

# Definition of Typical Textures of Sedimentary Grains Using Co-occurrence Features And *K*-means Clustering Technique

Aleš Křupka

**Abstract**— The paper deals with a definition of typical structure forms, which can be extracted from the surface of sedimentary grains. The co-occurrence features are used for this purpose. To find typical patterns, the *K*-means clustering technique is used to group related data in feature space. Then, it is visually investigated if related data in feature space are also related when being perceived by human. The scheme for a specific grain texture definition is proposed and three models of grain textures are experimentally created. The first model involves especially significant grain corners and edges, the second model involves homogeneous parts of a grain, the third model can be used for coarse and abraded surface recognition.

**Keywords**—Texture, sedimentary grains, co-occurrence matrix, *K*-means, clustering.

## I. INTRODUCTION

The goal of geomorphological research is to reveal a relief genesis of an investigated area. This research can be carried out by the methodology called exoscopy. Specifically, this process means the analysis of unlithified sedimentary grains. The grains are examined and typical structural features are sought on its surface. Such features can be the degree of roundness, presence of fractures, and so on [1]. When a set of grains (contained in one sample) is analyzed, the histograms of particular features are constructed and typical features for a given genesis are stated. Because these grains are of small sizes, electron microscope is utilized for magnifying and capturing them, see Fig. 1. After zoomed images are obtained, they are manually analyzed by experienced expert. This procedure is very time consuming. Because the analysis is done by using images, it could be possible to incorporate image processing techniques into the analysis to shorten the analysis time or to bring some new information suitable for genesis evaluation.

The objective of this paper is to define typical structure forms, which can be recognized on the surface of sedimentary grains by using a computer. As stated in the previous paragraph, the information about a grain surface character, described by typical structure forms, can be used to build statistics about a set of grains (one sample) and these statistics can be used for genesis estimation in the exoscopy analysis. The purpose of computer usage is to ease the routine work of experts. Typical structure forms, which are

sought on the grain surface, have been already defined by geomorphologists. However, the implementation of a procedure which would be able to extract the defined structures is not straightforward due to the complex nature of grain surface. Thus, structure forms, which can be easily obtained by using computers and also which can be interpreted by human expert need to be found.

The co-occurrence features and a clustering technique are used in this task. The computer analysis of grains is conducted from the perspective of texture analysis for which the co-occurrence features are widely used. The reason is that the grain surface structures exhibit some degree of randomness and generally, texture can be seen as a mixture of typical patterns with some degree of random variability [2]. Then, the *K*-means clustering technique is used for investigating the feature space presented by the co-occurrence features. Individual clusters, which represent a given texture type in feature space, are examined if they also represent a particular texture type when visually perceived by human. There are papers dealing with texture and clustering utilization [3][4]. However, they are motivated by using the texture analysis for segmentation. Moreover, these methods are evaluated on the images from general synthetic testing databases. In this study, the main point is to define grain structure forms, which can be described by co-occurrence features.

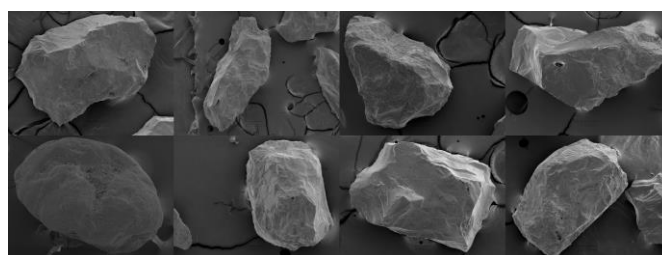


Fig. 1. Examples of grain images acquired by electron microscopy

## II. PRINCIPLES

### A. Co-occurrence matrix

Co-occurrence matrix [5] represents spatial relations between values of pixels in an image. Consider images  $p_1$  and  $p_2$ , where  $p_2$  is created by shifting  $p_1$  with distance  $(x,y)$  in Cartesian coordinates. These images are overlapped and number of pixel pairs with values  $(i,j)$ , where  $i$  is value in  $p_1$  and  $j$  in  $p_2$ , is written on the position of  $i$ -th row and  $j$ -th column in the co-occurrence matrix  $P(i,j)$ .  $P(i,j)$  is thus constructed for a given parameters  $(x,y)$ . When  $P(i,j)$  is

Manuscript received March 28, 2013.

