Journal of Theoretical and Applied Information Technology
15th August 2018. Vol.96. No 15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SOFTWARE TOOLS FOR CELL WALLS SEGMENTATION IN MICROPHOTOGRAPHY

[1]GULZIRA BAKHITBEKOVNA ABDIKERIMOVA, [2]FEODOR ALEXANDROVICH MURZIN, [3]ALEKSEY LEONIDOVICH BYCHKOV, [4]JAMALBEK TUSSUPOV, [5]SERGEY KHAYRULIN, [6]WEI XINYU, [7]ELENA IVANOVNA RYBCHIKOVA

[1]Senior lecturer of the Department of Information technology, Faculty of information technologies, L.N. Gumilyov Eurasian National University, Republic of Kazakhstan

[2]Deputy director for research part at A.P. Ershov Institute of Informatics Systems, Siberian Branch of the Russian Academy of Sciences, Russia

[3]Researcher Institute of solid state chemistry and mechanochemistry of the Siberian Branch of the Russian Academy of Sciences, Russia

[4]Head of the Department of Information technology, Faculty of information technologies, L.N. Gumilyov Eurasian National University, Republic of Kazakhstan

[5]Researcher at A.P. Ershov Institute of Informatics Systems, Siberian Branch of the Russian Academy of Sciences, Russia

[6]Postgraduate student at Novosibirsk State University, Russia

[7]The head of the Microscopy Researches Group at Institute of Molecular Biology and Fundamental Medicine, Siberian Branch of the Russian Academy of Sciences, Russia

E-mail: [1]gulzira1981@mail.ru, [2]murzin@iis.nsk.su, [3]bychkov_a@solid.nsc.ru, [4]tussupov@mail.ru, [5]s.khayrulin@gmail.com, [6]xinyuwei2016@163.com, [7]lenryab@niboch.nsc.ru

ABSTRACT

This article provides an analysis of how textural areas are segmented on the microphotography in the image processing systems. A comparative description of the main methods is provided. The purpose of our research is to develop and find algorithms for analyzing the textural characteristic of images and the segmentation of textural features obtained using the transmission electron microscopy using software, established under Treaty No. 16-13-10200/2. Work has been done for the Institute of solid state chemistry and mechanochemistry of the Siberian branch of the Russian Academy of Sciences. The focus is on the development of software tools of choice and the creation of features that describe texture differences; of selection and segmentation of the textural areas; of classifying textural areas and identifying the texture object to analyze the above-named microphotographs. The question of the applicability of textural sets to analyse experimental data in order to identify in the microphotography the patterns that could be linked in the future to porousness, chemical reactiveness, etc. Suitable algorithms have been selected and appropriate software tools have been established at Matlab and in the software package for statistical analysis of Statistica.

Key words: *Textural Features, K-Means Method, Agglomerative Method, Image Processing, R/S-Analysis, Microphotography Analysis, Electron Microscopy, Herbal Raw Material.*

1. INTRODUCTION

The basis for this scientific research is the treaty between the Institute of solid state chemistry and mechanochemistry of the Siberian branch of the Russian Academy of Sciences (ISSC SB RAS) and the A.P. Ershov Institute of Informatics Systems, Siberian Branch of the Russian Academy of Sciences (IIS SB RAS) No. 16-13-10200/2 from July 1, 2016. The work is carried out in order to performance the ISSC SB RAS of Agreement No. 16-13-10200 dated May 12, 2016 between the Russian Science Foundation, the project manager and the Organization to provide a grant for fundamental scientific research and exploratory research studies relating to the grant of a scientific project entitled "Controlled change in the structure and composition of herbal raw materials with the mechanochemical methods for intensification of extraction of biologically active compounds" No. 16-13-10200».

This work focuses on developing software tools for image analysis, such as textural analysis, textural feature segmentation, and R/S analysis programmes. The source is the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

microphotography of vegetable raw materials, which is shredded in special mills. The final objective of the work is to determine on the microphotography the porousness, chemical reactivity of raw materials, etc. This task in the work has not been fully solved, and this requires more in-depth studies to be carried out in the future. To do this, textural feature sets are used.

The following methods are used: Image and signal processing, programming, design and analysis of algorithms and programs, development of human-machine interfaces

2. METHODS OF TEXTURAL FEATURES ANALYSIS

Despite the ubiquitous presence of textures in images, a uniform and formal approach to describing the texture and its strict definition does not exist at this time. Texture analysis methods are typically designed for each individual case.

In [1], the texture is understood as "spatial organization of elements within a borders of some surface". It is also explained that this organization is due to a certain statistical distribution of the intensity of gray tones or tones of different colours. An area can be considered textural if the number of variations in intensity or color changes it contains is large enough. In [2] a texture is called "in someway organized surface area". In [3] a texture is defined as a matrix or a fragment of the spatial properties of the image areas with homogeneous statistical characteristics.

