

## EXAMINATION OF CLUSTERING IN EUTECTIC MICROSTRUCTURE

The eutectic microstructures are complex microstructures and a hard work to describe it with few numbers. The eutectics builds up eutectic cells. In the cells the phases are clustered. With the development of big databases the data mining also develops, and produces a lot of method to handling the large datasets, and earns information from the sets. One typical method is the clustering, which finds the groups in the datasets. In this article a partitioning and a hierarchical clustering is applied to eutectic structures to find the clusters. In the case of AlMn alloy the K-means algorithm work well, and find the eutectic cells. In the case of ductile cast iron the hierarchical clustering works better. With the combination of the partitioning and hierarchical clustering with the image transformation, an effective method is developed for clustering the objects in the microstructures.

*Keywords:* Eutectic, microscopy, image analysis, data mining, cluster analysis

### 1. Introduction

A correlation can be discovered in the properties, mainly mechanical properties and the microstructure of metals and alloys [1]. So the appropriate description of the microstructure is an important task. The microstructural descriptors are used to qualify the product in several cases. As complex the observed microstructure it become harder to find that measurable features which are in strong correlation with the behavior of the given metal or alloy. Nowadays the computational image analysis gives more possible solution which can easily make the descriptions robustly. Besides the high computational power of the computers makes these analysis extremely fast.

The modern imaging devices make big size images which themselves means a large amount of data. Besides this, it is possible to make more images in a short time period from the examined microstructure. The high resolution of the imaging devices means detailed visual information. Advisable to organize the images and the information to databases instead of separate handling.

Nowadays a large amount data is stored and the databases are grown large. To find the information about this type of database which contains numerical and qualitative data is not efficient with the classical methodology. This makes the rapid development of data mining. The data mining collects the efficient methods to search and finds the information and contexts in large datasets. If the images, image analysis workflow, and results of the analysis are organized by a database, the data mining can find the necessary information to describe the images.

Typical area of the image analysis is the qualification the microstructure of castings. This study focused on the microstructure not on the analysis of the fault/porosity structure but the method can be used to describe the porosity without any changes. In standard case it was enough to measure the ratio between the eutectic and primer phase volume in the analysis of hypoeutectic alloys. Some cases the “coarseness” is also determined through the distance between the phases or through the size of the phases. But in the descriptions of the solidification of eutectics it can be seen that the eutectics builds a complex structure which has an effect to the properties of the alloys. In that alloys which contains a low volume fraction of eutectics the spatial distribution of the phases are not uniform [2]. The examination of the microstructures it can be stated that the phases are clustered [3]. As it was mentioned above that the properties of the alloys are determined by not only the average size and distance of the phases but the spatial distribution also. The data mining examines the clustering [4,5]. The clustering algorithms can efficiently find the clusters in a large dataset [6]. There are two basic algorithm groups of clustering: partitioning and hierarchical.

In partition clustering algorithm a point represents the data in the parameter space. In this case in the plane of the images a point represent the phases/objects. The algorithms search for the clusters based on the predefined geometrical distance between the phases/objects. The principle of the search is that the algorithm minimizes the distances of the objects in a cluster and maximizes the distance between of the clusters. The objects are situated in the plane of the image so the partitioning methods are an appropriate choice. An example will be introduced in the following.

\* UNIVERSITY OF MISKOLC, INSTITUTE OF PHYSICAL METALLURGY, METAL FORMING AND NANOTECHNOLOGY, H-3515 MISKOLC EGYETEMVÁROS

\*\* FUX CO. H-3527 MISKOLC BESENYŐI STR. 8.

# Corresponding authors: kornel.bortnyik@gmail.com, peter.barkoczy@fux.hu

A general problem in image analysis is that in most cases one geometric point do not represent clearly one object in the plane of the image. It follows that the partitioning methods do not give an exact clear picture about the spatial distribution of the phases. The modified definition of the distances can give a solution if the relation of stereology remains correct. In this way the hierarchical methods become more effective. The basis of these algorithms the similarity between the objects, which is handled by a distance in the computations. In this methods can takes into consideration the size, the shape, etc. of the objects. The problem of these methods, that the description of the objects means a following data analysis – classification tasks which have a large effect to the efficiency and accuracy of the clustering.

