Image histogram as a tool for a coal stream homogeneity evaluation

A Heyduk¹, J Joostberens¹, A Nowrot¹, J Pielot¹ and K Fabiańczyk²

¹Department of Electrical Engineering and Automation in Industry, Faculty of Mining and Geology, Silesian Technical University, Akademicka 2, 44-100, Gliwice, Poland

²Polscargo International Ltd, Szczecin, Poland.

aheyduk@polsl.pl

Abstract. The paper presents selected results of the preliminary laboratory research on the application of histogram analysis to the problem of coal stream homogeneity evaluation. This is an important problem of industrial coal mine output stream quality testing. Image acquisition is fast and cheap, but needs complex analysis techniques. Histogram analysis can be considered to be one of them. There has been proposed a distance measure, and there have been presented analysis results which proved an ability to detect gangue, rock, schist, and even wooden block contamination of the homogenous coal stream. The method needs further improvement and comprehensive analysis of other available options of the RGB-to-grayscale conversion and of histogram comparison to make it more selective. These further research directions have also been described in the paper.

1. Introduction

Coal quality monitoring is an important part of the whole chain of power generation and utilization. It is important from commercial and ecological point of view. Stable coal quality is very important for the proper operation of coal-fired power plants in order to ensure stable combustion conditions and low level of pollutants emission [1]. Continuous quality testing meters (like ash-monitors, heat-quantity meters, sulphur meters) are expensive. This fact (in addition to legal regulations requiring expensive and time-consuming laboratory testing of samples taken from the coal stream) hinders their wide application and inspires the search for new relatively low-cost methods of coal quality assessment, easily integrable with an existing coal quality testing framework, based on discrete-time and limited-size sample laboratory tests. Image processing can be considered as one of this methods because one of the crucial factors is the number of samples necessary to analyze in laboratory conditions. When the coal stream is homogenous, the sample size can be reduced, therefore decreasing the cost and time of laboratory analyzes. This coal stream homogeneity is currently analyzed locally by human inspectors but the video monitoring with automatic image processing and analysis can further reduce costs and time of this analysis simultaneously improving the result (due to limited human visual perception capabilities). Visual assessment of the coal quality can be performed at various detail levels both in the time domain and in the spatial domain. In the spatial domain the most detailed analysis level may include the thorough evaluation of specific coal particles. These coal particle contours must be precisely separated - that is, the entire image must be subjected to a complex and time-consuming image segmentation process [2]. The most general analysis level in the space domain may include assessment of the whole image (or
selected image fragment) as a single entity in either colour/intensity or in texture domain. This is the method analyzed in this paper. Resolution in time domain - regardless of the selected image analysis method - can vary from several random images of the whole train (homogeneity of the whole train coal load) to several images for each railway wagon (homogeneity monitoring of each wagon content). These two extreme possibilities are presented in figure 1.

Figure 1. Two possible applications of the proposed method: a) analysis of the content homogeneity of a larger number of wagons throughout the whole train; b) detailed analysis of the coal stream homogeneity throughout one wagon.

In order to ensure stable lighting conditions, images necessary for further analysis should be recorded not directly in the train or railway wagon, but on the path of the belt conveyor transporting the coal to the loading point in a single place with a stable artificial lighting, shielded from the direct influence of the sunlight. Before implementing this concept in industrial practice, there have been conducted small-scale laboratory tests in controlled laboratory environment using selected samples of both homogenous clean coal and contaminated coal. Laboratory tests concerned two issues:

- homogeneity evaluation of the clean coal stream
- the ability to detect contaminated fragments of the coal stream

2. Image histogram acquisition

The images of coal samples have been recorded using a Logitech HD camera (2Mpix resolution). They contained several coal particles placed on a real conveyor belt. In the laboratory setup 1 pixel corresponded to approximately 1mm² of the real image area. Then –as a task was to analyze coal surface not a conveyor belt surface- only a small rectangular fragment of the image has been selected, containing only the particles and the minimum part of the background. Therefore the information value of the analyzed image has been significantly enhanced. The selection process is presented in figure 2. Particular attention has been paid to provide a uniform illumination of the coal surface, so a distributed system of several halogen bulbs has been used.