Textures can be divided into several classes, as follows:

1. by origin: Artificial- for example, graphic patterns and natural- such as grass, forest, ground;

2. by surface structure: structural of geometrically correct elements, and stochastic, formed by a sequence of random elements; of relative size of texture elements: Fine-grained and coarse-grained;

3. by the form of texture elements: wavy, spotted, incorrect, linear, and so forth [1].

From the above definitions and characteristics, it follows that a texture is a piece of an image, one that has homogeneous statistical characteristics. This means that each texture of this class can be described by a characteristic property common to all textures of this class [1]. These properties are called textural features. Textural features play an important role in dividing an image into separate areas. Consider a more detailed system of features.

In this task, gray (halftone) images are examined. Therefore, the image is set as a matrix whose values are the brightness of the pixels in the range of 0 to 255.

The standard approach for calculating textural features is: It's necessary to select a so-called scrolling window from odd-numbered side : 3, 5, 7 pixels. The feature is calculated inside the scrolling window. The size of the local fragment is a carrier of the texture property. The feature value is written to the new matrix of the same size as the original. In the new matrix, the value is written to the coordinate, which is equal to the center of the scrolling window coordinate. Elements of the new matrix are obtained in some interval [A, B]. Next, typically, this interval is linearly displayed in the leg [0.255]. Then you can visualize the result of the textural feature calculation.

The experiments have shown that the standard approach in our case is not informative. It was therefore decided to use a *non-standard approach*. Specifically, the textural windows are calculated by large windows (including, by nonsquare), that the user can specify by selecting an area that may be of interest to him/her. That is about calculating numerical characteristics that pertain to vast areas that include different kinds of artifacts.

Considered 18 of the most important textural features, all of which are implemented in the program.

3. BASIC TEXTURAL FEATURES

3.1. Features that take into account the location of each other

The adjacency matrix approach (another name is the gradient distribution matrix [4]) is used to generate textural features that take into account the mutual disposition of pixels within the scrolling window. In the future, we will use the name of the adjacency matrix.

Let the analyzed image be a rectangle and have N_x horizontal and N_y vertical elements. Whereby $G = \{1,2,...,N\}$ is a set of N quantized brightness values. The image is then described by the function of the brightness values from the set of G, that is $f : L_x \times L_y \to G$, where $L_x = \{1,2,...,N_x\}$ and $L_y = \{1,2,...,N_y\}$ are horizontal and vertical spatial areas, respectively. A set N_x and N_y is a set of elements of resolution in a bitmap image. The adjacency matrix contains the relative frequencies p_{ij} of the adjacent elements in

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the image, located at a distance of d from each other, with actual brightness values $i, j \in G$. Typically, horizontal $(\alpha = 0^\circ)$, vertical $(\alpha = 90^\circ)$, and diagonal ($\alpha = 45^\circ$ and $\alpha = 135^\circ$) pairs of elements are recognized. It should be noted that these matrices are symmetrical, namely $P(i, j, d, \alpha) = P(j, i, d, \alpha)$.

Based on calculated adjacency matrices, it is possible to calculate direct numerical estimates of a number of textural features [4,5]:

Average

$$T_1 = \mu_i = \mu_j = \sum_{i=0}^{N-1}\left[i \sum_{j=0}^{N-1} P(i,j) \right];$$

Energy

$$T_2 = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}[P(i,j)]^2;$$

Variation

$$T_3 = \sigma_i^2 = \sum_{i=0}^{N-1}\left[(i-\mu_2)^2 \sum_{j=0}^{N-1} P(i,j) \right];$$

Homogeneity

$$T_4 = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P(i,j)/(1+|i-j|),$$

where P (i, j) is the frequency of occurrence of two pixels in a scrolling window with a brightness of i and j at an angle α and distance d.

σ_i - mean-square deviation of the brightness in the scrolling window.

The statistical moments allow you to create textural features, that take into account the reciprocal location of neighboring pixels in a scrolling window, and are therefore effective for describing textures with a spatial frequency.

The second angular point $T_5 = \sum_{i=1}^{N}\sum_{j=1}^{N}(P(i,j)/M)^2$, where M is the total number of pairs of elements adjoining each other (for example, $d=1, \alpha=0, M = 2N_y(N_x - 1)$), is a measure of similarity (homogeneity) of the image and takes a minimum value in this case.

Contrast

$$T_6 = \sum_{n=0}^{N-1} n^2 \left[\sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j)/M \right], \qquad |i-j| = n,$$

is determined by the amount of local variations in the brightness of the image. With the increase in the number of local variation, the contrast increases.

Correlation coefficient

$$T_7 = \sigma_x^{-1}\sigma_y^{-1} = \sum_{i=1}^{N}\sum_{j=1}^{N}[ij(P(i,j)/M) - m_x m_y],$$

where m_x, m_y, σ_x, σ_y is the average value and mean-square deviations for $p_x(i) = \sum_{j=1}^{N} P(i,j)/M$ and $p_y(j) = \sum_{i=1}^{N} P(i,j)/M$ respectively. T_7 provides a measure of the linearity of the brightness regressional dependence on the image.

