

The Data Exploration System for Image Processing Based on Server-Side Operations

Magdalena Ładniak, Adam Piórkowski, and Mariusz Młynarczuk

Department of Geoinformatics and Applied Computer Science,
AGH University of Science and Technology,
al. Mickiewicza 30, 30-059 Cracow, Poland
mladniak@geol.agh.edu.pl,
{pioro,mlynar}@agh.edu.pl
<http://www.geoinf.agh.edu.pl>

Abstract. In this paper the possibilities for construction of an ad hoc search system to examine large-sized raster image data sets, e.g. rock images or medical images, for analysis of its characteristic parameters are presented. A new solution for image exploration based on any attributes extracted with computer image analysis by using extensions for server-side operations is proposed.

Keywords: image mining, image processing, databases, user defined functions, image clusterization.

1 Introduction to Image Exploration

Most of scientific disciplines belonging to natural sciences are based on experimental data. Although theories, obviously stimulate the development of new branches, but in doubtful cases, experimental verification is of pivotal importance [1]. In recent years an increasing need for supporting measurements of rock structures with image analysis methods is observed. These methods are being successfully implemented into researches related to geology [2], mining [3] and rock mechanics [4]. Image data exploration is a current issue. This includes a number of approaches to image mining [6, 7]. However, many authors often present no solutions enabling computations and restrict themselves to theoretical deliberations. A classification scheme for medical images has been proposed in [8], but this solution requires image processing beyond a database that may pose a problem when integrating with such data exploration systems as Weka or Statistica.

The aim of this paper is to consider a possibility for construction of a system for mining a large-sized set of rock microsections by using database-side processing. Such solution avoids a complex construction of heterogenic environment. The weak points of the existing solutions are strong dependence on interfacing with various versions of different software that can make uniform application impossible, especially when the software is updated. The use of external programs for image processing might cause difficulties with file arrangement, description and ordering in the case of large amount of photos. That is why the Authors propose

a system that provides data-base side raster image storage and analysis, thus allowing data exploration with the data mining software commonly used. Research on preliminary image clusterization by the describing parameters is proposed.

2 An Approach to Image Selection Using Server-Side Operations

The literature, which is the Authors known, on this subject is limited to the description of the image processing only on the local computer and using images in the computer local memory or imported from databases. The Authors think that the future of image analysis involves the calculation of the total transfer of database servers. However, desktop computers (in the form of desktops, laptops, tablets, smartphones etc.) will be used only to manage the calculations and to review the obtained results.

2.1 The Idea of Data Processing and Analysis by Using Database Technologies and Statistical Tools

Database queries can be processed both with a database-side engine and from the statistical tool level. At baseline the database contains only raw images with a basic description like for example their names. An effect of such query is a set of numerical data and its further analysis depends only on analyst needs. Based on observations he/she can request the database to make an additional measurement for a selected feature and to store the result permanently or temporarily in the database. Such solution enables easy and direct extraction of desired information without the use of surplus software. Information flow and schematic context of analysis for the presented example is presented below (Figure 1). The detailed description of the structure and performance issues of the proposed system are available in the literature [9].

The use of the mechanism of User Defined Functions in a database environment gives the user complete freedom in creating his own image processing operations and advanced algorithms. These algorithms were created by the Authors for the purpose of this study. Absolute freedom in designing a database schema should be emphasized here. The stored images can be taken out by DMBS at the user's request, processed and the results can be placed into a result table. Any number of operations can be involved each time in processing. This may be either a single operation that builds a complex algorithm or a set of cyclic image processing operations. Any number of input images can be processed. In addition, it is possible to write images being partial solutions in the database. The mechanisms used allows easy reference also to these images what is of utmost importance to the creation process and verification of proper operation of new image processing algorithms. The application of the proposed solution enables construction of knowledge base of very large number of parameters. This can be used in the next step in statistical data analysis.