A. Křupka is with Department of Telecommunications, Brno University of Technology, 612 00 Brno, Czech republic  
e-mail: akrupka@phd.feec.vutbr.cz

normalized by number of pairs, joint distribution  $p(i,j)$  of pair values is obtained. The co-occurrence matrix is square and its size is equal to number of possible intensity values in images  $p_1$  and  $p_2$ . The size of co-occurrence matrix thus can be reduced by stronger quantization of image. The co-occurrence matrix is still very raw representation of texture, thus different measures are used to describe texture [6]. The mostly used measures are the following

$$\text{contrast} = \sum_{i,j} |i-j|^2 p(i,j) \quad (1)$$

$$\text{correlation} = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i\sigma_j} \quad (2)$$

$$\text{energy} = \sum_{i,j} p(i,j)^2 \quad (3)$$

$$\text{homogeneity} = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \quad (4)$$

where the range of  $i$  and  $j$  is given by size of  $p(i,j)$ ,  $\mu_i, \mu_j$  are marginal means of  $p(i,j)$  distribution,  $\sigma_i, \sigma_j$  are marginal standard deviations of  $p(i,j)$  distribution. When  $p_1$  is a subpart of some bigger image, then the features can be considered as local. This way, the texture of image subpart is described. By changing the parameter  $(x,y)$ , different features for a given sub-window can be extracted using previous formulas so a feature vector for the sub-window is obtained. Because no prior information about a texture type of the extracted sub-windows is known, no labels belong to the extracted feature vectors and thus, the desired texture models cannot be formed using some supervised machine learning algorithm. Moreover, it is desirable to evaluate the possibilities of co-occurrence features obtained from the sedimentary grains. Therefore, the key idea is to investigate the feature space of co-occurrence features if it is possible to specify some of feature space subparts, which present some kind of texture. This way, typical structure forms of sedimentary grains could be defined and it could be then distinguished between them. For this reason, clustering can be used to inspect if similarity in data space corresponds to texture similarity perceived by human.

### B. K-means

The well-known  $K$ -means algorithm of clustering can be used [7]. This algorithm allows to group  $D$ -dimensional data consisting of  $N$  samples  $x_n$  into clusters according to their inter-samples distances, where  $n = 1, \dots, N$ . The goal is to construct centroids  $\{\mu_k\}$ , where  $k = 1, \dots, K$ , such that centroid  $\mu_k$  belongs to  $k$ -th cluster. The data point  $x_n$  is then assigned to the cluster, whose centroid is in the smallest distance from  $x_n$ . Euclidean metric  $\|x_n - \mu_k\|^2$  is usually used as the distance. During the procedure of searching for optimal centroids  $\mu_k$ , an optimization objective is given by

$$J = \sum_{n,k} r_{nk} \|x_n - \mu_k\|^2$$

where  $r_{nk} = 1$ , if data sample  $x_n$  is assigned to  $k$ -th cluster, otherwise  $r_{nk} = 0$ . Thus, the term  $J$  represents the sum of the squares of the distances between sample and its assigned cluster with centroid  $\mu_k$ . To minimize  $J$ , sets of the  $\{r_{nk}\}$  and the  $\{\mu_k\}$  need to be found. This is done by an iterative optimization procedure where each iteration consists of two steps. In the first step, the minimization is done with respect to  $\{r_{nk}\}$ , which means to assign data samples  $x_n$  to its closest cluster centroids  $\mu_k$ . In the second step, the minimization is done with respect to  $\{\mu_k\}$ , the value of  $\mu_k$  is computed as the

mean of data samples assigned to the  $k$ -th cluster. This optimization is repeated until convergence. Usually the initialization of the set  $\{\mu_k\}$  is done by assignment of randomly selected data samples to the cluster centroids  $\mu_k$ . The optimization procedure does not guarantee to reach the global optimum of  $J$ , so to bring up the chance of ending up in the best local optimum, the iterative optimization can be repeated multiple times with different random initialization of the centroids  $\mu_k$ , after that, the model with the lowest  $J$  is selected.

## III. METHOD

### A. Implemented procedure

This part describes the implemented procedure using previously mentioned principles. A primitive used for the texture evaluation is a square sub-window taken from an image. The square sub-windows are sequentially extracted from the image with a given step. Co-occurrence matrices for several offsets  $(x,y)$  are then constructed from the sub-windows, and measures (1)(2)(3)(4) are computed. This way a dataset from available image set is obtained, the feature extraction process is shown in Fig. 2.

