_F$ was calculated as calculation of the degree of multifractality $\Delta$ was performed with three different measures (SUM, MAX and BCD) according to Equation [4]. We decided to compare these characteristics and initially evaluate the possibility of their usage as image content descriptors based on the analysis of boxplots.

Comparison of classification performance

In this study we wanted to check the usefulness of calculated fractal and multifractal features for classification of image chips into prevailing land cover classes for content-based image retrieval purposes. Moreover, we also decided to compare these two kinds of parameters with other types of global image descriptors, based on image histogram as well as on five different texture analysis techniques (calculated using MaZda software [Szczypiński et al., 2009]). As result every image chip was described by 295 attributes, which may be grouped into 9 groups:

a) the label (land cover class);

b) four histogram-based characteristics (mean, variance, skewness and kurtosis);

c) fractal dimension $D_F$;

d) six multifractal parameters ($\Delta_{SUM}, \Delta_{MAX}, \Delta_{BCD}, \Delta_p^{SUM}, \Delta_p^{MAX}$ and $\Delta_p^{BCD}$);

e) 220 grey-level co-occurrence matrix-based parameters [Haralick, 1979; Walker et al., 1997]: angular second moment, contrast, correlation, sum of squares, inverse difference moment, sum average, sum variance, entropy, difference entropy; these parameters were computed 20 times, for (d, 0), (0, d), (d, d), (d, -d) where the distance d can take values of 1, 2, 3, 4, 5;

f) 20 run length matrix-based parameters [Galloway, 1975; Haralick, 1979]: run length nonuniformity, gray level nonuniformity, long run emphasis moment, short run emphasis inverse moment, fraction of image in runs; these parameters were computed 4 times (for vertical, horizontal, 45-degree and 135-degree directions);

g) 5 absolute gradient-based parameters [Haralick, 1979]: mean absolute gradient, variance of absolute gradient, skewness of absolute gradient, kurtosis of absolute gradient, percentage of pixels with nonzero gradient;

h) 5 first-order autoregressive model parameters [Jain, 1989; Hajek et al., 2006]: $\theta_1$, $\theta_2$, $\theta_3$, $\theta_4$, $\sigma$. In this model the weighted sum of four neighbouring pixel values is used for prediction of pixel value. Model parameters are the weights assigned to left, top, top-left and top-right neighbours ($\theta_1$- $\theta_4$ respectively) as well as the variance of minimized prediction error ($\sigma$) [Szczypiński et al., 2009]

i) 20 parameters derived from wavelet analysis [Mallat, 1989]. These are energies computed within frequency channels obtained from scaling the original image up to five times, both in horizontal and vertical directions, using discrete Haar wavelet [Szczypiński et al., 2009].

In content-based image retrieval framework supervised classification is used for fast and accurate image retrieval as well as automatic indexation of images in image database [Datta et al., 2008]. In our study we used machine learning approaches, namely support vector machines (SVMs) and classification trees (Random Forest (RF) as well as C5.0
algorithms). These classifiers has been reported as giving good results for a wide spectrum of classification problems [Fernández-Delgado et al., 2014].

Support vector machines is non-parametric statistical learning technique first developed by Vapnik and presented in [Vapnik and Chervonenkis, 1971]. Since then, the approach has evolved and now is considered as one of the most effective machine learning methods [Kuhn and Johnson, 2013]. A comprehensive presentation may be found in [Vapnik, 2010]. The method aims to find a hyperplane that separates training dataset into known classes in such way that the distance (called margin) between classification boundary and the closest training set point is maximized. As a result only the subset of training data (called support vectors) defining the hyperplane of maximum margin is used for definition of decision boundary. Originally linear classifier was extended to nonlinear classification boundaries thanks to implementation of kernel function [Boser et al., 1992]. Linear, polynomial, radial basis function (RBF) and sigmoid kernels are the most frequently used [Qian et al., 2015].

Random Forest [Breiman, 2001] is reported as representative of modern discriminative classification methods, having the prediction strength competitive also to other methods, such as support vector machines, AdaBoost or neural networks [Hapfelmeier and Ulm, 1993; Schindler, 2012]. The algorithm uses an ensemble of classification trees, usually built with CART algorithm. Each of these trees is built on randomly selected subsets of training data and using a random subset of classification features at each split. Final classification decision is based on the averaged responses of each tree.

Like Random Forest, C5.0 is a classification method based on classification trees. However, the approaches vary in many aspects and, in some cases, the algorithm is able to outperform the random forest approach [Kuhn and Johnson, 2013]. C5.0 is a more advanced version of C4.5 algorithm [Quinlan, 1993] and differs from CART in criteria used for constructing the trees and the pruning approach, among others. In our experiment we took advantage of C5.0 boosting capability, when ensemble of classifiers is constructed and they are voted for final classification decision. The C5.0 boosting procedure is based on AdaBoost algorithm [Freund and Schapire, 1997], however the prediction from constituent trees are combined in a different way [Kuhn and Johnson, 2013]. The fuzzy threshold option can also be chosen, which causes softening of split tests for continuous variables. If attribute value is close to the threshold both tree branches are evaluated and the results combined probabilistically.

Classification was done based on 16 different sets of image descriptors:

- a) histogram-based features;
- b) fractal dimension and histogram-based features;
- c) multifractal parameters;
- d) multifractal parameters and histogram-based features;
- e) grey-level co-occurrence matrix-based features;
- f) grey-level co-occurrence matrix-based features and histogram-based features;
- g) run length matrix-based features;
- h) run length matrix-based features and histogram-based features;
- i) absolute gradient-based features;
- j) absolute gradient-based features and histogram-based features;
- k) autoregressive model parameters;
- l) autoregressive model parameters and histogram-based features;
m) parameters derived from wavelet analysis;
  n) parameters derived from wavelet analysis and histogram-based features;
  o) all classification features;
  p) all textural classification features (without histogram-based ones).

In the case of SVMs we used the polynomial kernel. Because the classifier performance is very dependent on parameter values [Mountrakis et al., 2011; Ustuner et al., 2015], we decided to determine the appropriate values of polynomial degree as well as cost parameter and scaling factor using model tuning approach. Linear, quadratic and cubic models were tested with 11 values of cost parameter $C = \{2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6, 2^7, 2^8\}$ and 4 values of scale factor (0.001, 0.005, 0.01 and 0.05).