Dispersion $T_8 = \sum_{i=1}^{N}\sum_{j=1}^{N}(i-m)^2(P(i,j)/M)$ determines the variation in brightness relative to the mean value.

The reciprocal difference moment $T_9 = \sum_{i=1}^{N}\sum_{j=1}^{N}[1+(i-j)^2]^{-1}(P(i,j)/M)$ is closely related to the contrast and reflects the degree of scattering of the gradient matrix elements around the main diagonal. This feature is an alternative to contrast in the case of edge structures influence, because the relatively large differences in brightness values make the smallest contribution to the final result.

Total average

$T_{10} = \sum_{n=2}^{2N} n p_+(n)$ where $p_+(n) = \sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j)/M$ for $i + j = n$, $n = 2,3, \ldots, 2N$ is a histogram of the sum of brightness values. T_{10} is defined by a histogram of the sum of the brightness values $p_+(n)$ for pairs of image elements that are directly related to the adjacency matrix.

The total dispersion $T_{11} = \sum_{n=2}^{2N}(n - T_{14})^2 p_+(n)$ is a measure of the brightness variation relative to the total average.

The total entropy for a histogram of the sum of brightness values $T_{12} = \sum_{n=2}^{2N} p_+(n)\log p_+(n)$ is determined by the classical measure of the statistical theory of information and expresses the uneven distribution of the brightness properties of the image elements.

Entropy

$T_{13} = -\sum_{i=1}^{N}\sum_{j=1}^{N}(P(i,j)/M)\log(P(i,j)/M)$ is defined in the same way as the total entropy, but only for the adjacency matrix.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.2 Features based on the description of structural elements

In recent times, there has been an increase in the development of structural approach to texture descriptions based on an analysis of the shape and size of the elements that make up the texture, the computation of local features, and an analysis of the distribution of texture elements in the image field. [1] presents features based on the length of the series. The length of a texture series is the number of elements in a raster line that have constant brightness. Let $p(i,j)$ indicate the number of lines whose length is equal j and which is oriented in the direction d. These lines are made up of image points whose intensity levels lie in the interval i. The following characteristics can be highlighted.

Reciprocal moments

$$T_{14} = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{p(i,j)}{j^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)};$$

Momenta

$$T_{15} = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}j^2\,p(i,j)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)};$$

Brightness differences

$$T_{16} = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}(p(i,j))^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)};$$

Series length heterogeneity

$$T_{17} = \frac{\sum_{i=1}^{N_g}(\sum_{j=1}^{N_r}(p(i,j)))^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)};$$

Portion of the image in series

$$T_{18} = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}jp(i,j)}.$$

4. CLUSTERING ALGORITHMS

4.1 Hierarchical Cluster Analysis

Image segmentation can be implemented in a local and global way. The local method evaluates whether there is a boundary between the areas by the features behave in the neighborhood of the image point. The global method presupposes the provisional clustering of the features space, and then establishes a match between the image pixel and the cluster to which its vector of features falls [16].

One of the heuristic methods of clustering is method based on a sequential agglomerative procedure. The advantages of these methods is a simplicity of computational procedure and algorithms.

Let the analyzed population consist of p elements, each of which is characterized by n variable values.

In the first step of the iterative procedure, there are p clusters, each of which includes one element. Two of the closest or similar clusters are defined, combined into a single cluster, and the number of clusters is reduced by 1 ($p \to p-1$).

The proximity measure is determined by the distance between the elements in the symmetrical matrix of distances D:

$$D = \begin{pmatrix} 0 & d_{12} & \cdots & d_{1p} \\ d_{21} & 0 & \cdots & d_{2p} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ d_{p1} & d_{p2} & \cdots & 0 \end{pmatrix},$$

The closest are the objects with *the smallest distance*.

The distance between the points between the cluster centers are defined by different metrics, so the distance between the l and the r points in the Euclidean metric is equal to:

$$d_{rj}^{(2)} = \sqrt{\sum_{i=1}^{n}(x_{ri} - x_{li})^2}.$$

At each subsequent step of the agglomerative procedure, you will need to recalculate only one line (and one column) D, meaning the distances from the formed cluster to each of the remaining clusters. There are several methods for recalculating distances using the old distance values for the combined clusters that differ by the coefficients in the formula

$$d_{rs} = \alpha_p d_{ps} + \alpha_q d_{qs} + \beta d_{pq} + \gamma\left|d_{ps} - d_{qs}\right|$$

If the p and q clusters are merged into the r cluster and you need to calculate the distance

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

from the new cluster to cluster *s*, the application of certain method depends on how the cluster distances are determined, different methods differ in coefficient values

$$\alpha_p, \alpha_q, \beta \; u \; \gamma.$$

The hierarchical cluster analysis procedure provides the grouping of both objects (data matrix lines) and variables (columns). It can be assumed that in the latter case, the role of the objects is played by variables and the role of variable by columns.

4.2 K-means method

The k -means method is a cluster analysis method, the purpose of which is to separate m the observations (from the space R^n) into k clusters, and each observation belongs to the cluster to the center (the centroid) of which it's the closest.