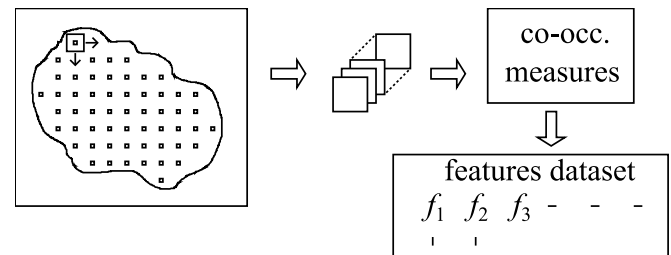


Fig. 2. Procedure of feature extraction: From the images, square subparts are obtained, which are used for construction of co-occurrence matrices. The measures of co-occurrence matrices constructed from a single sub-window form a single feature vector.

After the extraction, examination of feature data is performed. The  $K$ -means clustering technique is used to divide the dataset into parts homogeneous in feature space. Then the result of clustering is visually evaluated, see Fig. 3. From the first clustering result, visually recognizable type of texture, which seems to be consistently included in one or more clusters, is selected.

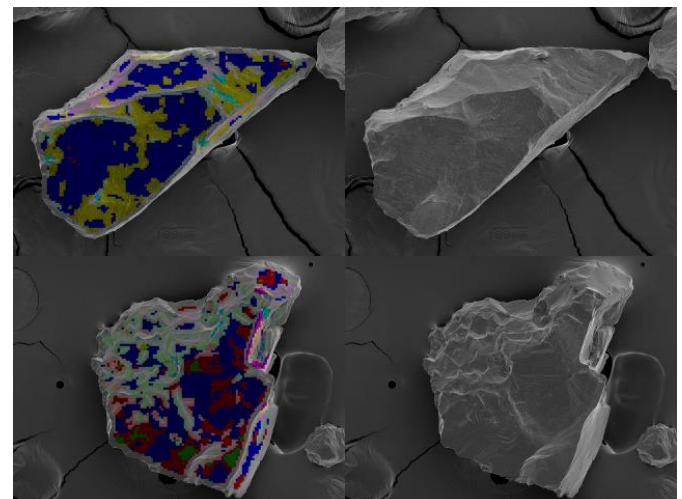


Fig. 3. Example of initial clustering for  $K = 9$ .

It is certainly possible that more than one type of texture perceived by human can be included into one cluster, therefore, the selection scheme demonstrated in Fig. 4 is applied. This selection scheme can be understood as a decision tree, where the clusters not fitting to a given texture are discarded. On the rest of data,  $K$ -means procedure is applied and the result is again visually examined. This way the subpart of feature space, which belongs to visually related textures, is determined more specifically.

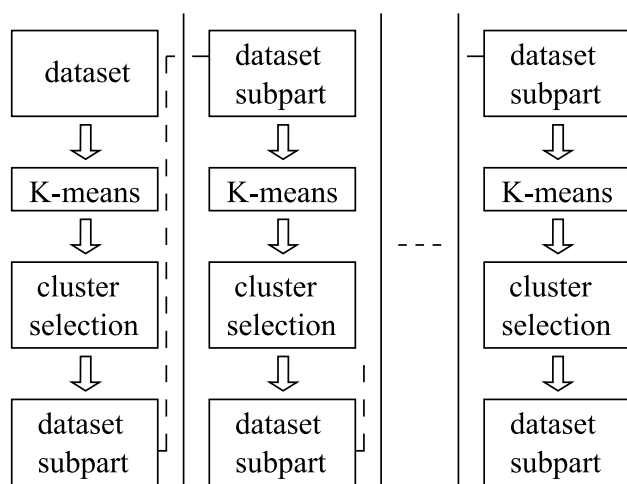


Fig. 4. Procedure for texture modeling. A model is specified by a subpart of feature space.

#### B. Experiment scheme and properties

Here follow the parameters of the experiment. The resolution of the images is 1280x960 pixels. A sedimentary grain is located in the middle of image and covers a significant part of the image. The square sub-windows for features extraction are of size 30 pixels and they are picked with horizontal and vertical step of 8 pixels. To avoid extracting sub-windows from the background parts of an image, manually prepared masks determining the area of a grain are used for controlling the extraction of the sub-windows. From every sub-window, co-occurrence matrices are extracted for parameters (1,0), (1,1), (0,1), (-1,1), (15,0), (15,15), (0,15), (-15,15), which represent shifts in directions of 0, 315, 270, 225 degrees, the opposite directions are ignored from the assumption of texture periodicity. The shift of 1 or 15 pixels in  $x$  and  $y$  directions is chosen to exhibit co-occurrences in closer spatial relations as well as in the distance of half sub-window width. The intensity values of sub-windows are uniformly quantized to 16 levels to get smaller size of co-occurrence matrices. From the prepared co-occurrence matrices, the feature vectors are computed as described in Section III a). The number of features in a vector is 32 because of 8 offsets of co-occurrence matrices multiplied by 4 measures computed from them. For getting texture data the set of 100 images is used.