The image has been recorded using an RGB camera, so for each pixel three different colour component values R(x,y), G(x,y), B(x,y) have been recorded. They have been converted to single grayscale intensity value GS(x,y) using an equation corresponding to human colour perception [3][4]:

\[ GS(x,y) = 0,299 \cdot R(x,y) + 0,587 \cdot G(x,y) + 0,114 \cdot B(x,y) \] (1)
Figure 2. Image fragment selection. a) whole recorded image b) selection of the meaningful part, c) selected part – zoomed.

There can be also used simplified formulas like

$$GS(x, y) = 0.3 \cdot R(x, y) + 0.6 \cdot G(x, y) + 0.1 \cdot B(x, y)$$  \hspace{1cm} (2)

or even

$$GS(x, y) = \frac{R(x, y) + G(x, y) + B(x, y)}{3}$$  \hspace{1cm} (3)

and nonlinear (max/min) formulas like

$$GS(x, y) = \text{Max}[R(x, y), G(x, y), B(x, y)]$$  \hspace{1cm} (4)

and

$$GS(x, y) = \text{Min}[R(x, y), G(x, y), B(x, y)]$$  \hspace{1cm} (5)

or

$$GS(x, y) = \frac{\text{Max}[R(x, y), G(x, y), B(x, y)] + \text{Min}[R(x, y), G(x, y), B(x, y)]}{2}$$  \hspace{1cm} (6)

Advantages and disadvantages of these approaches need further research (particularly in the case of nonlinear formulas – because they use nonlinear Max/Min functions it can be said that the equation (4) enhances brighter parts of the image and equation (5) enhances darker parts of the image).

3. Image histogram analysis

3.1. Histogram similarity and dissimilarity measure

As the analyzed histograms contain the great number of bins (natural colour resolution of the RGB camera CCD matrix is usually 8-bit so it covers the range between 0 and 255) their comparison can be a complex task. Because they are unimodal (they have a single dominating value – corresponding to the most frequently occurring colour of coal particles surface) the modal value comparison can be one of the conditions of histogram similarity. But the modal value describes only a histogram location not its shape. The distance between two histograms must be calculated as a function of all bins. Sometimes it should also consider a different number of samples in each histogram, but in this case this condition can be omitted as the number of image pixels can be established as constant, or histogram bins values can be rescaled to their relative value as a part of the total number of pixels in the whole analyzed image area. As a distance measure for two histograms $h_1[n]$ and $h_2[n]$ there has been proposed a normalized sum:
\[ D(h_1, h_2) = \frac{1}{N} \sum_{i=1}^{N} \frac{(h_1[i] - h_2[i])^2}{f_1[i]^2 + f_2[i]^2} \]  

(7)

where \( N \) is a number of histogram bins. This formula is normalized with respect to the bin number and in each bin, the distance between two histogram values is calculated relatively to the average value of both bin content. In the case of sequential analysis of the video stream, consecutive values of (7) form a numerical sequence and a rapid change in these distance values proves a change in the output coal stream qualitative parameters. There are also many other possibilities of histogram distance calculation methods, described in [5][6].

3.2. Analysis of the homogenous coal stream

Analysis of the homogenous stream has been performed using several times the same coal sample (without gangue) but with particles placed in different layouts. Sample images are presented in fig. 3. and their corresponding grayscale intensity histograms in fig. 4.

![Coal sample images representing a homogenous coal stream.](image1)

![Grayscale intensity histograms of a homogenous coal stream samples from figure 3.](image2)

On the basis of these several histograms of the homogenous coal stream samples, there has been calculated an averaged histogram. The other possibility was to calculate a reference histogram as a
median of sample histograms. These averaged and median histogram have been presented in figure 5. as thick lines on the background of thin lines of sample histograms from figure 4.

![Figure 5. An averaged and median grayscale intensity histograms on the background of sample histograms of the homogenous coal stream.](image)

This averaged histogram has been used as a reference vector for the analysis of contaminated coal stream samples, presented in the following subsections.

3.3. Analysis of the coal stream contaminated with rock/gangue

After analysis of the homogenous coal stream samples, there has been performed an analysis of coal samples contaminated with schist and pyrite particles (figure 6). Their histograms show greater share of the lighter bins, as the surface colour of these gangue particles surface is lighter than the colour of coal particle surface.