The Euclidean distance is used as a proximity measure:

$$\rho(x,y) = \|x-y\| = \sqrt{\sum_{\rho=1}^{n}(x_\rho - y_\rho)^2}$$, where

$$x, y \in R^n$$

So, looking at a series of observations

$$(x^{(1)}, x^{(2)}, ..., x^{(m)}), \; x^{(j)} \in R^n.$$

The k - means method divides m observations into k groups (or clusters) $(k \le m)$ $S = \{S_1, S_2, ..., S_k\}$ to minimize the total quadratic deviation of cluster points from the centroids of these clusters:

$$\min\left[\sum_{i=1}^{k}\sum_{x^{(j)} \in S_i}\left\|x^{(j)} - \mu_i\right\|^2\right]$$, where

$$x^{(j)} \in R^n, \; \mu_i \in R^n$$

μ_i - centroid for the cluster S_i.

So, if the proximity measure to the centroid is defined, then splitting the objects into clusters comes down to defining the centroids of those clusters. The number of clusters k is set by the researcher in advance.

Consider the initial set of k-means (centroids) $\mu_1, ..., \mu_k$ in clusters $S_1, S_2, ..., S_k$. In the first step, the centroids of the clusters are selected randomly or by a specific rule (for example, to select the centroids that maximize the initial distances between the clusters).

Include observations on the clusters whose means (centroid) is closest to them. Each observation belongs to only one cluster, even if it can be classified to two or more clusters. The centroid of each i cluster is then recalculated according to the following rule:

$$\mu_j = \frac{1}{s_j}\sum_{x^{(j)} \in s_i} x^{(j)}$$

Thus, the k -means algorithm involves the at each step of the centroid for each cluster obtained in the previous step.

The algorithm stops when the μ_i values do not change: $\mu_i^{шаг\ t} = \mu_i^{шаг\ t+1}$.

4.3 R/S image analysis

A description of the broad class of processes and phenomena, such as the processes of limited diffuse aggregation, the formation of viscous fingers in porous environments, turbulence, diffusion processes called flow, and the description of nature objects such as clouds, the earth surface and many others, in the terms of fractal geometry, defined a new direction in research- analysis of fractals. In this approach, the authors do not call an object - texture, but they call it a fractal [6]. Many authors note that most natural surfaces are spatially isotropic fractals and that the two-dimensional fields of intensity from such surfaces are also fractal. Fractal methods of analyzing image textural areas are based on the fact that a texture is a fractal, structure consisting of parts that are in a sense similar to a whole [9]. Thus, in the definition of fractal, a fractal self-similarity property is used. Many curves and surfaces are statistically similar, meaning that each part can be considered as an image of a whole in smaller size (Figure 1). The main feature of fractal characteristics is that images of man-made and natural objects have very different fractal dimensions. This allows for the successful use of fractal signs to detect artificial landscape changes from space photographs, the detection of artificial objects on images obtained from cameras and other tasks. For segmentation, use the characteristics that calculate the size of the fractal according to the formula (9).

Journal of Theoretical and Applied Information Technology
15th August 2018. Vol.96. No 15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

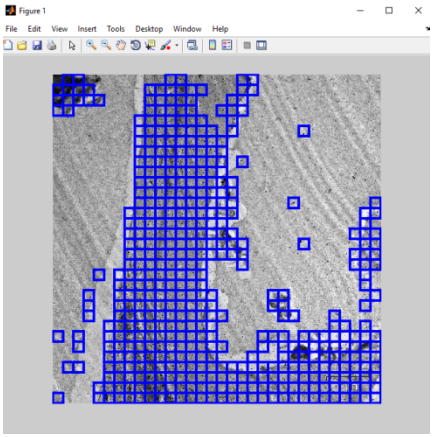

Figure 1: Selection of the area of study on the microphotosgraphy activated at 10 degrees.

The R/S-analysis method is a statistical method of analyzing time series or numerical sequences (primarily financial) to identify some important characteristics, such as the existence of non-periodic cycles, the "memory" of the process, degree of chaotic nature, etc. [10]. The method also applies to image analysis, because you can view the sequence of brightness functions along a line or curve. Interestingly, the method is used in chemistry to analyze the microphotographs (for example, lignin), and the method has shown some efficiencies.

An important characteristic calculated by R/S-Analysis is the Hurst exponent, usually referred H, which characterizes the degree of chaoticness of the process. H Hurst Coefficient is named after the British hydrologist of Harold Edwin Hurst. This coefficient allows a fractal dimension to be calculated and is therefore a fractal geometry instrument [11].

Note that when $0.5 < H < 1$, they say of the persistent (supported) behavior of the process, or that the process has a prolonged memory. In other words, if the process has been positive increment of the process for some time in the past, there has been a growth, the it will continue an increase in means. In other words, the probability that the process in the step deviates from the means in the same direction as the step is as large as the parameter is close to 1. Thus, persistent stochastic processes reveal a clear trend of change with a relatively small "noise".