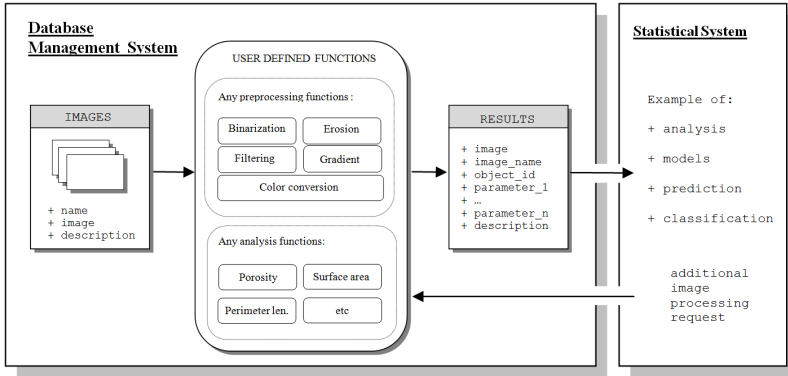


Fig. 1. Information flow and schematic context of analysis

There are many tools for the whole process of converting data into usable knowledge. In study described here Weka (Waikato Environment for Knowledge Analysis) was used. This is an open source software under General Public License. The possibility of direct communication with a database system and graphic presentation of data were utilized in this software. Thanks to provided communication with a database the user's request for image processing is sent directly by doing SQL (Structured Query Language) queries. In practice this indicates a capability to run image analysis routines directly from a statistical tool when the user sees a potential need for an additional analysis. This approach is well known in methods of gaining knowledge from a sets of numerical data, for example in time series analysis [10].

2.2 Combination of Search Criteria

It is often necessary to make a selection based on many criteria concurrently. An example is finding the images for which a group of parameters fulfills the specified standard conditions. The manual solution of this task induces a series of operations to be executed connected with switching between various solutions, e.g. 'get from the database, batch convert, save results to a file, import to statistical tool'. The Authors propose to move such operations onto database platforms, thus allowing uniform combination of search criteria. This consists of constructing appropriate conditions to narrow and group the query result. The solution unifies the data storage method, avoid problems related to entering various access paths to images, systemizes places where addition information on computed parameters are stored and allows return to results at the selected stage of analysis.

2.3 Input Data

The input data in the system under consideration are images of rock specimen microsections made by the Authors. The process of test material preparation

consisted of taking a rock specimen, cutting it into thin discs, impregnating with an adhesive with pigment addition, polishing, taking a photo with a digital camera integrated with a microscope at 100x magnification [11]. The impregnation process contributed in gaining clear color marking of pore spaces, its shape and size [12]. This fact was the basis for selection of segmentation algorithm based on color analysis.

2.4 Preprocessing

Image processing is a well known dynamically developing discipline, so its basic issues were omitted here, while referencing to the existing publications [13–18]. In computer image analysis the purpose of transformations is most often to gain a properly segmented binary image. To obtain such image a series of transformations of the input image is carried out. As a result a contrast imaging is obtained where objects (e.g. potential pore cross sections) are clearly distinguishable from the background [14].

Threshold Binarization. To prepare a binary image two segmentation algorithms were tested by using threshold binarization:

- In the first algorithm, the thresholds were chosen empirically based on color analysis and the YIQ color space model was employed. As an input signal represent the chrominance information (I channel) was used.
- In the second algorithm, the thresholds were chosen by maximum entropy method [16].

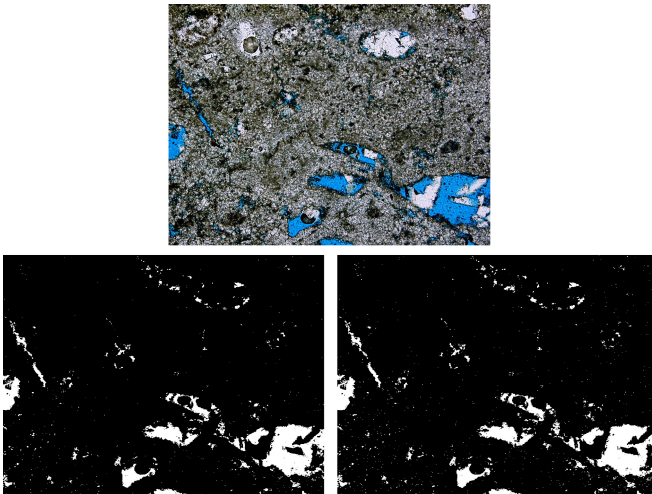


Fig. 2. Sample results of binarization by using methods 1 and 2

For further analyses the results of binarization with manually chosen thresholds were used because of better results of the method.

