Coal Images Database and Its Application for Selected Coal Grains Properties Analysis

S Iwaszenko and L Róg
Central Mining Institute, Plac Gwarków 1, 40-166 Katowice, Poland
s.iwaszenko@gig.eu

Abstract. The raw materials processing methods play important role in excavation industry. The area is nowadays a target for novel, innovative technology application. One of them is an image processing as an instrument for determination of different material properties. The paper proposes a tool – database system – dedicated for supporting images processing and computer vision methods application in the mining industry. The idea and implementation of the system is presented. The database was supplied with a set of coal images. The images were taken at a special stand, with the calibrated camera. The usefulness of system was tested by application of a pipeline of image processing algorithms. The goal was to estimate the grain size composition of coal and rock matter presented on images. Two algorithms for edges detection were tested: Sobel filter and Statistical Dominance Algorithm. After edges detection, a set of morphological operations were applied of determination of good starting markers for the watershed algorithm. The result showed that both algorithms tend to produce over segmented images. However, SDA performed noticeably better. The experiments proved the usefulness of developed database system dedicated to image processing.

1. Introduction
Coal and in general rock processing methods are one of the crucial steps in the raw material industry. The quality and classification of the solid fuel are strictly connected with its physicochemical properties. The determination of coal's properties usually involves a significant amount of human labor. Therefore, application of methods offered by information and communication technologies can help in raising the effectiveness of analysis and improve their results. The image processing and analysis is a discipline which application in mining industry becomes more and more important. The application of image processing in Run of Mine (ROM) and aggregate analysis is growing since the 1990s. The image processing was used for grain size distribution determination [1, 2], ore classification [3], coal gangue identification [4] and so on. Despite the machine vision method used, the effective and efficient way of information storage and retrieval is necessary. There is also need for preparation of rock grains image database, which can be used for development and testing of dedicated computer vision methods. Unfortunately, the images' sets used in an application of computer vision to mineral raw materials processing methods are hardly reused. One of the rear examples of such reuse is given by [5]. The reported research uses the same images and images’ preparation procedure as reported in [3]. In much other application of image processing techniques, the usage of common picture databases is usual approach. As an example of available databases the IMAGENET repository [6] or databases dedicated to emotion or face recognition. However, it seems that there is lack of such
a database in case of aggregate or crushed rocks images. The authors developed a database system for coal and coal accompanying rock images. Though the idea for application of database technology is not new – an excellent attempt was described in [7] – there is no images database directed to solid fuels. Authors believe that proposed solutions extend the idea presented in [7] and directs it towards coal mining industry. The proposition of database system is presented. The system was implemented and a feed with coal samples images. The application of the developed system to the aggregate grain sizes determination by batch images processing was presented. The goal of the work was to create a dedicated but elastic tool helping in image processing techniques application in mine industry.

2. Material and methods

2.1. General assumptions and ideas

The computer vision methods application in any discipline is involved with processing a certain number of images. The concept of application of database technology to support management of the images’ set given in [7] suggests following approach to the problem:

- Database systems are targeted in processing large amounts of data, therefore they are a natural choice as a platform for bulk machine vision application.
- The processing effort should be moved from client computers to database servers thus improving mobility and lowering requirements for end user’s equipment.
- Machine vision algorithms should be implemented in form of User Defined Functions, which repository should be stored at the database server.
- There should be possible to write images formed as a result of image processing method application back to the database.
- The results of machine vision application should be stored as a set of parameters in the database, forming the knowledge base.
- All data stored in the database system should be available for end users and for further processing.

The authors present the results obtained by application of proposed idea – the research on retrieval of information about grain sizes in copper concentrate as well as segmentation of grains in thin sections. However, there was no information about database system architecture nor information about its implementation. In particular, there was no clear statement if the same database system (though selecting an appropriate set of images) was used in both types of research. The proposed work addresses this area of database system dedicated to computer vision applications. Usefulness of such system is recognized in two areas:

- The system should be capable of batch processing of the big amount of visual information stored in the database by use of processing workflow (batch mode).
- The system should deliver the functionality for performing on demand, individual processing of selected images by use of desktop or web application (interactive mode).

The results of analysis should be stored in the same database. The presented areas are not separated. In fact, the interactive mode is a tool for prototyping and establishing the appropriate sequence of operations, which lead to achieving the desired result. Once the image processing and analyzing algorithm are established, the workflow for batch mode can be prepared. Then, the process is performed on the server side, for possibly a long time. During the process, the selected bunch of images are transformed and analyzed automatically. The results of operations are stored in the database for further investigations and usage. The schematic idea of operation is presented in figure 1.

The processing workflow, stored in the database system, can be easily reused. Not only does it allow efficient processing of any new images stored in the database, but also form the basis for further development.
2.2. Database system architecture and implementation

The proposed database system was divided into few modules, responsible for important aspects of functionality. The modules can be characterized as follows:

The data repository, responsible for storing all data and information processed and shared by the system. The data repository stores both: images and all kind of information describing the images. This includes the results of the application of image processing algorithms to images as well as results of the application of analytical tools. The data repository logs all operations and workflows which were used to process the images.