Random Forest was used for classification with different combinations of $m_{\text{tray}}$ and $n_{\text{trees}}$ parameters. The former is the number of randomly selected prediction variables to choose from at each split. The latter is number of trees. For particular classification tries $m_{\text{tray}}$ parameter was set as a square root or one third of the number of predictors and $n_{\text{trees}}$ as 500 or 1000.

In the case of C5.0, four attempts with different approaches to pruning and with or without fuzzy thresholding were tested. Boosting with ten trials was used in every classification run. Accuracy of classification was evaluated by repeated cross-validation. In each classification try the stratified random split was performed on the dataset and 80% of data were used for training and 20% for evaluation. The accuracy of particular classification approach was assessed based on repetitions of 5-fold cross-validation realizations (50 classification tries).

**Reduction of classification feature set**

In case of high-dimensional data, determining the set of classification features which should be used in the model may be very important [Kuhn and Johnson, 2013]. A model based on lower number of predictors is usually easier for interpretation. Moreover, in case of classification of VHR panchromatic images, calculation of textural parameters may be very computationally expensive. We decided to perform such kind of analysis and choose the most informative subset from the full set of 294 classification features. We were interested if any of proposed multifractal parameters would be included in.

Both classification trees algorithms implemented in our study have the ability to identify informative variables. In case of Random Forest, several feature selection methods have been suggested in literature (e.g. [Svetnik et al., 2004; Diaz-Uriarte and Alvarez de Andrés, 2006; Sandri and Zuccolotto, 2006; Wang et al., 2010; Kuhn and Johnson, 2013]). We decided to employ the approach of Diaz-Uriarte and Alvarez de Andrés [2006] used previously in remote sensing applications [Chehata et al., 2009; Guan et al., 2012]. It uses the variable importance measure calculated in Random Forest based on a decrease of classification accuracy when randomly permuting the values of each predictor (one at a time) for each tree. Based on this ranking the selection of most relevant features is done by iterative backward elimination of the least important variables. Random forests are fit iteratively and in each iteration some fraction (Díaz-Uriarte and Alvarez de Andrés [2006] suggest 20%) of least important features is dropped. The set of selected features may be determined as the one that yields the lowest OOB (out-of-bag) error or as the smallest set of features whose error rate is within 1 standard error of minimum error rate. In second case
usually fewer variables are selected while error rate is still close to the minimal one. In case of C5.0 for reducing the number of predictors we applied its winnowing capability. To winnow non-informative predictors the tree is created from the randomly chosen half of training samples. If the classification feature is not present in any split in this tree it is considered as unimportant. The training samples not used for creation of this tree are used to estimate the error rate of the tree. It is then compared to error rates estimated without each of predictors. If error rate without the predictor is lower, it is removed from the set of classification features [Kuhn and Johnson, 2013].

To verify if multifractal parameters will remain in the reduced sets of classification features we decided to perform the feature selection procedures for one RF and one C5.0 classification attempt. In the case of Random Forest the smallest set of features resulting in OOB classification error within 1 standard error of minimum error rate was determined. The \( m_{\text{tree}} \) parameter was set as a square root of predictors number and a number of trees as 500. For C5.0 fuzzy thresholds and global pruning were employed. Pruning CF parameter was set as 10\% and boosting with ten trials was used. Both approaches were applied to two sets of image descriptors - the full one and the one consisted of all textural classification features.

To select the most important features for SVMs we applied a backward selection algorithm (called recursive feature elimination - RFE) described in [Guyon et al., 2002]. In this approach after the model is created the classification variables are ranked according to their importance calculated based on SVM kernel weights. The least important feature is removed and the process repeated. To avoid the possible selection bias of RFE algorithm (reported by Ambroise and McLachlan [2002]), we used an resampling scheme for recursive feature elimination, according the algorithm presented in Kuhn and Johnson [2013].

**Results**

*Fractal and multifractal characteristics*

Figure 2 presents values of fractal dimension \( D_f \) determined for four land cover classes in the form of boxplots. More precisely, single boxplot displays three values: first quartile (bottom edge of the box), second quartile - median (line inside the box) and third quartile (top edge) [Frigge et al., 1989]. As illustrated in the Figure 2, values of the \( D_f \) are in the range 2.13 - 2.52 and for the agriculture (black), forest (green) and water (blue) land cover classes ranges of \( D_f \) values overlap. In the case of urban class, marked by red colour, we observe lower fractal dimension values and it seems that \( D_f \) could be used to discriminate this land cover class from three others. Anyway, we observe rather limited possibility to separate individual land cover classes using \( D_f \) exclusively.

The highest values are observed for water, then forest, agriculture, and the lowest for urban. As we may observe at Figure 2. water samples of WV-2 image are not smooth surfaces, but they consist of dense concentration of noisy pixels with glinting effect (described e.g. by Hedley et al. [2005]). This variability of pixels values is the reason why fractal approach detects complex structure of water. Following this approach, forest samples also present high irregularities in pixel values, what is caused by trees canopy cover complexity. The most regular shapes and artificial surfaces may be observed on urban samples, where the heterogeneity of pixels is not so dense as on the other types of samples. The deviations in
pixel values within tested types of land cover classes are strongly correlated with spatial resolution of the imagery. Lower spatial resolution would lead to strong homogenization of water and forest samples. Agriculture samples would behave similarly and result in less heterogeneous surfaces divided by lines of fields’ borders, while urban samples would be the most heterogeneous ones.

The results for the generalized dimensions $D_q$ as a function of $q$ are shown in Figure 3. The values of $D_q$ are calculated for four land cover types: agriculture (black), forest (green), urban (red), and water (blue) presented in Figure 1.