In the developed method it was combined the advantage of the partitioning and hierarchical methods with the image analysis algorithms. An efficient clustering method was developed which can analyze the spatial distribution during the image processing and analysis period. This solution reduces the computational time and the size of the meta-database.

## 2. Materials and methods

The analysis were performed on images made from the microstructure of commercial 3003 Aluminum-Manganese alloy and images made from the microstructure of ductile cast iron.

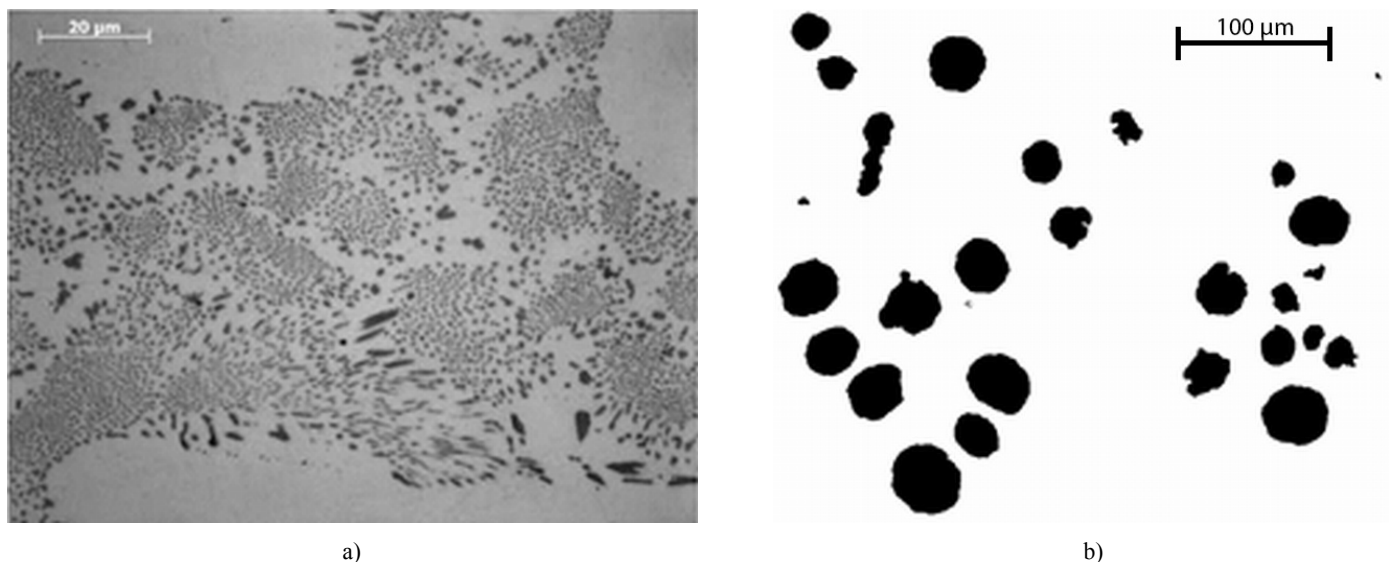


Fig. 1. Examples of the examined images, a) the micrograph of the AlMn alloy. The dark spots are the intermetallic AlMn phases, which forms during an eutectic solidification between the grains (dendrites) of the solid solution. The eutectic cell structure clearly visible in the micrograph. b) reference micrograph from the graphite structure of a ductile cast iron. Black objects are the graphite grains, white pixels means the background

## 3. Results and discussion

The main problem of the K-means algorithm that the number of clusters is unknown. This value must be set by the user before the clustering. If it could be determined by the K-means method the following steps will be made:

1. Set the number of the clusters ( $k$ );

The samples of AlMn alloy cut from a homogenized DC slab. The sample grinded and polished mechanically after the cut. The prepared surface etched by immersion to the HF reagent diluted with water. The micrographs made by an Axio Imager M1m microscope. The necessary image processing steps performed by the self-developed Cprob image analyzer [9].