Image processing and analysis routines repository stores software components in form of object-wrapped processing procedures. The solutions were preferred to suggested in [7] User Defined Functions. The objects implement a unified interface, thus facilitating the composition of complex processing sequences. The objects can be relatively easily exchanged, according to Liskov’s substitution principle [8].

The workflow engine is responsible for the execution of the sequence of image processing or analysis operations. The sequence (workflow) itself is managed by the workflow engine. Components of the sequence are taken from image processing and analysis routines repository. Workflows can have branches. It is suggested that workflow engine is based on one of available workflow generic solutions, e.g. Windows Workflow Foundations (WWF).

Batch processing manager is responsible for orchestrating of batch processing. It should start at scheduled time image processing tasks. Each task executes a workflow for a set of selected images given as parameters.

Application Programming Interface (API) forms an interface between system users and system. API provides a set of functions with well-defined parameters, which gives an access to database system functionality. Client applications, operated by end-users, interact with the system through this component.

The overall system architecture was presented in figure 2.
The image processing database system was implemented using Microsoft tools and technologies. The Microsoft SQL Server 2017 was used as a relational database management system. Workflow engine was developed with Windows Workflow Foundation and SQL Server agent was used for batch processing management. The image processing and analysis routines were organized as a filesystem stored dynamically linked libraries, using MEF (Managed Extensibility Framework).

2.3. Coal images acquisition
The coal images were taken at Central Mining Institute in the Department of Solid Fuels Quality Assessment. The images were taken using Nikon D80 digital camera. For images acquisition, the dedicated stand was prepared. All samples taken for analysis were photographed at different camera focal length. Pictures were taken in the ambient daylight conditions. The images have a resolution of 3872x2592 pixels with 8-bit RGB color space. The camera was calibrated using chessboard template [9]. The images were stored in the database along with the sample code for further reference. There were also a linear measure placed on each photograph. The images were saved in a lossless format.

3. Results and discussion

3.1. Coal images database
During several images, acquisition sessions over 500 of images were taken and stored in the database system. The photographed samples consisted of enriched and sieved coal, coal and barren rock with mixed granularity and coal originating wastes. There were also few pictures of crushed Run-Of-Mine (ROM). Because of imaging scene character, the automatic exposure parameters selection could not be directly used. For each sample, a basic information of its origin, as well as description, were stored. For the majority of samples laboratory, fuel analysis results are also available. The typical photograph gained is presented in figure 3.
3.2. *Images processing test*

The developed solution was used as a tool for investigations on selected grain sizes detection algorithms effectiveness. The overall procedure used was schematically presented in figure 4.

![Diagram of image processing sequence](image)

**Figure 4.** Image processing sequence.

The selected images stored in the database were cropped to ensure that there are no other objects than coal or rocks visible. The resulting pictures had the sizes of 3872x2096. Then the images were transformed to HSV color space and only the V (Value) channel, representing image brightness was taken for further processing. The median filter with square structuring element and sizes ranging from 3x3 to 11x11 were applied to each monochromatic picture. The resulting images were the basis for edges identification. Two algorithms were used for this purpose: the Sobel filter and Statistical Dominance Algorithm. The Sobel filter is one of the commonly known methods of edges detection. It exploits the observation that edges are usually the places of abrupt changes in pixels brightness. Therefore, the length of gradient vector can be treated as an edge indicator [10].

\[
|\nabla I(x,y)| = \sqrt{\left(\frac{\partial I(x,y)}{\partial x}\right)^2 + \left(\frac{\partial I(x,y)}{\partial y}\right)^2} \tag{1}
\]
As long as gradient magnitude is greater than the selected threshold, the points are treated as belonging to the edge. The method, though simple, is not resistant to the artifacts. There is also not always clear which value of thresholding should be used.

SDA is one of the recently proposed algorithms for edges and structures detection. The algorithm focuses on differences in grey level between the selected point and its neighborhood. The number of points exceeding the selected point value is the result value of the operation. There is also possible to take into account only this points from the vicinity, which values are greater of selected point value increased by given threshold. The formal description of the algorithm is presented in the equation below [11].