In the case where $0 < H < 0.5$ it is said of a antipersistence of process [13]. Here, the high values of the process follow the low, and vice versa. In other words, the probability that the step of the process deviates from the means in the opposite direction (relative to the deflection in the step) is as large as the parameter is close to 0. When H =0.5, process deviations from the means are really random and do not depend on previous values, which corresponds to the case of Brownian movement.

To calculate fractal dimension D_φ, the following technologies were used in the experimental research process: the surface grid coating method, surface coating design method, R/S analysis.

The easiest way to implement a coating method is to drape a square grid on a fractal image and count the number of cells $N(r)$ to which the fractal falls [9]. When the distance between parallel grid lines becomes small enough, the value converges to the final value, which is the fractal dimension. The fractal dimension is calculated according to the following formula:

$$D\varphi = \lim_{r \to 0} \frac{\ln(N(r))}{\ln(1/r)} \quad (1)$$

where $N(r)$ is the number of cells, r is the size of the cell.

[10] shows that the image can display fractal properties only in a limited area of optical density change. The width of this area (scaling areas) is a characteristic of the degree of fractalness of the image. If the scaling area is small (for example, only a few times more than the elementary step quantum), you should consider the weakness of the fractal features when analyzing the image.

To test the method described, 36 microphotographs were selected. The following are the relevant formulas and test results:

$$\frac{R}{S} = (aN)^H, \quad (2)$$

from where

$$H = \frac{\log(R/S)}{\log(aN)}, \quad (3)$$

where

H – Hurst exponent;

S – mean-square deviation of a number of observations x;

N – number of observation periods;

$a = \frac{\pi}{2}$ – the specified constant, a positive number,

$$S = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - X_{cp})^2}, \quad (4)$$

where

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

X_{cp} - arithmetic mean of a series of observations x for N periods

$$X_{cp} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (5)$$

The extent of the accumulated deviation R is the most important element of the Hurst exponent calculation formula. It is generally calculated in the following way:

$$R = \max_{1 \le u \le N}(Z_u) - \min_{1 \le u \le N}(Z_u), \quad (6)$$

where Z_u is the accumulated deviation of a series x from the average X_{cp}:

$$Z_u = \sum_{i=1}^{u}(x_i - X_{cp}) \quad (7)$$

The Hurst exponent calculation formula shows that its growth is influenced by:

- span extension of oscillation R;
- reduction of mean-square deviation S;
- reduction of the number of observations N.

The parameter in (2) H ($0 < H < 1$) is called the Hurst parameter. The surface that characterizes the change of magnitude X is a random fractal. At the same time, the fractal dimension D_φ of the reviewed surface, defined by the method of covering it with spheres or cubes [11], is associated with the Hurst parameter H as a ratio:

$$D_\varphi = 3 - H \quad (8)$$

Fractal image level lines (scan lines) will also have a fractal shape, but their size is determined by the expression:

$$D_\varphi = 2 - H \quad (9)$$

Most of the methods used in practice to estimate fractal dimensions of images are based on the ratio (9).

In (9) the Hurst parameter H can be defined by analogy with the procedure for defining it during the processing of signals, based on the structural function of the increment of the two-dimensional value being investigated.

5. CELL WALL SEGMENTATION CALCULATION PROGRAM BASED ON TEXTURAL FEATURES

1. The aim of the programme is to conduct computer experiments and to accumulate quantitative data (e.g. textural feature values) for cell wall images treated at different temperatures (illustrated area studies on photomicrography activated at 10 degrees Figure 2). The program was implemented on MatLab at 18 textural features and with the method k-means with the size of the window 9x9, with orientation 0°. The following is one of the test results.

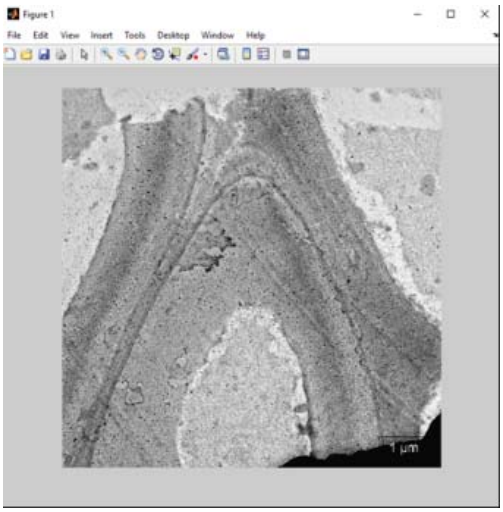

Figure 2: Original image

In the example, the size of the local fragment 9x9 is a carrier of the texture property. The textural feature values are written to the new matrix of the same size as the original. In the new matrix, the value is written to the coordinate, which is equal to the center of the scrolling window coordinate. Elements of the new matrix are obtained in some interval [A, B]. Next, clustering is performed, and then you can visualize the result of clustering (Figure 3).