The  $K$  is set to 9 to provide sufficiently fine clustering and still to allow good cluster visualization by different colors, see Fig. 3. Then the set of 100 pictures is visually examined and a type of texture, which seems to be well defined by clusters, is selected. Thus the subpart of dataset is selected for the next stage of clustering. This procedure is repeated several times to define the best possible subpart of feature space, which corresponds to visually consistent texture type.

The model created by scheme in Fig. 4 is then visually evaluated on the other independent test set of images.

## IV. EXPERIMENTAL RESULTS

### A. Created models

The experiments were conducted according to description in the previous part. During the experiments, three texture models using co-occurrence features were created.

The first model locates the sharp and contrast edges as can be seen in Fig. 5. The color marks located in different parts of the image represent the centers of sub-windows, whose content meets the given texture model.

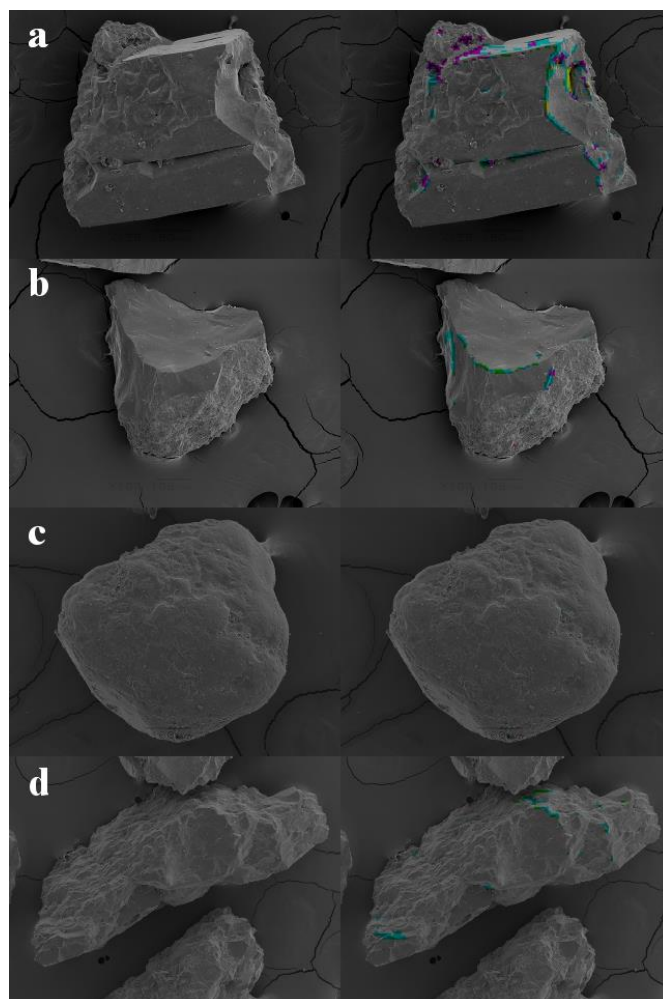


Fig. 5. Example of visualization using the first model: Left side presents the original image; right side highlights the sub-windows, whose texture belongs to the first model.

The second model was designed to address especially homogeneous part of the grain surface. This way especially smooth and plain parts of grain are detected which is visualized on samples in Fig. 6.

The third model is aimed to texture of rough and variable grain surface, which is illustrated in Fig. 7.

### B. Discussion

As can be seen, the clustering of dataset is being used for defining typical structural pattern contained in an examined image subpart. According to initial clustering results, three texture models were defined as subparts of the whole feature

space using procedure in Fig. 4. In the right part of Fig. 5, Fig. 6 and Fig. 7, the positions of sub-windows whose feature vectors belong to the subpart of feature space defined by a model are highlighted by color spots. Different colors stand for different clusters defined in the last stage of procedure shown in Fig. 4, however, all of these different clusters belong to one defined model.

