An Image Processing System for Char Combustion Reactivity Characterisation

Chaves, Deisy ; Trucco, Emanuele; Barraza, Juan ; Trujillo, Maria P.

Published in:
Computers in Industry

DOI:
10.1016/j.compind.2018.12.014

Publication date:
2019

Document Version
Peer reviewed version

An Image Processing System for Char Combustion Reactivity Characterisation

Chaves, Deisy ; Trucco, Emanuele; Barraza, Juan ; Trujillo, Maria P.

Published in:
Computers in Industry

Publication date:
2018

Document Version
Peer reviewed version

Accepted manuscript version of the article published in Computers in Industry, final version available via DOI: 10.1016/j.compind.2018.12.014

An Image Processing System for Char Combustion Reactivity Characterisation

Deisy Chaves[a,*], Emanuele Trucco[b], Juan Barraza[c], Maria Trujillo[a]

[a] *Multimedia and Computer Vision group, School of Systems Engineering and Computing, Universidad del Valle, Ciudad Universitaria Meléndez, Cali, Colombia*
[b] *Computer Vision and Image Processing group, School of Science and Engineering (Computing), University of Dundee, Queen Mother Building, Dundee DD1 4HN, Scotland, UK*
[c] *Coal Science and Technology group, Chemical Engineering School, Universidad del Valle, Ciudad Universitaria Meléndez, Cali, Colombia*

Abstract

Coal is the most used fuel source to generate electricity by pulverised coal combustion. During this process, volatile compounds are liberated giving rise to the formation of a variety of char particles. Char particles morphology can be classified into groups reflecting different coal reactivity levels which may be used to evaluate the effect of coal on the performance of burner. Char particles morphological classification may be automatically done with benefits in terms of speed, consistency and accuracy. However, the classification performance relies on correct identification of char particles. Moreover, broken walls, created during char generation process, blurriness and low contrast are factors that make the classification task a challenging problem. In this paper, we propose a system for particle detection and particle classification into two reactive groups. Initially, a set of candidate regions, that may contain particles, is selected by combining regions and edges. Then, regions containing particles are detected using texture features and a Support Vector Machine classifier. The particle classification is done based on the International Commission for Coal Petrology criteria. Experiments using coals from two Colombian regions —Valle and Antioquia—

*Corresponding author. Address: Multimedia and Computer Vision group, School of Systems Engineering and Computing, Universidad del Valle, Ciudad Universitaria Meléndez, Calle 13 No. 100-00, Cali, Colombia.
Email address: deisy.chaves@correounivalle.edu.co (Deisy Chaves)

showed that the proposed system, in most cases, correctly detect char particles. Regarding the classification of detected particles, analysed char samples were automatically classified similarly as manual classification did. Consequently, the system is found to be a successful first approach for char combustion reactivity characterisation.

Keywords: Char coal morphology; Particle detection; Particle classification; Image processing; Candidate regions; Machine learning

---

1. Introduction

Pulverised coal combustion is the most common method in coal-fired power plants. However, there are environmental issues associated with electricity generation, such as air pollution and compatibility with local land use. Increased combustion efficiency to convert coal into electricity may reduce CO_2 emissions as well as the amount of unburned coal, and can be achieved by setting combustion parameters correctly. These parameters may be tuned based on char morphology. Chars are produced in the first stage of the coal combustion process. Char morphology corresponds to the forms of char surfaces observed through a microscope. Morphological characteristics of chars correspond to wall thickness, porosity, shape and unfused material. Char morphology can be used to estimate coal reactivity, which determines combustion efficiency and the amount of ash and carbonaceous oxides released to the environment [1, 2]. Char reactivity depends on particle size and structure changes during the combustion. There are three interacting factors: (i) the chemical reaction of oxygen with the internal surface of a particle, (ii) the extent of this surface, and (iii) the extent to which oxygen diffusion through the pores —which form the internal surface— restricts the reaction. In morphological terms, particle reactivity depends on porosity, shape and wall thickness —where porosity corresponds to the ration between area porous and area particle, shape is related to sphericity and wall thickness is a measure of char particle internal walls size.

The Combustion Working Group in Commission III of the International

Committee for Coal and Organic Petrology (ICCP) published a classification of chars [3] considering nine types, shown in Fig. 1. This classification can be summarised in two main groups "high reactive" and "low reactive". Char particles with high reactive morphology are more desirable for coal combustion. The Mineroid char type is not considered in this work because the identification requires a quantification of the ash content which is related to mineral matter. The estimation of mineral matter is based on methods such as Rietveld-based X-ray powder diffractometry that we considered out of the scope of this research as it increases cost and processing time.