In the next step, we look at the degree of multifractality $\Delta$ as another quantitative parameter considered in the frame of this work and calculated with three different measures: SUM ($\Delta^{\text{SUM}}$), MAX($\Delta^{\text{MAX}}$) and BCD($\Delta^{\text{BCD}}$), according to Equation [4]. Figure 4 shows $\Delta$ values for SUM measure with division on four land cover classes. We observe the highest values of $\Delta^{\text{SUM}}$ for urban class reaching ten times higher values of degree of multifractality than forest and agriculture. The water samples with $\Delta^{\text{SUM}} = 0$ reveal no symptoms of multifractal character and can be considered as monofractal cases. In general, results show that urban and water classes can be easy to separate from two others by using parameter $\Delta^{\text{SUM}}$ only.
Figure 3 - Examples of generalized dimensions $D_q$ as a function of $q$ calculated with SUM (a, b), MAX(c, d) and BCD (e, f) measure, for cases presented in the Figure 1. Left column shows results for images from left column of Figure 1.
Figure 4 - The degree of multifractality $\Delta$ calculated for four land cover classes: agriculture (black), forest (green), urban (red) and water (blue) by using SUM measure.

Figure 5 presents the degree of multifractality calculated with MAX measure. We see that values of $\Delta_{\text{MAX}}$ parameter for urban class dominate over others. Very interesting are results for water class, which becomes as multifractal as agriculture and forest. This effect is strictly related with the use of measure MAX, which exposes maximum values of image digital numbers and increases the heterogeneity in a given class. Summarizing, by using $\Delta_{\text{MAX}}$ parameter it is possible to separate urban class only.

Figure 5 - The degree of multifractality $\Delta$ calculated for four land cover classes: agriculture (black), forest (green), urban (red) and water (blue) by using MAX measure.
Let us now look at results presented in Figure 6, which corresponds to values of parameter calculated with BCD measure. Again, urban class reveals the highest level of multifractality and can be separated from others. For water class the values overlap agriculture and forest cases, what makes these three classes difficult to separate. However, the concentration of $\Delta_{BCD}$ values for agriculture and forest suggests that these two land cover classes could be distinguished from each other. It is worth to underline that $\Delta_{BCD}$ is only multifractal parameter from all three studied, which allows to separate between agriculture and forest classes.

![Figure 6](image)

**Figure 6 -** The degree of multifractality $\Delta$ calculated for four land cover classes: agriculture (black), forest (green), urban (red) and water (blue) by using BCD measure.

**Classification accuracy**

The overall classification accuracies obtained for each classification approach were calculated based on the repeated cross-validation procedure. Table 1 and Figure 7 present for particular image descriptors sets the results achieved for the optimal set of SVMs parameters and the best of parameter sets tested for RF and C5.0 classifiers. When looking into overall accuracy values, support vector machines usually outperformed the approaches based on classification trees. Only for 3 of 16 tested sets of image descriptors C5.0 algorithm gave higher accuracies. At the same time the C5.0 results are the most stable, what is visible in overall accuracy standard deviation range.

Classification based just on four histogram characteristics (mean, variance, skewness and kurtosis) allowed us to achieve the overall accuracy level of 98.6% for SVMs, 98.2% for RF and 98.0% for C5.0 approaches. None of image descriptor sets consisted of features originated from single texture analysis technique was able to get a better result. Only the
grey-level co-occurrence matrix-based features set gave comparable levels of accuracy for classification trees algorithms (98.0 and 98.2%) and better one (99.4) for SVMs. Accuracies obtained for classification based on pure multifractal features were among the lowest ones.

Table 1 - Classification results - overall accuracy and its standard deviation based on repeated cross-validation

| Number of feature set | Image descriptors                                                                 | Classification results |
|-----------------------|-----------------------------------------------------------------------------------|------------------------|
|                       |                                                                                   | RF Accuracy [%]        |
|                       |                                                                                   | RF SD [%]              |
|                       |                                                                                   | C5.0 Accuracy [%]      |
|                       |                                                                                   | C5.0 SD [%]            |
|                       |                                                                                   | SVM Accuracy [%]       |
|                       |                                                                                   | SVM SD [%]             |
| 1                     | all classification features                                                       | 99.1                  |
|                       |                                                                                   | 0.6                   |
|                       |                                                                                   | 99.4                  |
|                       |                                                                                   | 0.2                   |
|                       |                                                                                   | 99.7                  |
|                       |                                                                                   | 0.4                   |
| 2                     | all textural classification features                                              | 98.7                  |
|                       |                                                                                   | 0.8                   |
|                       |                                                                                   | 98.6                  |
|                       |                                                                                   | 0.3                   |
|                       |                                                                                   | 99.3                  |
|                       |                                                                                   | 0.6                   |
| 3                     | gray-level co-occurrence matrix-based features                                    | 98.0                  |
|                       |                                                                                   | 1.0                   |
|                       |                                                                                   | 98.2                  |
|                       |                                                                                   | 0.3                   |
|                       |                                                                                   | 99.4                  |
|                       |                                                                                   | 0.6                   |
| 4                     | gray-level co-occurrence matrix-based features and histogram-based features       | 98.6                  |
|                       |                                                                                   | 0.8                   |
|                       |                                                                                   | 99.1                  |
|                       |                                                                                   | 0.3                   |
|                       |                                                                                   | 99.8                  |
|                       |                                                                                   | 0.3                   |
| 5                     | run length matrix-based features                                                   | 95.0                  |
|                       |                                                                                   | 1.5                   |
|                       |                                                                                   | 98.8                  |
|                       |                                                                                   | 0.2                   |
|                       |                                                                                   | 97.4                  |
|                       |                                                                                   | 1.0                   |
| 6                     | run length matrix-based features and histogram-based features                     | 98.7                  |
|                       |                                                                                   | 0.9                   |
|                       |                                                                                   | 98.8                  |
|                       |                                                                                   | 0.3                   |
|                       |                                                                                   | 99.5                  |
|                       |                                                                                   | 0.5                   |
| 7                     | absolute gradient-based features                                                  | 91.1                  |
|                       |                                                                                   | 1.6                   |
|                       |                                                                                   | 94.7                  |
|                       |                                                                                   | 0.5                   |
|                       |                                                                                   | 93.7                  |
|                       |                                                                                   | 1.6                   |
| 8                     | absolute gradient-based features and histogram-based features                     | 98.6                  |
|                       |                                                                                   | 0.8                   |
|                       |                                                                                   | 98.5                  |
|                       |                                                                                   | 0.4                   |
|                       |                                                                                   | 99.2                  |
|                       |                                                                                   | 0.6                   |
| 9                     | autoregressive model parameters                                                   | 96.7                  |
|                       |                                                                                   | 1.5                   |
|                       |                                                                                   | 98.0                  |
|                       |                                                                                   | 0.5                   |
|                       |                                                                                   | 95.5                  |
|                       |                                                                                   | 1.4                   |
| 10                    | autoregressive model parameters and histogram-based features                       | 99.5                  |
|                       |                                                                                   | 0.5                   |
|                       |                                                                                   | 99.4                  |
|                       |                                                                                   | 0.2                   |
|                       |                                                                                   | 99.6                  |
|                       |                                                                                   | 0.4                   |
| 11                    | parameters derived from wavelet analysis                                         | 97.8                  |
|                       |                                                                                   | 1.0                   |
|                       |                                                                                   | 98.0                  |
|                       |                                                                                   | 0.3                   |
|                       |                                                                                   | 98.4                  |
|                       |                                                                                   | 0.9                   |
| 12                    | parameters derived from wavelet analysis and histogram-based features             | 99.1                  |
|                       |                                                                                   | 0.7                   |
|                       |                                                                                   | 99.0                  |
|                       |                                                                                   | 0.4                   |
|                       |                                                                                   | 99.6                  |
|                       |                                                                                   | 0.5                   |
| 13                    | fractal dimension and histogram-based features                                    | 98.7                  |
|                       |                                                                                   | 0.7                   |
|                       |                                                                                   | 98.3                  |
|                       |                                                                                   | 0.5                   |
|                       |                                                                                   | 99.0                  |
|                       |                                                                                   | 0.6                   |
| 14                    | histogram-based features                                                          | 98.2                  |
|                       |                                                                                   | 0.9                   |
|                       |                                                                                   | 98.0                  |
|                       |                                                                                   | 0.2                   |
|                       |                                                                                   | 98.6                  |
|                       |                                                                                   | 0.6                   |
| 15                    | multifractal parameters                                                           | 95.6                  |
|                       |                                                                                   | 1.4                   |
|                       |                                                                                   | 95.2                  |
|                       |                                                                                   | 0.6                   |
|                       |                                                                                   | 96.6                  |
|                       |                                                                                   | 1.2                   |
| 16                    | multifractal parameters and histogram-based features                              | 99.1                  |
|                       |                                                                                   | 0.7                   |
|                       |                                                                                   | 99.0                  |
|                       |                                                                                   | 0.1                   |
|                       |                                                                                   | 99.6                  |
|                       |                                                                                   | 0.4                   |
Figure 7 - Classification results - overall accuracy and its standard deviation based on repeated cross-validation.
For each of textural features groups classification performance increased substantially when histogram-based features were added. The highest classification accuracy for RF and C5.0 algorithms (99.5% and 99.4% respectively) was achieved for the set of image descriptors consisted of histogram-based features and autoregressive model parameters. For SVMs this group of features gave the third result, but the accuracy was just 0.2% worse than the highest one (and in the standard deviation range of the best accuracy result). The best performance was noticed for the set of grey-level co-occurrence matrix-based features and histogram-based features, with overall accuracy of 99.8%, the highest one for all tested approaches. Also for C5.0 algorithm this set of classification features performed well. However, in case of RF its result was rather average.