Noise Reduction. The next stage of preprocessing was to remove the smallest objects that did not indicate any pore voids. The feature of microscopic measurements is that they are made at a specified magnification. On this basis the object elimination criterion was established. The isolated small areas (here of few pixels in size) were removed. The boundaries of the marked areas were smoothed. To do it the erosion and dilatation operations were used. When these operations are used separately they cause different changes in surface area of objects under transformation. Erosion decreases it, while dilatation increases it. To eliminate this effect the transformations being a superposition of the previous ones, namely opening and closing were used. Opening results in removal of small objects without significant object size changes. The aim of closing is to fill in narrow nicks and small holes inside the object, while having no significant effect on its shape [19]. The set of parameters used in analyses is presented below.

2.5 Search Parameters

The starting point for making automatic measurements is proper segmentation. In the case of binary images an object should be understood as a set of pixels of brightness 1 connected to each other. Despite the fact that measuring methods are quite well defined, its limitations should be always taken into account. To convert the values of some parameters into nominative values it is necessary to do image calibration, i.e. entering information about its actual dimensions. For reasons of clarity this process was omitted in this paper [20].

- Surface area - area was calculated by summing all pixels within the object.
- Coefficient of porosity - the value of this coefficient was assumed to be the ratio of pore volume to the total volume of tested specimen expressed in percent [21].
- Number of objects - the counting process for such objects was carried out by using the labeling algorithm. This consists of viewing the binarized image line by line until a point belonging to the object is encountered. Then it is assigned a label and the values of preceding pixels are analyzed. The labeled vertices are based on associativity to neighbor values. Depending on neighborhood type it is possible to analyze both higher or left neighbor (4-connected neighborhood) or three higher and one left neighbor (8-connected neighborhood) [22].
- Average object chord length - it is possible to determine the average chord lengths at vertical and horizontal. To do it the method consisting of computing the average length of the intersection between the segmented image and the regular set of horizontal/vertical lines was used (refinement was defined symmetrically).

- Specific surface area - specific surface area is understood as the surface area of the skeleton of porous medium per the total volume of this medium [23]. The knowledge of object chord lengths allows specific surface area to be computed from the following formula known from the literature [25]:

$$S = 4 * \frac{n}{N} \quad (1)$$

where:

n – number of objects (chords) being superposition of the image and the regular grid of vertical/horizontal lines

N – length of all lines belonging to the regular grid

- Object perimeter length - to determine approximate length of object perimeters in the photographs the erosion subject image was subtracted from the original one, and then summing of the marked pixels was done.

2.6 Analysis Results

The described measuring methodology was implemented in detection of potential pore cross sections in rock images. There following rocks were selected: dolomite from Buszewo, dolomite from Baszyna, dolomite from Koscian and limestone from Santok. The photographs were taken under optical microscope equipped with a specialized CCD camera at magnification 100x [9]. The set search criterion was strictly defined values of parameters describing the material shown below in photographs. Sample result of clasterization according to criteria: origin place (Buszewo, Baszyna, Koscian, Santok) and specific surface area (Figure 3), origin place (Buszewo, Baszyna, Koscian, Santok) and number of objects Figure 4).

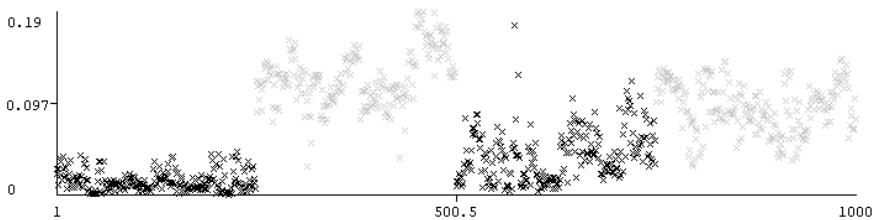


Fig. 3. Sample result of graphical clasterization, X axis - origin place (from the left in sequence: Santok (1-250), Koscian (251-500), Baszyna (501-750), Buszewo (751-1000)), Y axis - specific surface area

Efficiency. The efficiency tests were carried out on the set of 1000 photographs at the standard size of 1024x768 pixels. The Microsoft SQL Server 2012 engine was used based on the performed preminaly tests [9]. For programming the following technologies were used: Transact-SQL, and SQL CLR (CLR - Common Language Runtime) to create procedures that then to be used by the