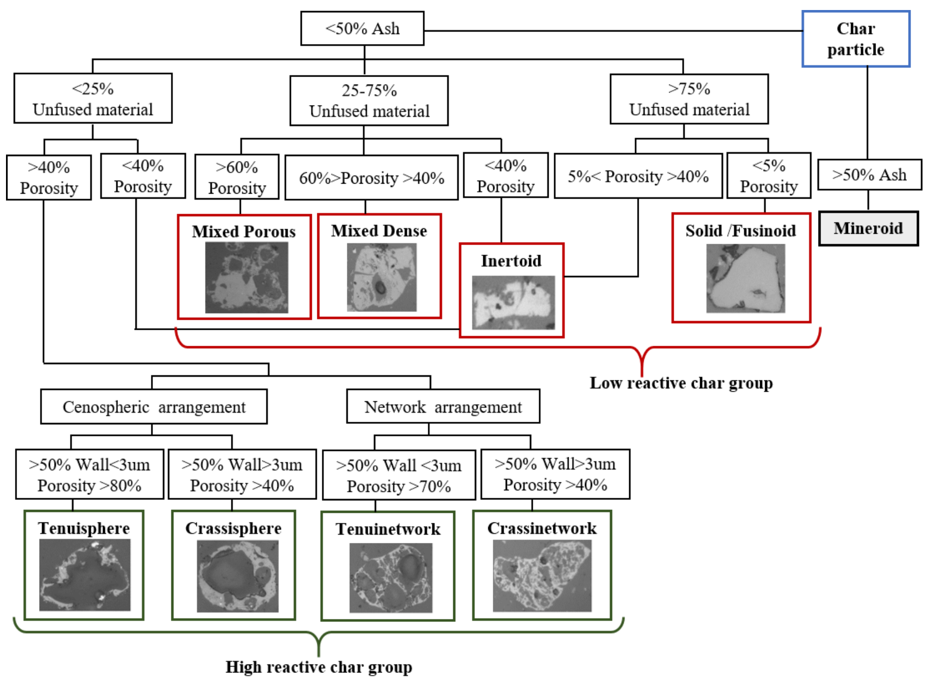


Figure 1: The ICCP classification tree for char types and char groups considered in this work.

Char morphology is commonly analysed by experts using a microscope. Before this can be done, char particles are immersed in resin to create a block. After curing the resin, the block is polished to expose char particles on the surface [4]. Then, char particles from the block surface are observed, through an air or oil microscope, counted and classified, based on morphological characteristics of the ICCP classification decision tree. Finally, frequencies per char type are used to define the reactivity of char samples. The char type with the highest frequency is adopted as an estimation of coal reactivity. An expert analysis of a char sample may be subjective, error-prone and requires a significant amount of time since it is necessary to observe and classify between 350 and 500 particles per sample, according to the recommendation of Wu *et al.* [5] in order to guarantee reproducibility of results.

However, several manual analysis in industrial processes may be done automatically by the means of image processing and using new technologies. In coal industry, systems for characterisation allowed to automate the estimation of coal quality based on texture and colour information considering three fixed categories —best, good and poor— [6], the prediction of ash content of coarse coal based on texture and colour features [7], and the estimation of particle size and particle size distribution on fine coal [8]. In a similar way, automatic tools using image processing would be beneficial to increase the accuracy of char analysis and reduce processing times.

Few systems for classifying coal samples into char types, following the ICCP decision tree —where char particles are automatically detected and morphological features are automatically quantified— have been reported [9, 10, 11, 5, 12]. In those systems, a special effort is done to detect particles since the classification heavily depends on morphological features. Changes in particle structure may occur during devolatilisation and block preparation, resulting in (i) two or more particles fused, and (ii) a particle is fragmented due to fragile walls are fractured. In both cases, char particles may not be correctly identified. Consequently, morphological features are wrongly measured. Thus, misclassified particles may affect char reactivity analysis, and introduce errors in coal quality characterisation [13, 5]. Char particles are detected using binary images along with morphological operations [10, 5, 11] or edge information [12] to refine the detection of broken particles. Binary images are obtained employing commercial software, such as KS400 [5, 11], histogram-based methods such as Isodata [14, 15] and the Triangle method [16, 12]. However, those methods fail to detect particles if the particle fragments are separated by considerable width gaps.

In other application domains, such as common object recognition, many detection and classification systems adopt a sliding window approach [17]. Small regions of variable size (windows) are swept over an image and feature vectors are used to represent the content of image regions. Feature vectors are used to train a classifier in order to distinguish windows containing target objects from others [18]. Recently, approaches using smaller numbers of windows (proposal regions) than full grids have been proposed to limit the number of candidate regions by varying parameters and criteria. Some methods use different strategies for identifying proposal regions, such as: colour contrast, edge information and superposition of super-pixels [19, 20]; number of edges within a region [21]; and hierarchical segmentation to identify initial regions using a similarity measure based on colour, texture and overlap areas for region merging [22]. High detection rates using at least 1000 proposal per image have been reported [22, 21]. However, most of those regions significantly overlap, lead to difficulties in selecting proposals containing particles without duplications. Deep learning [23] has also been used to identify and classify proposals [24, 25, 26]; first to refine results of proposal region methods [24, 25] and lately to identify and classify proposals [26, 27]. However, deep learning approaches require a large amount of training data, which is scarce in our case.