When combined with histogram-based features multifractal parameters gave the second best result for Random Forest with overall accuracy of 99.1% and the third one for C5.0 and SVMs classifiers (99.0 and 99.6%, respectively). Very similar results were obtained for the set of parameters derived from wavelet analysis and histogram-based features. The results achieved for combination of histogram-based characteristics and fractal dimension were 98.7% (RF), 98.3% (C5.0) and 99.0 (SVMs). Surprisingly, the classification approach based on all classification features, though performed very well for every classifier, did not outperform the other approaches.

Deeper insight into classification performance of particular image descriptor sets is possible based on analysis of error matrices and producer and user accuracy parameters calculated for each land cover category. As SVMs performed the best from tested classifiers, we decided to perform such thorough analysis for their results. We decided also to analyze only the results obtained for the sets of classification features where textural descriptors were supplemented by histogram-based ones. Error matrices were saved for each of 50 classification realizations and averaged. Then producer and user accuracy values calculated for particular land cover classes (Tab. 2).

| Number of feature set (see Tab. 1) | 1 | 4 | 6 | 8 | 10 | 12 | 13 | 14 | 16 |
|-----------------------------------|---|---|---|---|----|----|----|----|----|
| agriculture                       | 98.38 | 98.73 | 98.96 | 99.02 | 98.84 | 99.81 | 98.67 | 98.52 | 99.63 |
| forest                            | 99.97 | 99.91 | 99.67 | 99.27 | 99.85 | 100.00 | 98.99 | 99.14 | 99.60 |
| urban                             | 99.80 | 99.96 | 99.22 | 99.14 | 99.30 | 98.70 | 99.55 | 99.21 | 99.59 |
| water                             | 99.97 | 100.00 | 99.69 | 99.19 | 100.00 | 99.63 | 98.71 | 97.56 | 99.59 |
| agriculture                       | 99.70 | 99.82 | 98.96 | 98.48 | 99.09 | 98.66 | 99.21 | 97.7 | 99.21 |
| forest                            | 99.94 | 99.97 | 99.73 | 99.27 | 99.94 | 99.36 | 98.90 | 98.5 | 99.63 |
| urban                             | 99.88 | 99.71 | 99.59 | 99.30 | 99.06 | 99.96 | 99.14 | 98.7 | 99.75 |
| water                             | 99.18 | 99.52 | 99.32 | 99.35 | 99.97 | 99.93 | 98.84 | 99.2 | 99.66 |

Interestingly, classification based on multifractal features is the only one which allowed to achieve the producer accuracy value well over 99% for every land cover category. However, for any category did not give the best result. Usually, some agriculture samples were wrongly classified as both urban and water. In case of multifractals, they were
misclassified only to urban class. The highest producer accuracy values for agriculture and forest were obtained using wavelet-based features. For urban and water classes, grey-level co-occurrence matrix-based features performed the best. Multifractals gave also user accuracy values over 99% for every category. Again, there were not the highest values at any case.