\[
\bigwedge_{p(x,y) \in I} \sum_{p_{b}(x_{b},y_{b}) \in B(x,y)} \begin{cases} 
   p'(x,y) := p'(x,y) + 1, p_{b}(x_{b},y_{b}) \geq p(x,y) + t \\
   p'(x,y) := p'(x,y), otherwise
\end{cases}
\]

where \(p(x,y)\) is a pixel from the source image, \(p_{b}(x_{b},y_{b})\) and is a pixel belonging to neighborhood \(B(x,y)\) of point \(p(x,y)\). The \(p'(x,y)\) is resulting pixel and \(t\) stands for the grey level threshold. The resulting image is normalized and then, as in the case of Sobel filter, a threshold is applied. The threshold level was selected individually for each image. For that purpose selected, representative set of images were processed manually. The edges of aggregate grains were drawn by trained personnel. After that, the average edges’ area ratio was calculated yielding 9.3%. As the Sobel/SDA detected edges are fragmented it was assumed, that about half of the edges are missing. Therefore, the threshold level was chosen to ensure that no more than 4.5% of all pixels in given image will be selected. The thresholded images are processed by dilation algorithm to remove tiny elements. Again, the process was continued until at least 10% of initial thresholded areas were eroded. Then a watershed algorithms are applied, using remaining areas as seed. For the resulting segmentation, a histogram of grain sizes for each picture is calculated. The histograms obtained by Sobel filter edge detection method as well as SDA algorithm were compared.

3.3. Tests results
The exemplary result of proposed processing for the same image once using the Sobel filter and once SDA algorithm is presented in figure 5 and figure 6 respectively. The histograms for presented image are presented in the figure 7 and figure 8.
Figure 5. Results of coal image processing for Sobel filter (selected image). a – original image, b – image processed by 11x11 median filter, c – the normalized output of Sobel filter, d – watershed segmentation result.

Figure 6. Results of coal image processing for SDA filter (selected image). a – original image, b – image processed by 11x11 median filter, c – the normalized output of Sobel filter, d – watershed segmentation result.
It can be observed, that despite the method used the watershed algorithms tend to over-segment the image. The resulting distribution is therefore shifted towards lowest values in comparison to expert determined segmentation. The SDA algorithm seemed to perform slightly better, resulting in lesser over-segmentation. Despite that, SDA resulted in segmenting grains of extremely large areas. It suggests a modification to the algorithm used: after an object exceeding threshold size is detected, it is processed again for better segmentation. Nevertheless, the testing showed that developed database of images allows for relatively easy bulk processing of a big number of images. Though the processing times were relatively large, the server-oriented processing gives the possibility to use server resources for calculation. There is also possible to schedule the processing to off hours and collect the results in suitable time.

4. Conclusions
The coal images database was developed for coal images storage and processing. The database was supplied with coal and barren rock images taken during laboratory samples acquisition. The images were stored with information allowing their matching with measured coal properties. There are above 700 pictures of coal with different granularity and properties stored in the database. A set of image processing operations were performed showing the usefulness of the designed database. The images were filtered for noise removal by the median filter, then the grain edges were detected using Sobel or SDA filter. After that, the grains were segmented using the watershed algorithm. Though obtained results showed significant over-segmentation, they proved that the developed system can be efficiently
used for batch images processing. The further development of the system is planned as well as research aimed at obtained precise image segmentation methods. The further research will also consider other features of coal grains as color (in various color spaces), texture, grain shape properties and so on. It is expected that enhanced features set will allow not only shape and size of rock grains determination but also coal type recognition.

Acknowledgments
The works presented in the paper have been supported by the statutory activity of the Central Mining Institute: Application of image recognition and machine learning methods for Run-Of-Mine and waste mineral matter characterization – No. GIG: 11010117-144.

References
[1] Maerz N 1998 Aggregate sizing and shape determination using digital image processing Center For Aggregates Research (ICAR) Sixth Annual Symposium Proceedings pp 195–203
[2] Persson A L 1998 Engineering Geology 50 177 – 186 ISSN 0013-7952 URL http://www.sciencedirect.com/science/article/pii/S001379529800009X
[3] Tessier J, Duchesne C and Bartolacci G 2007 Minerals Engineering 20 1129 – 1144 ISSN 0892-6875 URL http://www.sciencedirect.com/science/article/pii/S0892687507001161
[4] Wang R and Liang Z 2011 Automatic separation system of coal gangue based on dsp and digital image processing 2011 Symposium on Photonics and Optoelectronics (SOPO) (IEEE) pp 1–3
[5] Perez C A, Estévez P A, Vera P A, Castillo L E, Aravena C M, Schulz D A and Medina L E 2011 International Journal of Mineral Processing 101 28 – 36 ISSN 0301-7516 URL http://www.sciencedirect.com/science/article/pii/S0301751611001037
[6] 2019 URL http://www.image-net.org/
[7] Mlynarczuk M, Ladiak M and Piorkowski A 2014 Physicochemical Problems of Mineral Processing 50(2) 563–573
[8] Liskov B and Wing J M 1993 Family values: A behavioral notion of subtyping Tech. rep. Carnegie-Mellon Univ Pittsburgh PA Dept of Computer Science
[9] Zhang Z 1999 Flexible camera calibration by viewing a plane from unknown orientations Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on 1 (IEEE) pp 666–673
[10] Szeliski R 2010 Computer vision: algorithms and applications (Springer Science & Business Media)
[11] Piorkowski A 2016 A statistical dominance algorithm for edge detection and segmentation of medical images (Springer) pp 3–14