Intensity compression and gamma correction were necessary to apply before analysis of the images. A Brensen automatic segmentation was used to segment the objects.

As Fig. 1 shows the majority of the AlMn intermetallic phases in the eutectic cells are compact so one point can be used to represent the spatial position of the phases. K-means algorithm was chosen to analyze the spatial distribution of the phases. K-means is the mostly used partitioning algorithm.

Image processing of the reference images taken from the microstructure of a ductile cast iron was not necessary. The graphite particles are black the background is white. A simple segmentation was enough before the analysis of the objects. Fig. 1b shows that not all of the graphite particles can be described by one geometrical point in the plane of the image. For this the K-means method cannot gives an appropriate result in principle. In this case a self-developed hierarchical clustering based method was used. The algorithm builds the dendrogram which is the basis of the hierarchical clustering. The analysis of the dendrogram was made as similar to the original hierarchical clustering methods.

2. Build up  $k$  cluster in random manner, and calculate the center of the clusters OR, directly set up  $k$  cluster center in random manner.
3. Clustering the point set with the predefined cluster centers.
4. Calculate the new cluster centers.
5. Repeat the steps 1-4 till one convergence criteria is not valid (generally that the clusters don't change).

Generally the number of the clusters is unknown during the analysis of images. To solve this problem the following method is often used: the introduced K-means method is performed in the point set with  $k = 1$ . The number of clusters is increased by one till a predefined  $n$  positive integer value. It can be seen in Fig. 2 that the inner distance in the clusters (1) decreases with the increasing of  $n$  at the small  $n$  values. After a specific  $n$  number of clusters the inner distance in the clusters remains constant. To fit a straight line to the decreasing section of the curve, and the constant values, the intersection shows the real number of clusters.

$$d_{inner} = \frac{\sum_{i=1}^n \sum_{j=1}^n (1 - d_{ij}) \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{(n-1)^2} \quad (1)$$

It can be found the clusters with this method at that case when the partitioning algorithms works. In the case of the examined AlMn alloy the majority of the intermetallic phase are compact, the K-means algorithm is applicable (Fig. 3). It can be seen in the Fig. 1 that the eutectic cells are distinguishable. In the eutectic cells the intermetallic phases build up clusters. The aim of the analysis is to find the eutectic cells.

But the Fig. 3c shows two different intersection in the curve instead of theoretical one. One intersection shows 7, the other intersection shows 20 clusters. Fig. 3a shows the clustering with 7 centers and Fig. 3b shows the results with 20 clusters. It can be seen that the 20 clusters means the eutectic cells. After the clustering the size, the shape, the inner distance, etc. can be measured of the eutectic cells. Not only the microstructure can be described in more detail but the effect of the circumstances of solidification or phase transformations of the microstructure can be studied more deeply.

The first intersection on the curve shows that the eutectic cells are also clustered, with gives the possibility of more de-

tailed analysis of the microstructure. The great advantage of the method that it works robustly with that alloys which contains low volume fraction of eutectics too.

As it was mentioned in the introduction in the case of the ductile cast irons, where not only nodular graphite builds up the microstructure the K-means algorithm not work properly in principle. In these cases the non-nodular graphite particles cannot be clearly described by one point. In that case the hierarchical clustering gives better result, but it needs to declare the similarity of the graphite particle as a distance.

There are a several cellular automata in image processing and analysis. The convolutions are totalistic automata and the gray and binary morphological operators are synchronous deterministic automata. It is natural because the images easily transform to a cellular automata universe. In the field of morphological transformation stochastic automata are used where the operation of transformation rule is depends on given range of a random variable.

