Dendritic crystallogram images classification

Rustam A. Paringer[1,2]*, Alexander V. Kupriyanov[1,2], Nataly Y. Ilyasova[1,2]

[1] Samara State Aerospace University, 34 Moskovskoe shosse, Samara, 443086, Russian Federation
[2] Image Processing Systems Institute of the Russian Academy of Sciences, 151 Molodogvardejskaya st., Samara, 443001, Russian Federation

e-mail: RusParinger@gmail.com

Abstract. A computer classification system of dendritic crystallogram images is presented in this paper. To improve the quality of classification we use an algorithm for the informative features formation, using methods of discriminant analysis. The method for receiving an informativeness estimation was used. As basic features are seven geometric characteristic was calculated. The research confirming the efficiency of the formed features for classification of dendritic crystallogramms images was conducted by means of classification by support vector machine. Of these, was selected most informative basic five features and one new feature was formed. The error classification decreased from 0.081 to 0.061. The algorithm possesses a sufficient level of universality and may be applied to increase the informativeness of any feature set. © 2015 Samara State Aerospace University (SSAU).

References

1. R. A. Paringer, and A. V. Kupriyanov, "Methods For Estimating Geometric Parameters of The Dendrite's Crystallogramms," Proceedings of 8th Open German-Russian Workshop "Pattern Recognition and Image Understanding", 226-229 (2011).
2. R. A. Paringer, "Separation and Analysis of Dendrites on Images of Diagnostic Crystallograms of Biological Liquids," Biomedsistems-2012, 105-108 (2012).
3. N. Y. Ilyasova, A. V. Kupriyanov, and R. A. Paringer, "Formation features for improving the quality of medical diagnosis based on the discriminant analysis methods," Computer Optics 38(4), 851-856 (2014).
4. R. A. Paringer, and A. V. Kupriyanov, "The Method for Effective Clustering the Dendrite Crystallogram Images," Electronic on-site Proceedings of 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding (OGRW 2014) (2014).
5. R. A. Paringer, and A. V. Kupriyanov, "Research Methods for Classification of the Crystallogramms Images," Proceedings of the 12th international conference "PRIP'2014", 231-234 (2014).
6. K. Fukunaga, Introduction to the Image Discrimination Statistical Theory, Nauka, Moscow (1979) [in Russian].

1 Introduction

The analysis of medical crystallogram images is an important part of medical diagnostics. Medical crystallograms are the structures formed at crystallization of salts as a result of drying the biological liquids (tears, blood, saliva, etc.). Computerization of the crystallogram image processing will enable to improve the quality of diagnostics and will reduce its time expenditures. The use of different devices and techniques of crystallograms imaging shall result to the fact that the structures density in crystallogram images may essentially differ. Therefore, in order to improve the overall quality of diagnostics it is offered first to estimate the scale of the crystallogram image shooting. The features that evaluate geometric characteristics of dendritic crystallograms have been developed [1, 2]. To evaluate the efficiency of the developed features, for a classification problem of the

presented class of on-scale images, a respective algorithm has been developed, based on the discriminant analysis algorithm [3]. The developed algorithm also enables to create spaces of new features ensuring the best possible separation in the definite classification problem [4, 5]. To evaluate the partibility of classes the discriminant analysis criteria are used [6]. Thus, for a well-defined set of features and the strictly defined image classes, a new space has been formed in which the image classes being of interest were divided in the best way. The effectiveness of the used features may be also evaluated by using the proposed approach.

2 Methodology

Features. Let us consider a dendrite model shown in Fig. 1. A, B, C, D – the "tops" of dendritic processes, E, F, G – the "roots" of dendritic processes, EF, FG – distances between processes, AE, BE, CF, DG – dendrites processes.

The algorithm consists of the following sequence of actions: thresholding, median filtering, skeletonization, identification of key elements and construction of a "map of dendrites."
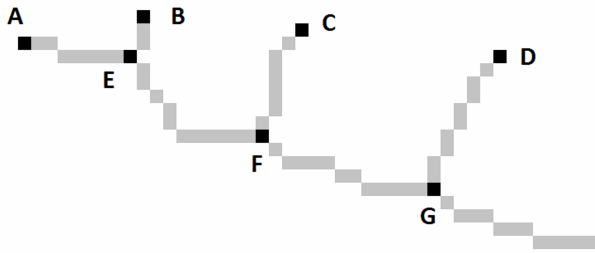

Fig. 1 Dendrite model.

