MULTIVARIATE ANALYSIS FOR IMAGE RECOGNITION SYSTEM TO ASSESS THE QUALITY OF THE MINERAL SPECIES

Olga Baklanova[1], Alexander Baklanov[2], Olga Shvets[3]
D.Serikbayev East-Kazakhstan State Technical University
Ust-Kamenogorsk, KAZAKHSTAN
e-mail: [1]OEBaklanova@mail.ru, [2]ABaklanov@ektu.kz,
[3]Olga.Shvets75@gmail.com

Abstract

This paper contains development of methods and algorithms of image recognition for mineral rocks. It is described algorithms of the cluster and morphological analysis for definition of rocks composition on colors and shapes. This approach is explained by the fact of the possible presence of objects with similar color-brightness characteristics, but with different shapes and there are objects with similar color-brightness characteristics also. Preliminary definition of group membership allows reducing the computational complexity of classification. It is determined the sorting into groups according to color of the object at the stage of segmentation. It is described and discussed the example using multivariate analysis for mineral rocks recognition.

1 Introduction

Minerals are homogeneous in composition and structure of the rocks and ores. They are natural chemical compounds resulting from various geological processes. Historically minerals initially determined by color and shape [1].

The development of computer vision system for mineral rocks is discussed in offered work in order to assess the qualitative composition of mineral rocks, in particular some problems of a technique and image recognition technology.

2 Materials and Methods

2.1 Methods of Identification of Mineral Rocks Images

Let us consider a sample of anode copper slag as an example (Figure 1. Micrographs of this sample were kindly provided by Eastern Research Institute of Mining and Metallurgy of Non-ferrous Metals (Kazakhstan, Ust-Kamenogorsk).

According to experts on microscopy of minerals from Eastern Research Institute of Mining and Metallurgy of Non-ferrous Metals at this picture there is no minerals having dependent on the direction of the plane of polarization of light. In this picture you can detect metallic copper and the following minerals: cuprite Cu_2O, magnetite Fe_3O_4, Delafosse $CuFeO_2$, silicate glass.

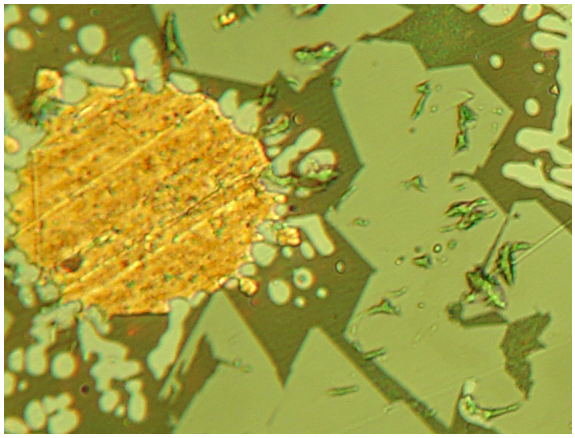

Figure 1: Micrograph of a sample of slag copper anode, increasing in 500 times.

Cuprite Cu_2O can be identified as follows: it is characterized by the shape of a round shape, color - it is light gray (sometimes with a slight bluish tint). Fe_3O_4 magnetite on micrographs may also be detected by color and shape. Color of magnetite on micrographs is dark gray. Shape is angular, as expressed by technologists, "octahedral".

Delafossite $CuFeO_2$ micrographs can allocate to the needle shape and gray (with a brownish tint) color.

Metallic copper on the micrographs can be found on the following criteria: color - yellow, shape - round, without flat faces.

Silicate glass - is a dark gray mass fills the rest of the space that is left of the other minerals. These data indicate that for real micrographs slag samples (and some other minerals) it is possible to use automated qualitative assessment of the mineral composition. After receiving the full image it is often needed to treat it, mainly to simplify further analysis.

2.2 Methods of Cluster Analysis for Mineral Rocks Images

Clustering - is the automatic partitioning of a set of elements into groups according to their similarity. Elements of the set can be anything, for example, data or characteristics vectors. Themselves groups are also called clusters [2].