If the image is transformed to a universe, where the state of the background is 0 and the state in the graphite particles is 1 and a stochastic dilatation is applied, the graphite particles will start to growth as in a microstructural simulation [10]. During the growth the dilatation performed as a distance transformation, so all cell has state variable which shows its distance from the original contour. If the dilatation is performed till the growing particles are impinged, in the first impingement event the given cell shows the distance between the impinged objects. From the impingement events, and from the distances the dendrogram can be plotted (Fig. 4). When all cells change its state from 0 to one the whole dendrogram is ready from the properly organized data. Additionally to search for the common contours of the impinged particles the SKIZ structure is got. This SKIZ structure is the closest to the Voronoi tessellation of the points [11]. The analysis of the dendrogram gives the clusters, so not only the distance of the particles has an effect to the clustering but the size and

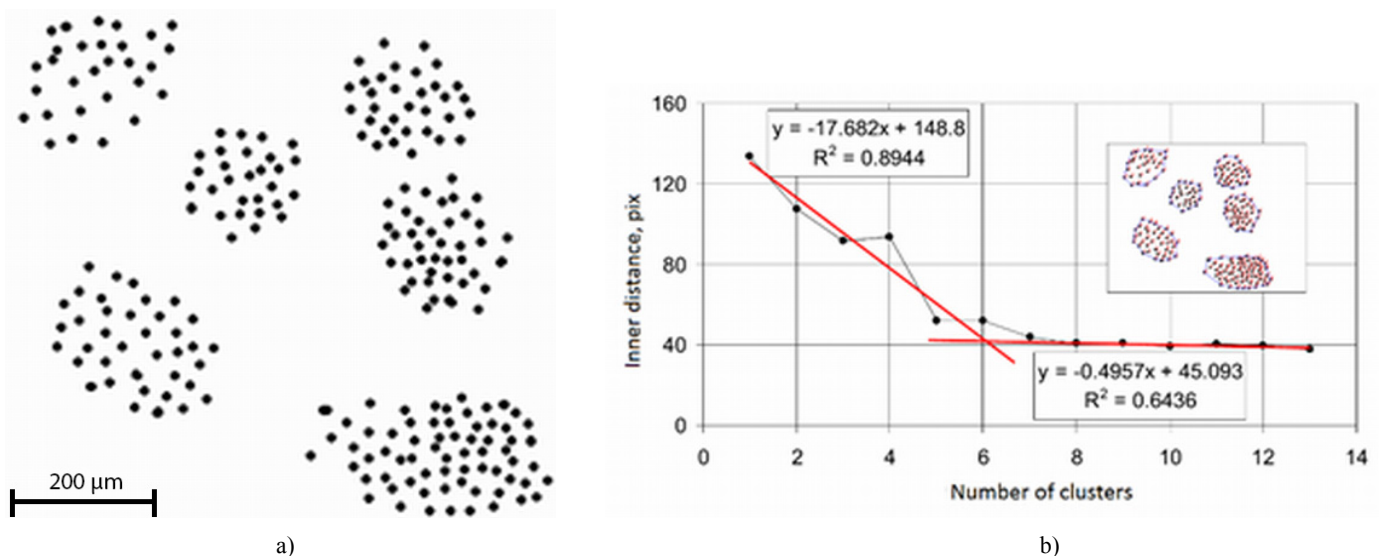


Fig. 2. Search for the number of the clusters with K-means method. The a) image contain 6 clusters. During the analysis the number of the clusters is increased from 1. In all cases the inner distance is calculated. The inner distance decreases at the small number of clusters, and till 6 it remains constant. It can be clearly seen that the two fitted straight line intersect each other at 6