The first constructed model defines the feature space subpart, which includes especially the parts of grain surface containing significant rapid intensity changes, see Fig. 5. The corners connecting two well visible plains are the most probable to be involved by the model, see Fig. 5a, b. On the other hand, well rounded grains do not contain a lot of these corners and thus a small number of sub-windows belongs to the first model, see Fig. 5c. A grain with coarse surface can miss well distinct corners, therefore, a small number of color spots can be seen in Fig. 5d.

The second model is designed for inclusion typically homogeneous parts of texture. Fig. 6a,b demonstrate highlighted homogeneous plains. Also parts of a well rounded grain without coarse structure are captured by the second model, see Fig. 6c. Conversely, a grain with rough structure has a minimum of homogeneous parts which is also reflected in small number of highlighted sub-windows in Fig. 6d.

The third model is aimed to coarse structure generated e.g. by surface abrasion, which is clearly visible in Fig. 7 where the sub-windows containing rough texture are highlighted.

The distinctive corners as well as smooth plains on the grain surface are omitted by the third model.

The classified sub-windows according to created models can be used for percentage computation of a given texture occurrence on the grain surface. This way, statistics can be evaluated by geomorphologists and these results can be included to the conclusions stated about particular geomorphological genesis. This classification scheme can be also used as one particular step in possible multi-stage grain processing. As can be noticed in the presented figures, some sub-windows can be highlighted by a model in spite of their visual non-similarity to the model, which can be caused by intersection of visually different textures in feature space. However, the experiments showed that typical textures are densely filled with correctly classified sub-windows, thus areas with a high concentration of highlighted sub-windows can be considered as areas of texture given by the used model. This offers e.g. to determine grain parts for further specific processing dependent on texture. The sub-windows highlighted by the first model are typically positioned into thin lines so they represent the corners of the grain. If these lines would be properly extracted by further processing, the roundness of a grain could be evaluated not only from the shape borders of 2D projection (for which some methods have been done) but also from the presence of corners inside the grain.

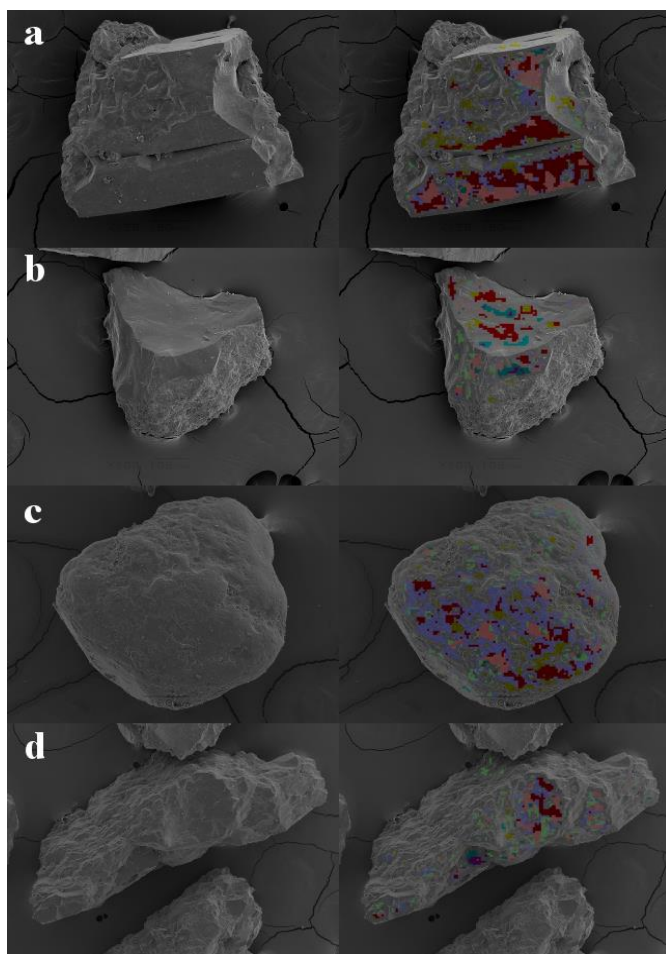


Fig. 6. Example of visualization using the second model: Left side presents the original image; right side highlights the sub-windows, whose texture belongs to the second model.