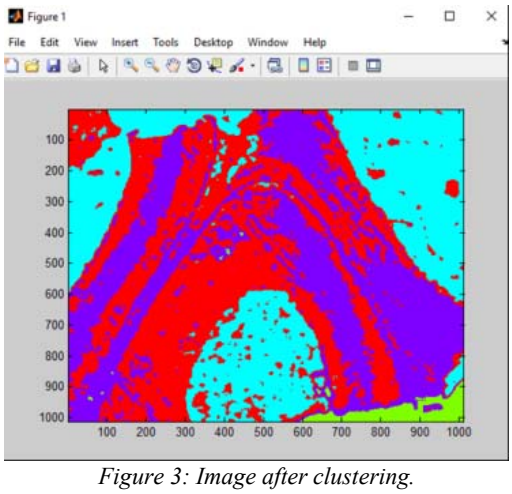

Figure 3: Image after clustering.

2. The program was implemented in the software package for the Statistica statistical analysis. Rectangular areas are randomly selected

Journal of Theoretical and Applied Information Technology
15th August 2018. Vol.96. No 15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

in Figure 4. Textural features are calculated for each area: $T_{10}, T_{12}, T_{13}, T_{14}, T_{15}$.

The most important result from the arboraceous clustering is the hierarchical tree. Ward's method is different from all other methods because it actually uses dispersive analysis methods to estimate the distances between clusters. The method minimizes a certain amount of square distances. More precisely, the distance between the clusters is increased by the sum of the squares of the object distances to the cluster centers resulting from their union. In general, the method seems to be very effective, but it seeks to create clusters of small size (Figure 5).

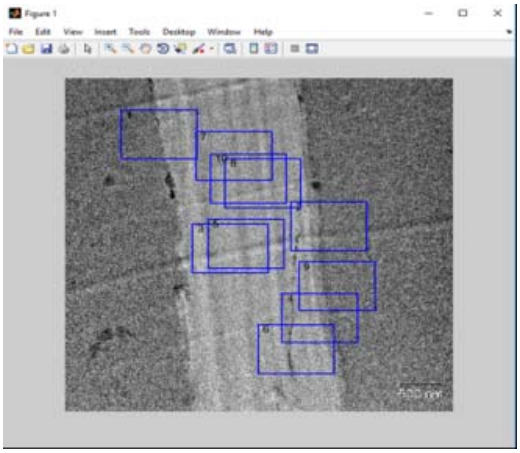

Figure 4: Original image.

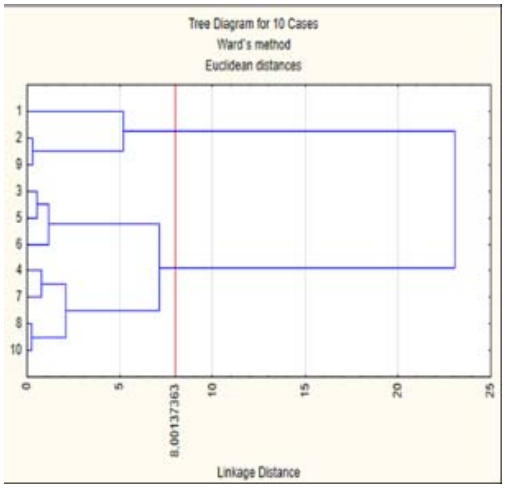

Figure 5: Dendogramma for the selected regions. Ward's method. Euclidean distance.

Each node in the diagram above represents the union of two or more clusters, the position of nodes on the horizontal axis determines the distance at which the respective clusters were combined. Table 1 shows the results of the dendrogramma numerically.

Table 1. Result of arboraceous clustering.

	Cluster Membership (11_template) Linkage distance = 8,00137 Ward's method Euclidean distances
	Cluster Membership
1	1
2	1
3	1
4	2
5	2
6	2
7	2
8	2
9	2
10	2

Graph of means allows you to view average values for each cluster in a line graph. (Figure 6).

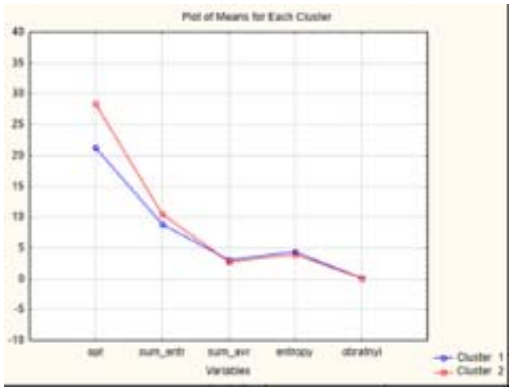

Figure 6: Average textural feature values for each cluster.

On the average textural feature values in Table 2, we can say that they do not intersect, and are far from each other well. This location of cluster center coordinates indicates a good clustering capacity.