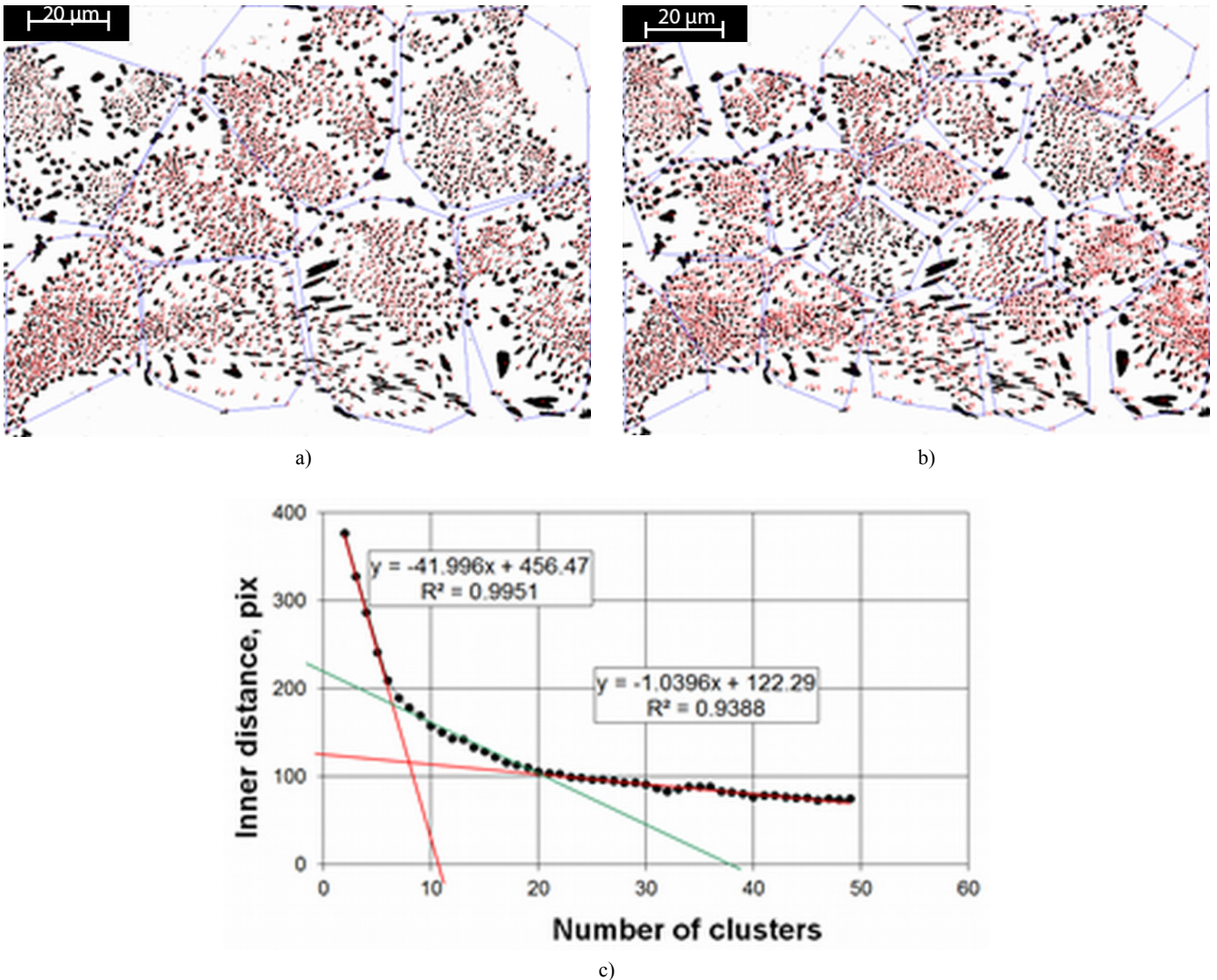


Fig. 3. Results of the cluster analysis of the microstructure of AlMn alloy. The clustered phases are the dark intermetallic phases. K-means algorithm is used. During the determination of cluster centers two intersection (7 and 20 clusters) can be observed. The real number of the eutectic cells are 20. At 7 the intersection shows that the eutectic cells are also clustered

the shape of the objects. In that cases where these parameters has a significant effect to the clustering the hierarchical method gives better result.