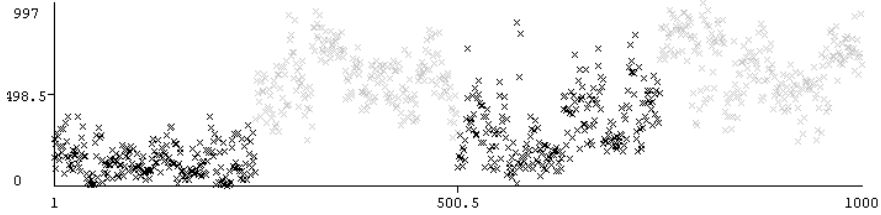


Fig. 4. Sample result of graphical clasterization, X axis - origin place (from the left in sequence: Santok (1-250), Koscian (251-500), Baszyna (501-750), Buszewo (751-1000)), Y axis - number of objects

Table 1. Function execution time for selected user defined functions

Operations:		Times [ms]					
		TSQL		CLR			
		1	2	1		2	
		s	s	s	p	s	p
preprocessing	Erosion/Dilatation	33167	25741	156	75	64	33
	Opening/Closing	65967	34331	304	160	125	53
	Logical: AND, OR	7873	2590	12	7	5	2
	Subtracting/Summing	7827	2481	14	8	6	3
	Conversions between color spaces	9437	4321	8	5	11	6
	Threshold binarization	9388	4305	26	17	11	6
	Maximum entropy method	10451	5692	28	18	12	8
analysis	Determination of surface area	1925	975	8	6	4	2
	Determination of coefficient of porosity	1947	1036	8	5	4	2
	Determination of specific surface area	10986	4087	29	21	13	10
	Counting the number of objects	10662	6015	22	-	9	-
	Counting the chord diameter length	18302	9965	31	19	13	8

database engine as an external library [22]. The procedures were stored in The Microsoft.NET 4.5 using C# language. Two computer environments were used:

- 1. PC, HDD Seagate SATA ST3500418AS (500GB), access time 14,9 ms, CPU Intel Celeron E3200 2.40 GHz 2 Cores, 2 threads, RAM 4GB DDR2
- 2. PC, HDD Seagate 7200 RPM ST2000dl003-9VT166 (2000 GB), access time 11,9 ms, CPU Intel Core i7 2600 4 Cores, 8 threads, RAM 4GB DDR3.

The preprocessing performed by using a Transact-SQL language is an extremely time-consuming task. By adding external authorial libraries CLR to the database system and using parallel computing it was possible to speed up the research considerably, thus the system executes the tasks in an adequate time. The Parallel Class (Microsoft.NET 4.5) that provides support for parallel loops and regions was used [24].

The lead times, serial as s and parallel as p , are listed in the table below.

3 Conclusions

The tests we carried out demonstrate possibilities for construction and practical use of the system (understood as a set of user's database functions) that enables analysis to be carried out based on the results of searching by the specified parameters. The Authors' intention was to present possibilities for construction of a dedicated system to meet the needs of the specified problem related to automatic measurements of rock geological structures where parametrization plays an important role. Execution of complicated conditionally point computations in image analysis is a time-consuming task for the database server, so the Authors undertake research on timing optimization of such solutions. Execution of complicated conditionally point computations in image analysis is a time-consuming task in general [26, 27] so the Authors undertake research on timing optimization of such algorithms implemented on database servers by using more advanced methods of database systems and parallel programming. It was accepted that benefits resulting from independence from various systems, its cooperation (also in the context of versioning) are so such large that further studies are justified and should be continued.

Acknowledgement. This work was financed by the AGH - University of Science and Technology, Faculty of Geology, Geophysics and Environmental Protection as a part of statutory project.

References

1. Tadeusiewicz, R.: Data mining as a chance for relative cheap scientific discoveries obtained by looking up seemingly fully explored empirical data. *Statystyka Data Mining w Badaniach Naukowych*, StstSoft Polska, Krakow (2006)
2. Mlynarczuk, M.: Some Remarks on the Application of Image Analysis and Image Processing for the Description of the Geometrical Structures of Rock. *Physico-chemical Problems of Mineral Processing* 33, 107–116 (1999)
3. Wierzbiicki, M., Mlynarczuk, M.: Structural aspects of gas and dolomite outburst in Rudna copper mine, Poland. *International Journal of Rock Mechanics and Mining Sciences* 57, 113–118 (2012)
4. Mlynarczuk, M.: Description and classification of rock surfaces by means of laser profilometry and mathematical morphology. *International Journal of Rock Mechanics and Mining Sciences* 47(1), 138–149 (2010)
5. Lay, B.J.: Image processing software - multiple solutions for a single problem. In: *Proceedings of the 9th European Congress on Stereology and Image Analysis*, Zakopane, pp. 55–70 (2005)
6. Wang, S., Mingquan, Z., Guohua, G.: Application of fuzzy cluster analysis for medical image data mining. In: *2005 IEEE International Conference on Mechatronics and Automation*, vol. 2. IEEE (2005)
7. You, F.C., Yong, B.Z.: Elliptic Object Features Extraction and Measurement in Image Data Mining. In: *International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2009*, vol. 1. IEEE (2009)
8. Mazurkiewicz, A., Krawczyk, H.: A parallel environment for image data mining. In: *Proceedings of the International Conference on Parallel Computing in Electrical Engineering, PARELEC 2002*. IEEE (2002)