![Figure 6. Grayscale images of the coal sample contaminated with schist (left) and pyrite (right).](image)
3.4. Analysis of coal stream with wooden blocks
Accidentally in the coal stream there can also occur some other pollutions—like wooden blocks. They can be also detected by grayscale histogram analysis, as the wood is lighter than coal, but probably in this case even better results can be obtained using analysis in single colour channels. More general method can be based on calculation of hue value for each pixel, as they also form a one-dimensional histogram.

3.5. Influence of the particle size on the coal stream image histogram
Shape of the grayscale intensity histogram is a function of not only the surface colour (connected to mineralogical properties of the particles) but also of the surface shape, connected to the particle shape and size. So there has been carried out an analysis of fine coal image (figure 9). It can be seen that the modal value of the histogram is approximately the same (as the surface colour of finer and coarser coal particles is the same) but there is a significantly greater share of darker histogram bins. This fact can be explained by the much more frequent dark inter-particle spaces in the case of finer particles.
3.6. Statistical signal processing

In cases where the averaged histogram for clean coal is unknown, it would be useful to apply a chi-square distance and statistical homogeneity test [7][8], where all the recorded $N_b$ histograms (each with $N_b$ bins) form a $N_h \times N_b$ size contingency table with

$$D_F = (N_h - 1) \cdot (N_b - 1)$$

(8)

degrees of freedom. The chi-square statistics $Q$ is calculated as a weighted sum (across all the contingency table elements) of squared differences between recorded bin counts $f$ and an expected reference histogram $e$

$$Q = \sum_{i=1}^{N_h} \sum_{j=1}^{N_b} \frac{(f_i[j] - e_i[j])^2}{e_i[j]}$$

(9)

where a reference histogram $e_i$ is calculated for each $i$-th row (in general case e.g. if they have different image sizes and different number of pixels) on the basis of the vectored column-wise total sum of the contingency table as:

$$e_i[j] = e[j] \sum_{k=1}^{N_b} f_i[k]$$

(10)

where

$$e[j] = \frac{\sum_{k=1}^{N_h} f[k][j]}{\sum_{k=1}^{N_h} \sum_{i=1}^{N_b} f[k][i]}$$

(11)

Then the homogeneity assessment simplifies to hypothesis testing (null hypothesis $H_0$ that all the image histograms come from the same grayscale intensity distribution, against the alternative hypothesis $H_A$) if the calculated for the whole histogram set chi-square statistics is greater than the value of chi-square distributed random variable with $D_F$ degrees of freedom the null hypothesis $H_0$ should be rejected and the stream probably needs more precise investigation (e.g. more frequent sampling and laboratory sample analysis).

4. Conclusions

Preliminary experiments carried out on the laboratory stand have proved that an image histogram can be a possible tool for evaluating the coal stream homogeneity. There are significant differences in the case of a coal stream contaminated with gangue, or other, e.g. wooden elements. A more reliable analysis requires a much larger number and size of samples, possibly in environmental conditions closer to industrial (e.g. on moving conveyor belt). There will be also tested different formulas for RGB to...
grayscale intensity conversion (e.g. enhancing lighter or darker parts of the image) and different formulas of histogram distance calculation. Image grayscale histogram analysis cannot be directly compared with an analysis based on sample sieving, because its main application purpose is not a granulometric analysis, but only a stability assessment of a coal blend qualitative parameters. Influence of coal particles edge density (closely related to the particle size) on the histogram shape is only a side effect, secondary to the primary purpose of the coal stream monitoring system.

References
[1] Luo Y Wang Y 2015 International Conference on Materials, Environmental and Biological Engineering (Paris: Atlantis Press) pp 847-50
[2] Heyduk A 2017 2-Dimensional and 3-dimensional image acquisition and processing methods in machine vision granulometric analysis (Gliwice: SUT Editions).
[3] Ware C 2004 Information Visualization: Perception for Design (Burlington: Morgan Kaufmann)
[4] ITU-R BT.709-6: 2015 Parameter values for the HDTV standards for production and international programme exchange.
[5] Cha H S 2007 Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. Int. J. of Math. Models and Meth. in App. Sci. 1 pp 300-7
[6] Bityukov S I, Maksimushkina A V and Smirnova V V 2016 Comparison of histograms in physical research. Nucl. Energy and Techn. 2 pp 108-13
[7] Greenacre M 2007 Correspondence analysis in practice (Boca Raton: Taylor & Francis)
[8] Gagunashvili N D 2010 Chi-square tests for comparing weighted histograms Nucl Instr Meth. Physics Res A 614 pp 287-96