In our case, using algorithms of cluster analysis will be the identification of ore minerals by color and texture characteristics of color-coded minerals identified in images taken in reflected light using a microscope [3].

In general, the K-means method segments the image on K different clusters (areas) located far away from each other based on certain criteria [4].

Segmentation method "K-means" is implemented through a two-step algorithm that minimizes the sum of distances "point-to-centroid" obtained by summing over all K clusters. Another words, the purpose of the algorithm is to minimize variability within clusters and maximize variability between clusters [5].

The purpose of cluster analysis - to implement such a partition of the n-dimensional feature space for k-clusters, in which the length between centroids of the resulting

clusters would be greatest, it is shown in the expression (1).

$$d_{i,j} \to max, \tag{1}$$

where $d_{i,j}$ is the distance between centroids of i-th and j-th clusters, $i, j = 0, \ldots, k$.

In this case, the most appropriate method of solving the problem of clustering is classic algorithm of unsupervised learning - a method of k-means (k-means method). Clustering incrementally in this case is as follows:

1. Lets specifie the number of clusters K, you want to find.
2. It is randomly selected K vectors $'$ from the set of vectors in selected space. These vectors are centroids of the clusters on the initial calculation stage.
3. Lets calculate the distance from each vector space used to each of the obtained centroids in step 2. It can be used metric (2)-(3) to determine the distance.

$$D_{(x,y)k} = \sqrt{\sum_{p=1}^{n} (P_{x,y}^{p} - P_{k}^{p})^2}, \tag{2}$$

$$D_{(x,y)k} = \sum_{p=1}^{n} |(P_{x,y}^{p} - P_{k}^{p})|, \tag{3}$$

   where:

   - (x, y) – coordinates of the observation,
   - $k \in [1, K]$ - cluster index,
   - n – dimensionality of the used feature space,
   - $p \in [1, n]$ – index of the feature observations.

4. Than we determine the centroid of the cluster to which the distance from the observation is the smallest. This cluster matched the observation.
5. Going through all available vectors we can recalculate centroids for each resulting cluster according (4).

$$P\prime_{(x,y)k}^{n} = \frac{1}{S(k)} \sum_{s=1}^{S_k} (P_{(x,y)s}^{n}), \tag{4}$$

   where:

   - k - cluster index,
   - $S(k)$ – number of observations related to the cluster index k,
   - s - indexes of the observations,
   - $P_{k}^{\prime n}$ - new value n-th feature of centroid cluster k.

6. Iterative process stops on steps 3-5 when the process of centroids changes stops or centroids will be fluctuate around some stable values. If the step of centroids change reaches a predetermined value it is possible to stop iterations.

Algorithm of the program can provide additional information after completion of the segmentation such as:

- a sum of distances "point-to-centroid";
- coordinates of centroid as well as some other data.

Algorithm of K-method can converge to a local optimum, when the separation points move any point to another cluster it increases the resultant sum of the distances. This problem can be solved only by a reasonable (successful) choice of initial points [7].

2.3 Methods of the Morphological Analysis of Mineral Shapes

Identification of the classification parameters is one of the primary task in pattern recognition [6].

It is offered the following description of the basic model of the object on the basis of morphological features (5–8):

$$M = \langle C, F, G \rangle, \tag{5}$$
$$C = \langle H, Sc, V \rangle, \tag{6}$$
$$F = \langle A \rangle, \tag{7}$$
$$G = \langle S, \beta \rangle, \tag{8}$$

where:

- C - cortege of metrics color of the object;
- F – cortege of morphological metrics of the object;
- G – geometrical metrics of the object;
- H - tone, Sc - saturation; V - value;
- A - number of allocated erosion circles;
- S - area of the object , β - the ratio of the long axis to the short one.

Proposed formalized description is focused on the entire spectrum of morphologically recognizable object parameters.