Another advantage of the hierarchical method that it can handle the so called outliers. These objects are not belong to either clusters. These are situated separately. As it can be seen in the image this phenomenon can be occur. If the Figs. 3a and 3b deeply examined it can be found particles mainly in the boundary of the eutectic cells which worse the result of the clustering. The clustering would be better if these objects handled as outliers. But K-means clustering cannot handle the outliers it assign all objects to the clusters. One solution is to use the Fuzzy based C-means algorithm.

Complex microstructures can be analyzed by the introduced method and robustly calculates the complex results for the research work, and the qualification of the microstructures.

#### 4. Summary

The article showed that clustering occur in the eutectic microstructures. The description of the clustering gives an additional features from the microstructure. That's for it is an important task to determine methods for clustering the spatial distribution of the particles. After the clustering easily can be determine the size, the shape, etc. of the clusters.

In the case when the particles are compact, and the spatial position can be clearly described by one point, the K-means algorithms work well. This was demonstrated by the analysis of the Al-Mn eutectic structure. The disadvantage of the K-means method that the number of clusters has to be set by the user before the analysis. One simple method exist to determine the number of clusters with the repeated application of the K-means method. The analysis of the structure of Al-Mn eutectics, two results were get. The higher number of clusters gives the eutectic cells, the smaller number of clusters shows that the eutectic cells