Table 2: Mean value, mean-square deviation and dispersion of textural features

Variable	Descriptive Statistics for Cluster 2 (11_template)		
	Mean	Standard (Deviation)	Variance
opt	28,37031	1,536784	2,361704
sum_entr	10,45043	0,293490	0,086137
sum_avr	2,78292	0,179215	0,032118
entropy	4,0256	0,281526	0,079257
obratnyi	0,04285	0,004660	0,000022

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Analysis of variance displays the dispersion analysis tabulation (Table 3). The following table shows the values of the intergroup (between SS) and intra-group (Within SS) dispersion of the features. The smaller the value of the intra-group dispersion and the greater the value of the intergroup dispersion, the better the feature characterizes the objects accessory to the cluster and the more qualitatively clustering Q as clustering can also be measured by the value of the F-criteria (the higher, the better) and the level of relevance of p (the smaller the better). For two clusters, the results of the dispersion analysis show a good clustering quality: The importance level p is less than 5% everywhere.

Based on the visual presentation of the results and the Euclidean distance in table 4, we can determine the composition of each cluster (table 5).

Table 3: Dispersion analysis by textural features.

Variable	Analysis of Variance (11_template)					
	Between (SS)	df	Within (SS)	df	F	signif. (p)
opt	104,5598	1	24,15045	8	34,63614	0,000368
sum_entr	5,0888	1	1,05271	8	38,67160	0,000254
sum_avr	0,2567	1	0,19496	8	10,53225	0,011783
entropy	0,3748	1	0,50497	8	5,93718	0,040778
obratnyi	0,0021	1	0,00047	8	35,14316	0,000350

Table 4: The clusters to which the observation, the Euclidean distance, and the observation number belong.

	11_template		
	CASE_NO	CLUSTER	DISTANCE
1	1	1	1,18
2	2	1	0,54
9	9	1	0,65
3	3	2	0,67
4	4	2	0,83
5	5	2	0,44
6	6	2	0,98
7	7	2	0,78
8	8	2	0,35
10	10	2	0,24

Table 5: The composition of the classes allocated by Ward's methods and k-means.

Class number	Number of objects in the class	Class composition
1	3	1, 2, and 9 selected areas
2	7	3, 4, 5, 6, 7, 8 and 10 selected areas

6. R/S ANALYSIS CALCULATION PROGRAMME

The method is programmatically implemented in the MATLAB environment and an image analysis is performed. Images are divided into separate ares by software tools. If each area of the image belongs to the scaling region, it will be possible to identify, from the point of view of estimating the fractal dimension of the neighboring sections, the smaller electron-dense (black) formations that are likely to be the residue intracellular of lipid or protein content (fractional research method). In this case, the fractal dimension of the microphotographs of the cell walls of the wheat straw will lie in the interval of the norm, and the fractal dimension of the sections with the lignified layers is below the norm boundary (Figure 1). The image was divided into 1024 parts. For each part of the image, a fractal dimension calculation was performed.

The results of the fractionary method in the form of circles calculate the fractal dimension of the microphotographs. The blue circles correspond to the blue squares in the original image in Figure 7. And white circles point to the background in Fig. 7.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

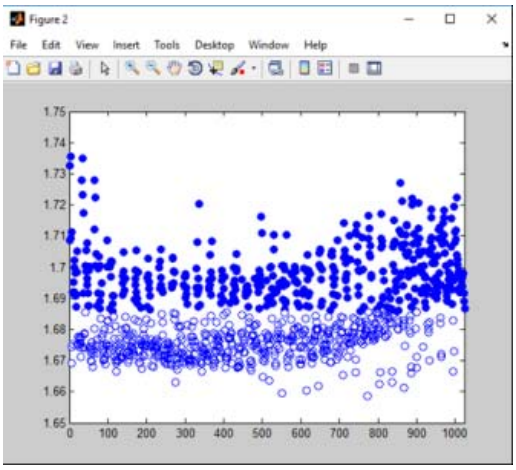

Figure 7: Calculation result.

7. CONCLUSION

The work focuses on texture image studies, texture segmentation, and R/S analysis. The source is the microphotography of vegetable raw materials, which is shredded in special mills. The work was done for the Institute of solid state chemistry and mechanochemistry of the Siberian branch of the Russian Academy of Sciences (ISSC SB RAS)

The work is about calculating different kinds of textural features and other image characteristics. Further, in principle, vectors of indicators can be "linked" to chemical reactivity. For example, a software system can be trained on examples based on neural-computer or other similar approaches. A large set of texture-study methods was analyzed, and the most informative ones were chosen from the methods reviewed. In the end, a program was developed to calculate various textural features. The aim of the programme is to conduct computer experiments and to accumulate quantitative data (textural feature values) for cell walls images treated at different temperatures. A total of 18 textural features are calculated. The results of calculation can be saved. The main purpose of the work is to create software tools. As a result, the following software tools were developed.

1. A program to calculate various textural features. The aim of the programme is to conduct computer experiments and to accumulate quantitative data (textural feature values) for cell walls images treated at different temperatures. A total of 18 textural features are calculated. The results of calculation can be saved.

2. Programmes for cluster analysis have been implemented, in the standard version and hierarchically, and experiments have been carried out also in a non-standard version. The first method used 18 textural signs with the 9x9 pixels scrolling window, and in the second 5 textural features with the window of 400 x 300 pixels. In the second approach, after carrying out research on textural features, five texture features were selected $T_{10}, T_{12}, T_{13}, T_{14}, T_{15}$, on which we conducted segmentation by area. By using $k-means$ and agglomerative methods, the signs clusterized well the areas in the images. The results show that these characteristics are informative.