**Feature selection**

Feature selection experiment resulted with reduced sets of image descriptors presented in Table 3. When features were selected from full set of available predictors (including both textural and histogram-based characteristics), reduced sets consisted of four or five features only. Mean pixel brightness was the only feature chosen irrespective of classifier. In every case autoregressive model parameter was included into the set. However, for RF and SVMs classifications the variance of minimized prediction error ($\sigma$) was selected, while $\theta_2$ parameter (the weight associated with top neighbour) was chosen by C5.0 algorithm. The degree of multifractality calculated by using BCD measure were chosen both in RF as well as C5.0 approaches. C5.0 winnowing procedure added another multifractal parameter - degree of multifractality calculated by using SUM measure. The multifractal parameter was also included into reduced set of features selected with SVMs. However, it was neither $\Delta_{BCD}$ nor $\Delta_{SUM}$, but $\Delta_{MAX}$. In case of RF the fourth parameter was derived from wavelet analysis. For SVMs, the fourth and the fifth features were grey-level co-occurrence matrix-based image descriptors.

| Image descriptors | Selected features |
|-------------------|-------------------|
| **RF**            | **C5.0**          | **SVM**          |
| all classification features | $\Delta_{BCD}$, $\sigma$, mean brightness, WavEnHH_s.3 | mean brightness, $\theta_2$, $\Delta_{BCD}$, $\Delta_{SUM}$ | mean brightness, $\sigma$, $\Delta_{MAX}$, S(1, 0) Correlation, S(1, 1) Correlation |
| all textural classification features | $\Delta_{BCD}$, skewness of absolute gradient, $\Delta_{MAX}$, $\sigma$, WavEnHH_s.3, WavEnLH_s.4 | $\Delta_{SUM}$, $\Delta_{BCD}$, $\Delta_{MAX}$, S(2, -2) Correlation, skewness of absolute gradient, $\theta_2$, $\sigma$, WavEnLL_s.1 | $\sigma$, $\Delta_{MAX}$, S(1, 0) Correlation, S(1, 1) Correlation, WavEnHL_s.5 |

In reduced sets of classification features derived for RF and C5.0 algorithms solely from textural characteristics mean brightness was replaced by skewness of absolute gradient and degree of multifractality calculated by using MAX measure. In the case of RF the set was completed with another wavelet analysis-based parameter. In C5.0 approach also $\sigma$ and one of grey-level co-occurrence matrix features were added. For SVMs the only difference was the replacement of mean brightness by the wavelet analysis-based parameter.
In most cases the overall accuracy of classification based on reduced sets of features was on the same level as from non-reduced ones (Tab. 4). In case of C5.0 the accuracy of classification based on sets of four features selected from all available predictors equals to the best results, however the standard deviation of the value is a little bigger. For RF the overall accuracy ranks this approach as the third one and for SVMs as fourth (although just 0.3% from the best one, in the range of standard deviation value).

Table 4 - Comparison of overall classification accuracy for reduced and non-reduced datasets (p stands for number of predictors; for SVMs the optimal set of parameters is provided).

| Image descriptors                      | Classification accuracy [%] |
|----------------------------------------|----------------------------|
|                                        | RF                         | C5.0                       | SVM                        |
|                                         | full set | reduced set | full set | reduced set | full set | reduced set |
| all classification features             | Acc [%] | SD [%]      | Acc [%] | SD [%]      | Acc [%] | SD [%]      |
|                                        | 98.9    | 0.7         | 99.0    | 0.8         | 99.3    | 0.2         |
|                                        | 99.4    | 0.4         | 99.7    | 0.4         | 99.5    | 0.6         |
| all textural classification features    | Acc [%] | SD [%]      | Acc [%] | SD [%]      | Acc [%] | SD [%]      |
|                                        | 98.7    | 0.8         | 97.7    | 1.0         | 98.6    | 0.3         |
|                                        | 98.5    | 0.3         | 98.5    | 0.3         | 99.3    | 0.6         |
|                                        | 93.0    | 1.5         | 93.0    | 1.5         | 93.0    | 1.5         |
| multifractal parameters and histogram-based features | Acc [%] | SD [%]      | Acc [%] | SD [%]      | Acc [%] | SD [%]      |
|                                        | 99.1    | 0.7         | 99.2    | 0.1         | 99.6    | 0.4         |

Discussion and conclusions

The aim of this study was to evaluate the applicability of multifractal measures as the global descriptors of panchromatic VHR satellite images. In particular we researched if this kind of textural features may be used as efficient descriptor of WorldView-2 image content in the context of dominating land cover types. For this purpose we compared the values of fractal and multifractal characteristics calculated for image chips cut from panchromatic WorldView-2 scenes. We also checked if it is possible to adequately classify these image chips into categories related to the land cover type prevailing inside the imaged area using these characteristics and/or several kinds of other global image descriptors.

The obtained results have shown that the degree of multifractality values differ depending on the land cover type. Regardless the measure used during the process of multifractal parameter calculation, the highest multifractality level is observed for the urban area. When SUM measure is used the level of multifractality calculated for water samples is close to 0 and images of water can be treated as monofractals.

On the other hand fractal analysis showed that urban cover has the lowest fractal dimension while water presents high values of $D_F$. This can be related to the limitation of fractal...
dimension used in characterization of a set’s texture [Myint, 2003] but also to the fact that fractal analysis compared with multifractals seems to treat texture complexity in different way. More precisely, during the $D_f$ determination the most important values are the minimum and maximum pixel values considered in a given box and used BCD method seems to concentrate on dynamical changes. In multifractal analysis, during the process of measure construction and its scaling all values of pixels in a given box are taken into account. It means that not only a span of pixels dynamic is considered but also their spatial relations [Wawrzaszek et al., 2014]. Degree of multifractality values calculated using different measures differs from each other. These differences influences their potential usefulness as classification features. In the case of MAX measure only the urban class may be separated. The same is true if the class discrimination is based on fractal dimension. Urban and water classes may be easily recognized based on the degree of multifractality calculated using SUM measure, while BCD measure enables differentiation of agriculture and forest images, provided the water samples are separated before.

Performed classification tests confirmed that for panchromatic VHR WorldView-2 satellite images multifractal parameters should be considered as valuable textural features. However, for efficient implementation these features should be accompanied by simple histogram-based characteristics. Our experiment has shown that the last statement is true also for other types of textural features.