The key elements include the "tops" and the "roots" of dendritic processes. The previously developed set of features [1, 2] has been added with four new features. The following features are calculated while processing:

- Density of the original image C_{ps}: is calculated after the median filtering as a ratio of the number of pixels in an object to the total number of pixels in the image.
- Thickness of dendrites C_t: is calculated as the number of iterations of the skeletonization algorithm.
- The other features shall be calculated according the "map of dendrites."
- The average distance between processes C_l: is the ratio of the sum of all distances between processes to the total number of processes. For Fig. 1 – $C_l = \frac{EF+FG}{4}$.
- The density of skeletonized dendrites C_{pd}: the ratio of the sum of lengths of all processes and all distances between processes to the total number of pixels in the image.

- The growth rate C_g: the ratio of the sum of lengths of all processes to the sum of distances between processes. For Fig. 1 – $C_g = \frac{AE+BE+CF+DG}{EF+FG}$.
- The average angle C_a: the ratio of the sum of angles of all processes to the total number of processes. For Fig. 1 – a process angle CF equals to an angle CFE in a triangle CFE.
- The symmetry factor C_s: C_s: the ratio of the number of the "tops" of dendritic processes to the number of the "roots" of dendritic processes. For Fig. 1 – $C_s = \frac{4}{3}$.

Algorithm of forming new features. Suppose we are given the sampling of n items divided into g classes and containing the proper p features. In discriminant analysis the measure of the sampling effectiveness are the partibility criteria which are calculated by the following formulas:

$$J_1 = tr(\boldsymbol{T}^{-1}\boldsymbol{B}),$$

where $\boldsymbol{T} = \boldsymbol{B} + \boldsymbol{W}$, $\boldsymbol{B}$ – is the among-groups scattering matrix, the elements of which are calculated by the formula:

$$b_{ij} = \sum_{k=1}^{g} n_k(\bar{x}_{ik} - \bar{x}_i)(\bar{x}_{jk} - \bar{x}_j), i, j = \overline{1, p},$$

$\boldsymbol{W}$ – is the intragroup scattering matrix, the elements of which are calculated by the formula:

$$w_{ij} = \sum_{k=1}^{g} \sum_{m=1}^{n_k}(x_{ikm} - \bar{x}_{ik})(x_{jkm} - \bar{x}_{jk}), i, j = \overline{1, p},$$

where x_{ikm} – is the value of the i feature for the m element in the class k, $\bar{x}_{ik} = (1/n_k)\sum_{m=1}^{n_k} \bar{x}_{ikm}$ – the mean value of the i feature in the class k, $\bar{x}_i = (1/n)\sum_{k=1}^{g}(n_k\bar{x}_{ik})$ – the mean value of the i feature in all classes, n_k – the number of elements in the class k.

The more the criterion value, the more the partibility of classes.

Suppose $x = [x_1, x_2 \dots x_p]^T$ – is the original feature vector. Let us consider the algorithm of forming new features of $y = [y_1, y_2 \dots y_p]^T$.

1. For the matrix $T^{-1}B$ let us define the values of eigen vectors $\boldsymbol{v}_i, i = \overline{1, p}$.

2. Let us define the vector of standardized coefficients $\boldsymbol{\beta}_i = [\beta_0, \beta_1 \dots \beta_p]^T, i = \overline{1, m}$, where the elements $\beta_i, i = \overline{0, p}$ are computed as follows:

$$\beta_0 = -\sum_{i=1}^{p} \beta_i\bar{x}_i, \beta_i = v_i\sqrt{n-g}, i = \overline{1, p}.$$

3. Let us compute the elements of the vector of new features by the formula:

$$y_i = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p, i = \overline{1, m}.$$

The exhaustive algorithm forms a group of new features using the discriminant analysis algorithm, thus

keeping the value of the intragroup criterion of features partibility. In order to form new features, all possible combinations of the original features are used. Besides, to compute the criteria, all possible combinations of the formed features are also used. The group of the formed features is to be selected, based on the value of the intragroup partibility criterion. The exhaustive search of all possible combinations of features guarantees the best division in the definite classification problem [3].

3 Results

A feature vector is calculated for each image. The classification error value is evaluated, based on the obtained data. Then, the basic features are subjected to the discriminant analysis procedure, and a set of new features is formed. After that, the classification error value is evaluated according to the formed features set. The classification error estimate was carried out by means of a support vector machine. The synthesis of a classifier was conducted by using the U-method. In order to implement the U-method, we have used the one-object elimination method that lies in the fact that all objects are included in a learning sample. By excluding one object, the classifier is synthesized by the remaining objects, and a test sample is formed by an unused object. This procedure is repeated for all sample objects, and the number of incorrectly classified objects is calculated. The classification error is defined as the ratio of the number of incorrectly classified objects to the total number of objects in the whole sample [6].

The basic data for the experiments is a set of diagnostic crystallogram images with a resolution of 256 by 256 pixels and consisting of 256 half-tone color tones. The set of basic data was divided into two classes, based on an increase value, which was used while shooting. The first class includes 76 images of dendritic crystallograms obtained at 100 times magnification (an example is shown in Fig. 2(a)). The second class includes 72 images of dendritic crystallograms obtained at 200 times magnification (an example is shown in Fig. 2(b)).