3 Results and Discussion

Nowadays developed automated image recognition system for assessing the qualitative composition of mineral rocks consists of 7 main subsystems [3]:

1. Research and getting micrograph rock.
2. Input and identification micrograph rock.
3. Pre-processing: improving the quality.
4. Definition of image reduction threshold [8].
5. Select the feature vector for cluster analysis.
6. Cluster analysis of color image to determine the mineralogical composition of rocks.
7. Morphological analysis of mineral shape to determine the mineralogical composition of rocks.

Each cluster includes a certain number of points. Given the ratio of the number of points allocated in each cluster with a number of common points can be displayed relative rates of minerals in rock samples. Various minerals marked in different colors. In this case, the metallic copper is red, magnetite - blue cuprite - orange. The result of the segmentation is shown in Figure 2.

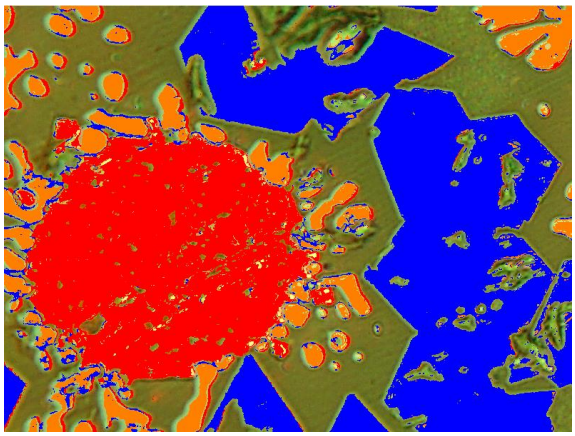

Figure 2: DFD - diagram decomposition subsystem “The result of cluster analysis”.

Considered sample has the following content of useful elements:

- Magnetite - 28.45%;
- Metallic copper - 18.45%;
- Cuprite - 7.92%.

4 Conclusion

In this article we have considered the development of segmentation algorithms for solving tasks of geological material analysis. We have proposed two different methods of ensuring the stability of results, based on pre-selection of centroids according to a few established principles in order to increase the stability of the segmentation pattern. The method of single-component searching consists of preliminary segmentation of the image based on a single variable vector. Every cluster has to get a matching segment assigned to it after the stage of segmentation. We have taken the average of every component of given variable space inside the assigned segment as initial values of the centroids. The basis of another method includes defining a point situated inside clusters defined by variables in N-dimensional space. We have proposed non-uniform partition of analyzed variable space followed by selection of initial values of centroids with maximum difference of color brightness characteristics in order to ensure stability. The program complex has been written in the language C# Visual Studio 2015. It was developed for results of research checking.

References

[1] Farndon J. (2006). *The practical encyclopedia of rocks and minerals.* Lorenz Books, London.

[2] Mandel J. (1988). *Cluster analysis.* Moscow: Finance and statistics.

[3] Baklanova O.E., Shvets O.Ya., Uzdenbaev Zh.Sh. (2014). Automation System Development For Micrograph Recognition For Mineral Ore Composition Evaluation In Mining Industry. *IFIP*. Vol. **436**, pp. 604–613.

[4] Odell P.L., Duran B.S. (1974). *Cluster Analysis: A Survey*. Springer-Verlag, NY.

[5] Huang Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *D. Mining and Knowledge Disc.* Vol. **2**, pp. 283–304.

[6] Gonsalez R.C., Woods R.E. (2011). *Digital image processing*. Pearson Education.

[7] Baklanova O.E., Shvets O.Ya. (2014). Methods and Algorithms of Cluster Analysis in the Mining Industry. Solution of Tasks for Mineral Rocks Recognition. *Proc. 11th Int. Conf. Sig. Proc. and Multimedia App. (SIGMAP'2014)*, pp. 165–171.

[8] Baklanova O.E., Shvets O.Ya. (2014). Development of Methods and Algorithms of Reduction for Image Recognition to Assess the Quality of the Mineral Species in the Mining Industry. *LNCS*. Vol. **8671**, pp. 75–83.