9. Ładniak, M., Piórkowski, A., Młynarczuk, M.: Structure of systems for data exploration for raster images. *Studia Informatica* 33(2B), 7–20 (2013)
10. Chuchro, M.: The structure of influent time series in wastewater treatment plants. *Environmental Engineering III A 3rd Congress of Environmental Engineering, Lublin* (2009)
11. Młynarczuk, M.: Application of image analysis and mathematical morphology to quantitative description of rock fracture surface. *Archives of Mining Sciences, Monography* 3 (2008)
12. Młynarczuk, M.: Stereological description of sedimentary rocks with methods of image analysis. *Polskie Towarzystwo Mineralogiczne, Prace Specjalne* 27 (2005)
13. Matheron, G.: *Random Sets and Integral Geometry*. Wiley, New York (1975)
14. Serra, J.: *Image Analysis and Mathematical Morphology*. Academic Press, New York (1982)
15. Serra, J., Młynarczuk M.: Morphological merging of multidimensional data. In: *Proc. STERMAV 2000, Cracow* (2000)
16. Tadeusiewicz, R., Korohoda, P.: *Komputerowa analiza i przetwarzanie obrazów. Wydawnictwo Fundacji Postępu Telekomunikacji* (1997)
17. Korkosz, M., Bielecka, M., Bielecki, A., Skomorowski, M., Wojciechowski, W., Wójtowicz, T.: Improved fuzzy entropy algorithm for X-ray pictures preprocessing. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) *ICAISC 2012, Part II. LNCS, vol. 7268*, pp. 268–275. Springer, Heidelberg (2012)
18. Bernaś, M.: Objects detection and tracking in highly congested traffic using compressed video sequences. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) *ICCVG 2012. LNCS, vol. 7594*, pp. 296–303. Springer, Heidelberg (2012)
19. Coster, M., Chermant, J.L.: *Precis d'Analyse d'Images*. Press du CNRS (1989)
20. Młynarczuk, M.: Potential applications of image analysis and mathematical morphology to stereological analysis of rock structures. *Archives of Mining Sciences* 49, 117–140 (2004)
21. Jeske, T., Przędęcki, T., Rosinski, B.: *Mechanika gruntów*. Państwowe Wydawn. Naukowe (1966)
22. Radomski, P., Jarosinski, A.: Determination of specific surface area of the granular materials in aspects of its use in selected technological processes. *Technical Transactions, Chemistry* (10) (2010)
23. Lay, B.J.: Image processing software - multiple solutions for a single problem. In: *Proceedings of the 9th European Congress on Stereology and Image Analysis, Zakopane*, pp. 55–70 (2005)
24. Toub, S.: *Patterns of parallel programming. Understanding and applying parallel patterns with the .Net Framework 4 and Visual C#*. Parallel Computing Platform, Microsoft Corporation. Version (February 16, 2010)
25. Bodziony, J., Konstankiewicz, K., Pukos, A.: Stereology as image analysis method of agricultural materials. *Int. Agrophysics* 9, 293–309 (1995)
26. Bielecki, A., Buratowski, T., Smigielski, P.: Recognition of two-dimensional representation of urban environment for autonomous flying agents. *Expert Systems with Applications* 40, 3623–3633 (2013)
27. Bielecka, M., Skomorowski, M., Bielecki, A.: Fuzzy syntactic approach to pattern recognition and scene analysis. In: *Proceedings of the 4th International Conference on Informatics in Control, Automatics and Robotics, ICINCO 2007. ICSO Intelligent Control Systems and Optimization, Robotics and Automation, vol. 1*, pp. 29–35 (2007)