In our classification tests multifractal parameters used without supplemental histogram-based features performed poorly and were beaten by most of analysed groups of textural characteristics. However, together with simple image statistics, they gave the second or third best result depending on classifier used. It is worth noting that such approach was the only one able to give the producer accuracies values over 99% for every land cover category. In every feature selection approaches very limited and very efficient sets of image descriptors were determined. The initial set of 294 features was reduced to 4 or 5 characteristics without increasing the classification error rate. Experiments with feature selection shown, that regardless the classifier multifractal parameters were present in reduced sets of predictors. It means they may be considered as one of the most informative classification features derived for analysed images. Combined with the mean brightness and up to three other features selected from first-order autoregressive model parameters and wavelet or grey-level co-occurrence matrix-based features (depending on classification algorithm used) they constitute extremely efficient set of image descriptors.

The highest of all obtained classification accuracies was achieved with SVMs approach based on grey-level co-occurrence matrix-based features combined with histogram-based characteristics. However, when analysing the performance of particular groups of textural image descriptor one has to notice the results achieved based on first-order autoregressive model parameters. Combined with histogram-based features they gave the best results for RF and C5.0 classifiers and performed very well in case of SVMs, as well. Approaches of this kind are generally considered as successful in micro-texture analysis, but not very useful for classification of remote sensing images [Coburn and Roberts, 2004]. Results of the present study may be considered as a stimulus for further investigations of its usefulness for characterization of texture of VHR remote sensing images.

The research presented in this paper have shown the usefulness of multifractal features as
textural characteristics of VHR WorldView-2 panchromatic satellite images. The novelty of presented study consists in the application of different measures constructed in the frame of multifractal analysis. In our opinion the opportunity to define various capacities is an important advantage of presented approach to implementation of multifractal formalism in image texture analysis. We can conclude that multifractal description may be considered as a promising tool for analysis of remotely sensed images. It seems to be especially important when applied to panchromatic images for purpose of image classification as well as in content-based image retrieval (CBIR) or image information mining (IIM).

The applicability of multifractal features should be proven during further research extended for images of different areas and other VHR remote sensing sensors, including digital and analog (archive) aerial photos. Change detection on panchromatic remote sensing images seems to be a promising field of application [Aleksandrowicz et al., 2016]. The study could be also continued to assess the sensitivity of proposed parameters to eg. image patch size or class definition.

Utilisation of multifractal features to derive local spatial information from the image may be considered as another direction of future research. Aleksandrowicz et al. [2016] used Hölder exponents to characterise texture at pixel neighborhood level. Incorporation of such textural (spatial) features into classification of multispectral and hiperspectral images constitutes one of actual research fields in remote sensing [Wang et al., 2016].

Acknowledgments
Wojciech Drzewiecki and Katarzyna Bernat acknowledge the support by AGH University grant No. 11.11.150.949. Anna Wawrzaszek, Michał Krupiński and Sebastian Aleksandrowicz acknowledge the support by the Polish National Science Centre (grant NN 526 1568 40).

References
Aleksandrowicz S., Wawrzaszek A., Drzewiecki W., Krupiński M. (2016) - Change Detection Using Global and Local Multifractal Description. IEEE Geoscience and Remote Sensing Letters, 13 (8): 1183-1187. doi: http://dx.doi.org/10.1109/LGRS.2016.2574940.

Ambroise C., McLachlan G. (2002) - Selection Bias in Gene Extraction on the Basis of Microarray Gene-Expression Data. Proceedings of the National Academy of Sciences of the United States of America, 99 (10): 6562-6566. doi: http://dx.doi.org/10.1073/pnas.102102699.

Blaschke T., Burnett C., Pekkarinen A. (2004) - Image segmentation methods for object-based analysis and classification. In: Remote sensing image analysis: Including the spatial domain. De Jong S.M., Van der Meer F.D. (Eds.), Springer, pp. 211-236. doi: http://dx.doi.org/10.1007/978-1-4020-2560-0_12.

Boser B., Guyon I., Vapnik V. (1992) - A Training Algorithm for Optimal Margin Classifiers. Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, pp. 144-152. doi: http://dx.doi.org/10.1145/130385.130401.

Breiman L. (2001) - Random Forests. Machine Learning. Proceedings of the fifth annual workshop on Computational learning theory, 45: 5-32.

Cerra D., Datcu M. (2010) - A multiresolution approach for texture classification in high
resolution satellite imagery. Italian Journal of Remote Sensing, 42 (1): 13-24. doi: http://dx.doi.org/10.5721/ItJRS20104212.

Chakraborty B., Haris K., Latha G., Maslov N., Menezes A. (2014) - Multifractal Approach for Seafloor Characterization. IEEE Geoscience and Remote Sensing Letters, 11 (1): 54-58. doi: http://dx.doi.org/10.1109/LGRS.2013.2245856.

Chehata N., Guo L., Mallet C. (2009) - Airborne lidar feature selection for urban classification using random forests. In: Laser scanning, Bretar F., Pierrot-Deseilligny M., Vosselman G. (Eds.), International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XXXVIII, pp. 207-212. Available online at: http://www.isprs.org/proceedings/XXXVIII/3-W8/papers/207_laserscanning09.pdf

Cheng Q. (1999) - Multifractality and spatial statistics. Computer and Geosciences, 25: 949-961. doi: http://dx.doi.org/10.1016/S0098-3004(99)00060-6.

Coburn C.A., Roberts A.C.B. (2004) - A multiscale texture analysis procedure for improved forest stand classification. International Journal of Remote Sensing, 25 (20): 428-4308. doi: http://dx.doi.org/10.1080/0143116042000192367.

Datta R., Joshi D., Li, J., Wang J.Z. (2008) - Image Retrieval: Ideas, Influences, and Trends of the New Age. ACM Computing Surveys, 40 (5): 2-60. doi: http://dx.doi.org/10.1145/1348246.1348248.

De Souza P. (1982) - Texture Recognition via Autoregression. Pattern Recognition, 15: 471-475. doi: http://dx.doi.org/10.1016/0031-3203(82)90025-5.

Díaz-Uriarte R., Alvarez de Andrés S. (2006) - Gene selection and classification of microarray data using random forest. BMC Bioinformatics, 7 (3): 1-13. doi: http://dx.doi.org/10.1186/1471-2105-7-3.