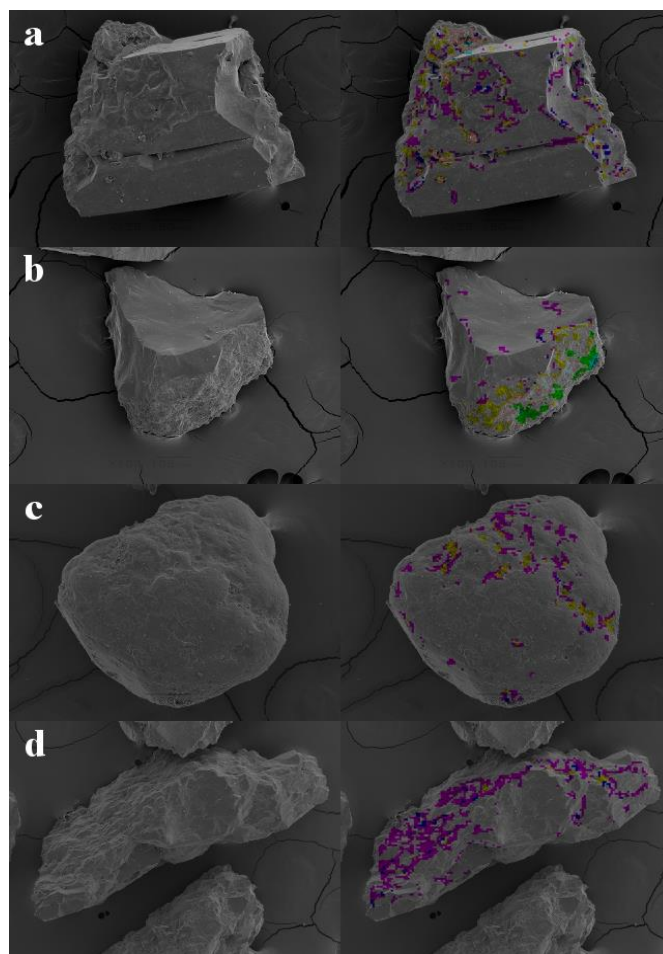


Fig. 7. Example of visualization using the third model: Left side presents the original image; right side highlights the sub-windows, whose texture belongs to the third model.

## V. CONCLUSION

The main purpose of this work was to find typical structure patterns of sedimentary grains, which can be described by co-occurrence features. The possibilities of the main co-occurrence features were explored using *K*-means clustering technique instead of a prior definition of textures and utilization of some supervised machine learning techniques. The reason was that the structure forms defined by geomorphologists cannot be easily extracted from images using computers because of their complex nature. Thus the feature space of co-occurrence features was examined to find typical subspaces representing visually consistent texture, for which *K*-means algorithm was used. By visual inspection of clustering result the typical classes were defined. The separate models were constructed for detection of sharp edges, homogeneous surface and rough surface. The sharp edges positions can be further processed to locate the corners or sharp lines. The parts with frequent detections of homogeneous sub-windows can be considered as smooth plain of surface without significant changes. Frequent detections of rough surface can determine the extensively abraded part of grain. The degree of presence of these three defined structure forms can be then used as input data for the exoscopic analysis.

The future work will be aimed at utilization of extracted texture as a base for more specific texture classification and also for an evaluation of statistical occurrence of these computer extractable surface structures in different geomorphological geneses.

## REFERENCES

- [1] L. Křížová, M. Křížek, L. Lisá, Applicability of quartz grains surface analysis to the study of the genesis of unlithified sediments, *Geografie*, vol. 116, no. 1, 2011, pp. 59–78.
- [2] M. Mirmehdi, X. Xie, J. Suri, Handbook of texture analysis, Imperial College Press, 2008, pp. 1-6.
- [3] D. A. Clausi, *K*-means Iterative Fisher (KIF) unsupervised clustering algorithm applied to image texture segmentation, *Pattern Recognition*, vol. 35, no. 9, 2002, pp. 1959-1972.
- [4] Ch. Li, R. Chiao, Multiresolution genetic clustering algorithm for texture segmentation, *Image and Vision Computing*, vol. 21, no. 11, 2003, pp. 955-966.
- [5] R. M. Haralick, K. Shanmugam, I. Dinstein, Textural Features for Image Classification, *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, 1973, pp.610-621.
- [6] R. Jain, R. Kasturi, B. G. Schunk, *Machine Vision*, McGraw-Hill, 1995, pp. 236-238.
- [7] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006, pp. 423-430.

**Aleš Křupka** is currently a student in Ph.D. programme Teleinformatics at Faculty of Electrical Engineering, Brno University of Technology. He is especially interested in image processing and machine learning. His research is aimed to discovering new methods of machine analysis of sedimentary grain images.