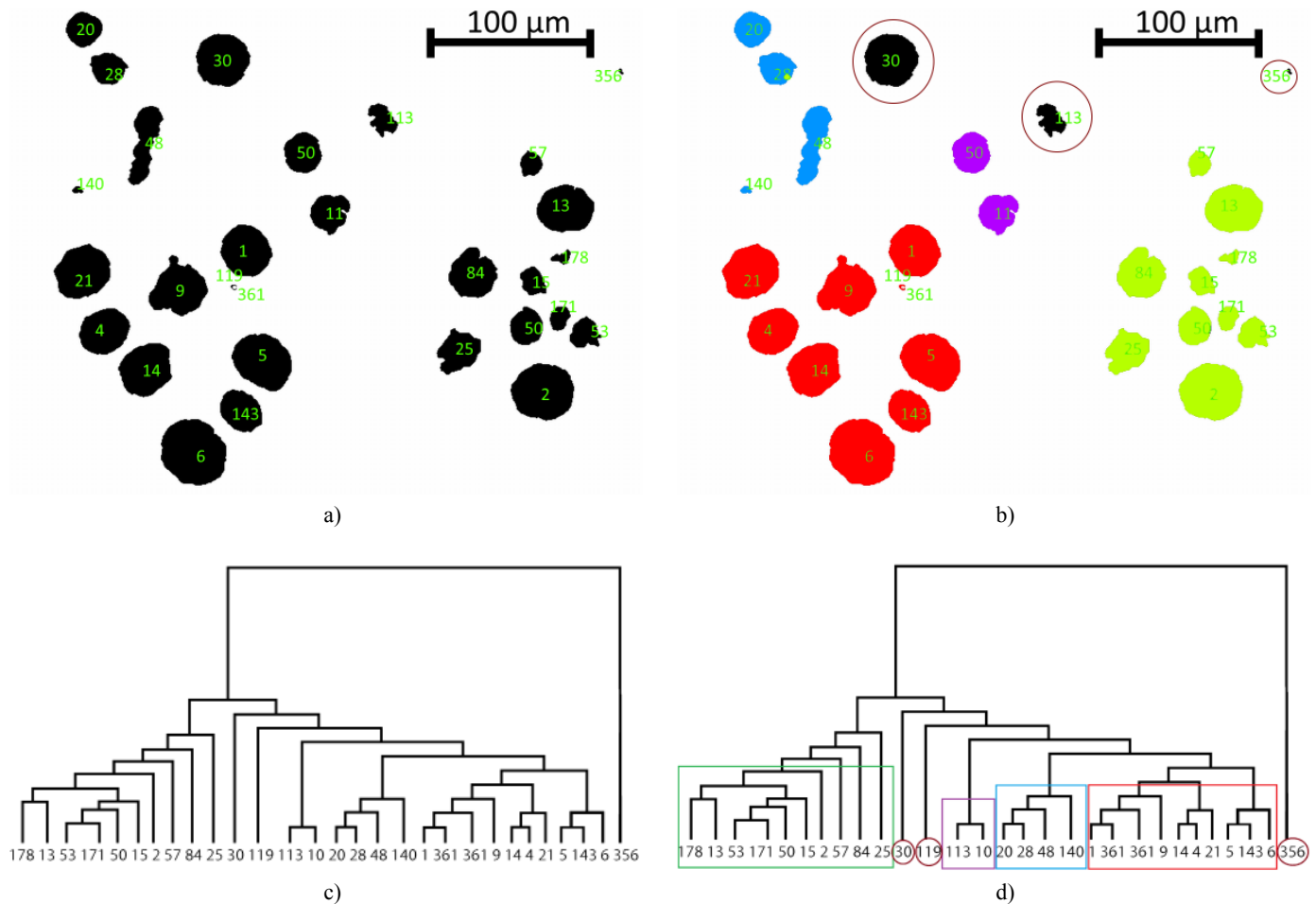


Fig. 4. Spatial clustering of the graphite particles. a) the reference image, b) the clusters signed by different gray tone, c) the result dendrogram, d) the clustered dendrogram with the outlier objects

are clustered. The K-means algorithm cannot handle the outliers, all particles are assigned to clusters.

There are several case where one geometrical point cannot describes the spatial position of one objects. Microstructure of a nodular cast iron shows an example to this case. In this case the hierarchical clustering method gives better result. Continuous stochastic dilatation cause impingement of the particles during the transformation. From this impingement the dendrogram of the hierarchical clustering can be built, and the clustering can be performed. One advantage of the introduced method that it can handle the outliers and gives a more accurate result.

#### REFERENCES

- [1] Gácsi Z., Sárközi G., Réti T., Kovács J., Csepeli Zs., Mertinger V., Sztereológia és képelemzés. Miskolc, MicroPress Nyomda, 2001.
- [2] R. Nunes et al, ASM Metals Handbook **02**.
- [3] Dr. Köves Elemér, Alumínium kézikönyv, Műszaki Könyvkiadó, Bp., 1984.
- [4] J. Han, M. Kamber, Data Mining Concepts and Techniques 3rd edition. Waltham USA, Elsevier, 2012.
- [5] M. Anderberg, Cluster Analysis for Applications. Morgan Kaufmann Publishers, 1973.
- [6] Knowledge-Based Clustering, New Jersey, John Wiley & Sons, 2005.
- [7] T.H. Cormen, C.E. Leiserson, R.L. Rivest: Introduction to Algorithms 3rd edition, MIT 2009.
- [8] S. Wolfram, A new kind of science, Wolfram Media, 2002.
- [9] Cs. Pólska, Z. Gácsi, P. Barkóczy, The Effect of Melt Flow on the Dendrite Morphology Materials Science Forum **508**, 169-174 (2006).
- [10] Sz. Gyöngyösi, P. Barkóczy, Scaling cellular automaton simulations of short-range diffusion processes Materials Science Forum **729**, 150-155 (2013).
- [11] P. Barkóczy, Gyors szomszédság kereső algoritmus véletlen rácsú cella automatához In: Gácsi Zoltán, Barkóczy Péter, Sárközi Gábor (szerk.) Képfeldolgozók és Alakfelismerők IV. Konferenciája. NJSZT-KÉPAF. 308 p.