Drzewiecki W., Wawrzaszek A., Aleksandrowicz S., Krupiński M., Bernat K. (2013) - Comparison of selected textural features as global content-based descriptors of VHR satellite image. IEEE International Geoscience and Remote Sensing Symposium - IGARSS, Melbourne, pp. 4364-4366. doi: http://dx.doi.org/10.1109/IGARSS.2013.6723801.

Du G., Yeo T.S. (2002) - A novel multifractal estimation method and its application to remote image segmentation. IEEE Transactions on Geoscience and Remote Sensing, 40 (4): 980-982. doi: http://dx.doi.org/10.1109/TGRS.2002.1006395.

Emerson C.W., Quattrochi D.A., Lam N.S.-N. (2004) - Spatial Metadata for Remote Sensing Imagery. In: NASA's Earth Science Technology Conference, Palo Alto.

Erner A., Düzgün H.S. (2009) - A methodology for land use change detection of high resolution pan images based on texture analysis. Italian Journal of Remote Sensing, 41 (2): 47-59. doi: http://dx.doi.org/10.5721/ItJRS20094124.

Fernández-Delgado M., Cernadas E., Barro S., Amorim D. (2014) - Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? Journal of Machine Learning Research, 15: 3133-3181. Available online at: http://jmlr.org/papers/volume15/delgado14a/delgado14a.pdf.

Freund Y., Schapire R.E. (1997) - A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55 (10): 119-139. doi: http://dx.doi.org/10.1006/jcss.1997.1504.

Frigge M., Hoaglin D.C., Iglewicz B. (1989) - Some implementations of the boxplot. The American Statistician, 43 (1): 50-54. doi: http://dx.doi.org/10.2307/2685173.
Galloway M.M. (1975) - *Texture Analysis Using Gray Level Run Lengths*. Computer Graphics and Image Processing, 4: 172-179. doi: http://dx.doi.org/10.1016/S0146-664X(75)80008-6.

Grassberger P. (1983) - *Generalized dimensions of strange attractors*. Physic Letters A, 97 (6): 227-230. doi: http://dx.doi.org/10.1016/0375-9601(83)90753-3.

Guan H., Yu J., Li J., Luo L. (2012) - *Random Forests-Based Feature Selection for Land-Use Classification Using LIDAR Data and Orthoimagery*. International Archives of the Photogrammetry. Remote Sensing and Spatial Information Sciences, 39: B7. doi: http://dx.doi.org/10.5194/isprsarchives-XXXIX-B7-203-2012.

Guyon I., Weston J., Barnhill S., Vapnik V. (2002) - *Gene Selection for Cancer Classification Using Support Vector Machines*. Machine Learning, 46 (1): 389-422. doi: http://dx.doi.org/10.1023/A:1012487302797.

Hajek M., Dezortova M., Materka A., Lerski R. (2006) - *Texture Analysis for Magnetic Resonance Imaging*. Med4publishing, Prague.

Halsey T.C., Jensen M.H., Kadanoff L.P., Procaccia I., Shraiman B.I. (1986) - *Fractal measures and their singularities: The characterization of strange sets*. Physical Review A, 33 (2): 1141-1151. doi: http://dx.doi.org/10.1016/0920-5632(87)90036-3.

Hapfelmeier A., Ulm K. (2013) - *A new variable selection approach using Random Forest*. Computational Statistics and Data Analysis, 60: 50-69. doi: http://dx.doi.org/10.1016/j.csda.2012.09.020.

Haralick R. (1979) - *Statistical and Structural Approaches to Texture*. IEEE Proceedings, 67 (5): 786-804. doi: http://dx.doi.org/10.1109/PROC.1979.11328.

Haralick R.M., Sternberg S.R., Zhuang X. (1987) - *Image Analysis using Mathematical Morphology*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 4: 532-550. doi: http://dx.doi.org/10.1109/TPAMI.1987.4767941.

Hedley J., Harborne A., Mumby P. (2005) - *Simple and Robust Removal of Sun Glint for Mapping Shallow-Water Benthos*. International Journal of Remote Sensing, 26: 2107-2112. doi: http://dx.doi.org/10.1080/0143116050034086.

Hentschel H., Procaccia I. (1983) - *The infinite number of generalized dimensions of fractals and strange attractor*. Physica: Nonlinear Phenomena, 8 (3): 435-444. doi: http://dx.doi.org/10.1016/0167-2789(83)90235-X.

Howarth P., Ruger S. (2007) - *Evaluation of Texture Features for Content-Based Image Retrieval*. Lecture Notes in Computer Science, 3115: 326-334. doi: http://dx.doi.org/10.1007/978-3-540-27814-6_40.

Jain A. (1989) - *Fundamentals of Digital Image Processing*. Prentice Hall International.

Kuhn M., Johnson, K. (2013) - *Applied Predictive Modeling*. Springer, New York. doi: http://dx.doi.org/10.1007/978-1-4614-6849-3.

Lam N.S.-N., De Cola L. (1993) - *Fractals in Geography*. Prentice Hall: Englewood Cliffs.

Li M., Zang S., Zhang B., Li S., Wu C. (2014) - *A Review of Remote Sensing Image Classification Techniques: the Role of Spatio-contextual Information*. European Journal of Remote Sensing, 47: 389-411. doi: http://dx.doi.org/10.5721/EuJRS20144723.

Macek W.M., Wawrzaszek A., Carbone V. (2012) - *Observation of the multifractal spectrum in the heliosphere and the heliosheath by Voyager 1 and 2*. Journal of Geophysical Research: Space Physics, 117: A12. doi: http://dx.doi.org/10.1029/2012JA018129.