3. Prototypes of programmes implementing R/S-analysis and fractal analysis [12-14] have also been implemented. Primary experiments which showed interesting results that could be useful in practice, namely, for the support of the work in the field of chemistry. Note that similar utterances are found in the scientific literature on the study of lignin, but without particular details [14, 15]. Segmentation based on the Hurst exponent in the microphotographs was performed. And with the help of these studies, we have come to the fact that the method has shown some efficiencies, where the background and the cell wall clearly differ from each other in images.

In the future, machine learning algorithms, for example, based on a brain-computer approach, can be used to determine chemical reactivity. After the system in the data set has being trained, you can then make predictions about chemical reactivity.

REFERENCES:

[1] H.A. Andreev, O.V. Bazarskii, A.S. Hlauberman, A.I. Kolesnikov, Yu.V. Korzhik and Ya.L. Khliavich, Analysis and synthesis of random spatial textures. *Foreign radio electronics*. Vol. 2, 1984, pp. 3-33.

[2] R.M. Kharalik, Statistical and structural approaches to the description of textures. *TIIER*. Vol. 67(5), 1979, pp. 98-119.

[3] A.A. Poapov, New information technologies based on probabilistic textural and fractal features in radar detection of low contrast targets. *Radiotechnology and electronics*. Vol. 48(9), 2003, pp. 1101-1119.

[4] N.V. Kolodnikova, *Overview of textural features for image recognition tasks*. Reports of the Tomsk State University of Control Systems and Radioelectronics Automated systems for information processing,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

management and design, pp. 117-118. Tomsk, 2004.

[5] A.V. Haidel and S.S. Pervushkin, Study of textual features for the diagnosis of bone tissue diseases by X-ray images. *Computer optics*, Vol. 37(1), 2013, pp. 113-119.

[6] V.T. Fisenko and T.Yu. Fisenko, Fractal methods for segmentation of texture images. *Instrumentation*. Vol. 56(5), 2013, pp. 63-70.

[7] M. Kholl, *Combinatorics*. Moscow: World, 1970.

[8] V.P. Ivannikov, V.V. Bielykh, V.A. Stepanov and R.T. Sufiianov. X-ray Fractal Analysis. *The Messenger of Izhevsk State Technical University*. 3, 2009, pp.150-154.

[9] P.V. Korolenko, M.S. Maganova and A.V. Mesniankin. *Innovative methods for analyzing the stochastic processes and structures in optics. Fractal and multifraktal methods, wavelet transformations: Training aid*. Moscow: Lomonosov Moscow State University Skobeltsyn institute of nuclear physics, 2004.

[10] E. Feder, *Fractals*. Moscow: World, 1991.

[11] E. Peters, *Chaos and order in capital market*s. Moscow: World, 2000.

[12] R.H. Barantsev, *Synergy in modern history*. Moscow: Editorial URSS, 2003.

[13] V.V. Petrov, *The structure of the teletraffic and the quality of service algorithm in the effects of self-similarity*. Moscow: The Moscow Energy Institute publishing house, 2004.

[14] A.P. Karmanov, Lignin: Structural Organization and Self-Organization. The theses of the report of the III All-Russian Conference: *Woodchemistry and Organic Synthesis*, p.9. Syktyvkar: Komi Scientific Center of UrO RAS, 1998.

[15] A.P. Karmanov and D.V. *Matveiev, Problems of timber chemistry and wood chemistry*. Syktyvkar: Institute of Chemistry of Komi Scientific Centre Ural Branch of the Russian academy of Sciences, 2001.

[16] V.S. Sidorova, Hierarchical Cluster Algorithm for Remote Sensing Data of Earth. *Pattern Recognition and Image Analysis*. Vol. 22(2), 2012, pp. 373-379.

[17] G. Abdikerimova, A. Bychkov, W. Yu, F. Murzin, N. Russkikh, S. Khairulina and E. Ryabchikova. Algorithms and software for the analysis of disordering the structure of cellular walls. *Bulletin of the Novosibirsk Computing Center. Series: Computer Science*. 40, 2016, pp. 1-14.

[18] G.B. Abdikerimova, A.L. Bychkov, S. Wei, F.A. Murzin, N.E. Russkikh, E.I. Riabchikova and S. Khairulin, Algorithms and software tools for texture analysis in microphotography. Reports of the International Conference: *Current Problems of pure and applied mathematic*s, pp. 124-126. Almaty, 2017.

[19] G.B. Abdikerimova, A.L. Bychkov, S. Wei, F.A. Murzin, N.E. Russkikh, E.I. Riabchikova and S. Khairulin, Library of software for texture analysis in microphotography. Russian scientific and practical conference with international participation, school of young scholars of memory V.V. Hubarev: *Data mining of signals, data and knowledge: Methods and tools*. Novosibirsk, 2017. In Press.