As a result of the experiment a new features was formed by multiplying basic features on vectors of standardized coefficients (β). "The Formed Feature 1" has been formed from five basic features. The features "Density of the Original Image C_{ps}" and "Thickness of Dendrites C_t" were not used for the forming. "The Formed Feature 2" was formed from all seven basic features and added for the comparison. The value of partibility criteria for all characteristics are presented in Table 1.

From the table above it can be concluded that the feature "Thickness of Dendrites C_t" was not used, thus, as judged by the value of the partibility criterion, it does not contain any meaningful information for classification. The feature "Density of the Original Image C_{ps}" was not used to form a new feature, since it significantly correlates with many other features that are used for the formation (Table 2). Classification error

with using all seven basic features is 0.081, for "The Formed Feature 1" 0.061 and for "The Formed Feature 2" same 0.061.
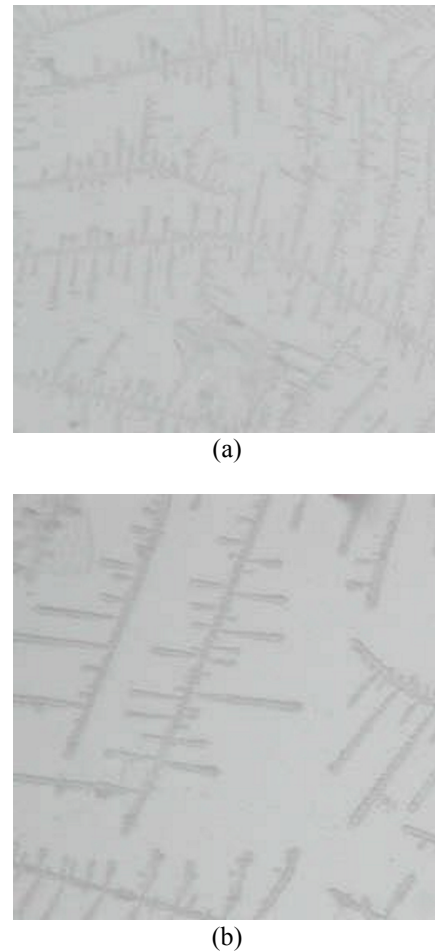

(a)

(b)

Fig. 2 Image of dendritic crystallograms (a) at 100 times magnification, (b) at 200 times magnification.

Table 1 The value of features partibility criteria and classification error.

Feature	J_1	Classification error
Density of the original image (C_{ps})	0,38397	
Thickness of dendrites (C_t)	1,6E-07	
The average distance between processes (C_l)	0,64174	
Growth rate (C_{pd})	0,03821	0.081
Density of skeletonized dendrites (C_g)	0,36071	
Angle factor (C_a)	0,23939	
Symmetry factor (C_s)	0,00854	
The Formed Feature 1	0,81747	0.061
The Formed Feature 2	0,81798	0.061

Despite the fact that the value of the partibility criterion "The Formed Feature 1" is little less than "The Formed Feature 2", their classification error values are

equal. Hence, it follows that the proposed algorithm of forming features does not only allow to evaluate the effectiveness of original features, but also allows to create new, more effective, criteria to solve the proper classification problem.

Table 2 Correlation matrix of the original features.

	C_{ps}	C_t	C_l	C_{pd}	C_g	C_a	C_s
C_{ps}	1,00	-0,02	-0,63	-0,33	0,78	-0,24	-0,45
C_t	-0,02	1,00	0,08	0,11	0,46	0,05	0,13
C_l	-0,63	0,08	1,00	-0,12	-0,55	0,44	0,12
C_{pd}	-0,33	0,11	-0,12	1,00	-0,31	-0,04	0,66
C_g	0,78	0,46	-0,55	-0,31	1,00	-0,12	-0,39
C_a	-0,24	0,05	0,44	-0,04	-0,12	1,00	-0,25
C_s	-0,45	0,13	0,12	0,66	-0,39	-0,25	1,00

4 Conclusion.

The algorithm of forming the space of efficient features based on the discriminant analysis was used in the classification problem of images of dendritic crystallograms. The algorithm allows to improve the efficiency of the classification of dendritic crystallograms taken at different magnifications, whereas the classification error of crystallograms is maximum 6.1%.

It is shown that the algorithm allows to evaluate the effectiveness of the used features to solve classification problems and to eliminate the linear dependence between the features. It is also worth pointing out that the more the initial features are used in the algorithm, the more accurate links would be set up between the features. The used algorithm of forming features possesses a sufficient level of universality and may be applied to increase the informativeness of any feature set. Implementation of the algorithm of forming efficient features enables to reduce the classification error.

Acknowledgments