Mallat S.G. (1989) - *Multifrequency Channel Decompositions of Images and Wavelet
Models. IEEE Transaction on Acoustics, Speech and Signal Processing, 37: 2091-2110. doi: http://dx.doi.org/10.1109/29.45554.
Mandelbrot B.B. (1983) - The fractal geometry of nature. Freeman and Co., New York.
Meneveau C., Sreenivasan K.R. (1991) - The multifractal nature of the turbulent energy dissipation. Journal of Fluid Mechanics, 224: 429-484. doi: http://dx.doi.org/10.1017/S0022112091001830.
Mountrakis G., Im J., Ogole C. (2011) - Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 66: 247-259. doi: http://dx.doi.org/10.1016/j.isprs jprs.2010.11.001.
Myint S.W. (2003) - Fractal Approaches in texture analysis and classification of remotely sensed data: Comparisons with spatial autocorrelation techniques and simple descriptive statistics. International Journal of Remote Sensing, 24 (9): 1925-1947. doi: http://dx.doi.org/10.1080/01431160210155992.
Ott E. (1993) - Chaos in Dynamical Systems. Cambridge University Press, Cambridge.
Qian Y., Zhou W., Yan J., Li W., Han L. (2015) - Comparing Machine Learning Classifiers for Object-Based Land Cover Classification Using Very High Resolution Imagery. Remote Sensing, 7: 153-168. doi: http://dx.doi.org/10.3390/rs70100153.
Quinlan J.R. (1993) - C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., San Francisco.
Samal A., Bhatia S., Vadlamani P., Marx D. (2009) - Searching satellite imagery with integrated measures. Pattern Recognition, 42: 2502-2513. doi: http://dx.doi.org/10.1016/j.patcog.2009.01.035.
Sandri M., Zuccolotto P. (2006) - Variable selection using random forests. In: Data Analysis, Classification and the Forward Search, Zani S., Cerioli A., Riani M., Vichi M. (Eds.), Springer, Berlin Heidelberg, pp. 263-270. doi: http://dx.doi.org/10.1016/j.patrec.2010.03.014.
Sarkar N., Chaudhuri B.B. (1994) - An efficient differential box-counting approach to compute fractal dimension of image. IEEE Transactions on Systems, Man and Cybernetics, 24: 115-120. doi: http://dx.doi.org/10.1109/21.259692.
Schindler K. (2012) - An Overview and Comparison of Smooth Labeling Methods for Land-Cover Classification. IEEE Transactions on Geoscience and Remote Sensing, 50 (11): 4534-4545. doi: http://dx.doi.org/10.1109/TGRS.2012.2192741.
Shyu C.R., Brodley C.E., Kak A.C., Kosaka A., Aisen A., Broderick L. (1998) - Local versus global features for content-based image retrieval. Proceedings IEEE Workshop on Content-Based Access of Image and Video Libraries, pp. 30-34. doi: http://dx.doi.org/10.1109/IVL.1998.694482.
Smeulders A.W.M., Worrting M., Santini S., Gupta A., Jain R. (2000) - Content-based Image Retrieval at the End of the Early Years. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (12): 1349-1318. doi: http://dx.doi.org/10.1109/34.895972.
Stojić T., Reljin I., Reljijn B. (2006) - Adaptation of multifractal analysis to segmentation of microcalcifications in digital mammograms. Physica A: Statistical Mechanics and its Applications, 367: 494-508. doi: http://dx.doi.org/10.1016/j.physa.2005.11.030.
Sun W., Xu G., Gong P., Liang S. (2006) - Fractal analysis of remotely sensed images: A review of methods and applications. International Journal of Remote Sensing, 27 (22): 4963-4990. doi: http://dx.doi.org/10.1080/01431160600676695.
Svetnik V., Liaw A., Tong C., Wang T. (2004) - Application of Breiman’s random forest to modeling structure-activity relationships of pharmaceutical molecules. In: Multiple Classifier Systems. Lecture Notes in Computer Science, Roli F., Kittler J., Windeatt T. Springer, 3077: 334-343. doi: http://dx.doi.org/10.1007/978-3-540-25966-4_33.

Szczypiński P.M., Strzelecki M., Materka A., Klepaczko A. (2009) - MaZda - A Software for image texture analysis. Computer Methods and Programs in Biomedicine, 94: 66-76. doi: http://dx.doi.org/10.1109/ISITC.2007.15.

Tso B., Mather P.M. (2009) - Classification Methods for Remotely Sensed Data. CRC Press.

Ustuner M., Balik F., Dixon B. (2015) - Application of Support Vector Machines for Landuse Classification Using High-Resolution RapidEye Images: A Sensitivity Analysis. European Journal of Remote Sensing, 48: 403-422. doi: http://dx.doi.org/10.5721/EuJRS20154823.

Vapnik V., Chervonenkis A. (1971) - On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability & Its Applications, 16: 264-280. doi: http://dx.doi.org/10.1137/1116025.

Vapnik V. (2010) - The Nature of Statistical Learning Theory. Springer.

Voorons M., Germain M., Benie G.B., Ko F. (2003) - Segmentation of high resolution images based on the multifractal analysis. In: IEEE Proceedings of International Symposium on Geoscience and Remote Sensing, 6: 3531-3533. doi: http://dx.doi.org/10.1109/IGARSS.2003.1294844.

Walker R.F., Jackway P.T., Longstaff I.D. (1997) - Recent Developments in the Use of the Co-occurrence Matrix for Texture Recognition. 13th International Conference on Digital Signal Processing Proceedings, 1: 63-65. doi: http://dx.doi.org/10.1109/ICDSP.1997.627968.

Wang L., Shi C., Diao C., Ji W., Yin D. (2016) - A survey of methods incorporating spatial information in image classification and spectral unmixing. International Journal of Remote Sensing, 37 (16): 3870-3910. doi: http://dx.doi.org/10.1080/01431161.2016.1204032.

Wang M., Chen X., Zhang H. (2010) - Maximal conditional chi-square importance in random forests. Bioinformatics, 26 (6): 831-837. doi: http://dx.doi.org/10.1093/bioinformatics/btq038.

Wawrzaszek A., Macek W.M. (2010) - Observation of the multifractal spectrum in solar wind turbulence by Ulysses at high latitudes. Journal of Geophysical Research, 115: A07104. doi: http://dx.doi.org/10.1029/2009JA015176.

Wawrzaszek A., Krupiński M., Aleksandrowicz S., Drzewiecki W. (2013) - Fractal and multifractal characteristics of very high resolution satellite images. IEEE International Symposium on Geoscience and Remote Sensing, pp. 1501-1504. doi: http://dx.doi.org/10.1109/IGARSS.2013.6723071.

Wawrzaszek A., Aleksandrowicz S., Krupiński M., Drzewiecki W. (2014) - Influence of image filtering on land cover classification when using fractal and multifractal features. PFG Photogrammetrie, Fernerkundung, Geoinformation, 2: 101-115. doi: http://dx.doi.org/10.1127/1432-8364/2014/0212.