In this paper, a system is proposed for particle detection and particle classification into two reactive groups. Particle detection is performed using a three-step process. First, a set of candidate regions is selected by combining regions and edges. This method generates a small amount of overlapped and well-located regions. Second, regions are represented using texture features to discriminate regions that contain particles. Third, regions are classified as "particle" or "non-particle" using texture features and a Support Vector Machine (SVM) [28] classifier. Finally, detected particles are classified into reactive groups following the ICCP decision tree, as a way to characterise the reactivity of an analysed char sample.

An experimental evaluation was conducted to validate the proposed system for char combustion reactivity characterisation using coals from two Colombian regions: Valle and Antioquia. Results shown that the proposed system accurately identifies char particles. Concerning char particle classification, the automatic classification of a char sample agrees with the manual classification. Thus, coal reactivity characterisation, by the proposed system, is a way for setting combustion parameters in a power plant, since high reactive coals require lower temperatures and residence times than low reactive coals.

2. Materials and methods

The coal characterisation is useful for setting combustion parameters in power plants. Depending on the coal characteristics, temperature and residence time may be optimised and pollution emissions can be reduced. Coal characteristics are revealed either by laboratory analysis (ultimate and proximate analysis) or by char morphology classification. We propose a system to characterise char reactivity based on char classification, which is composed of three main processes: (i) Char image acquisition from a char-block using a microscope with an attached camera (Section 2.1). (ii) Particle detection based on candidate regions, which are classified into "particle" and "non-particle" using SVM and texture descriptors, that are presented in Table 1, (Section 2.2). (iii) Classification of char particles into high or low reactive using the ICCP decision tree and morphological features. The morphological features are calculated using image-processing techniques (Section 2.3). Fig. 2 shows the modules and data flow in our system.

2.1. Char images acquisition

Cross-section images are acquired from a char-block surface using the two-step process described next (Fig. 2a-b).

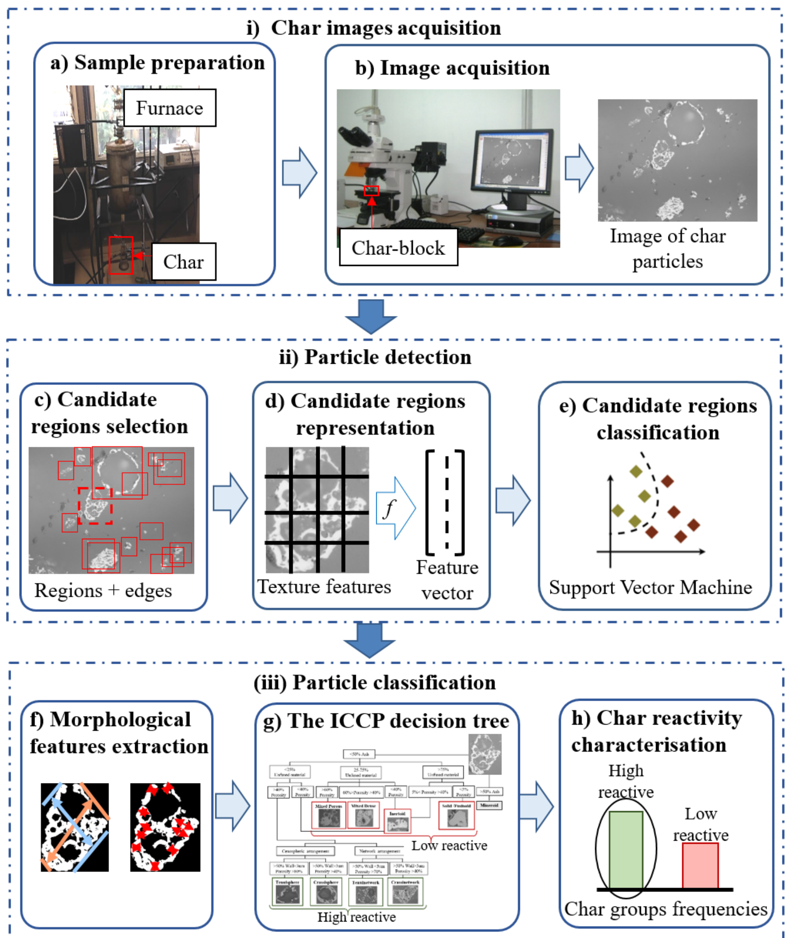


Figure 2: Main modules and data flow of the proposed char reactivity characterisation system. (i) Char images acquisition: (a) Sample preparation, (b) Image acquisition. (ii) Particle detection: (c) Selecting candidate regions, (d) Representing candidate regions using texture features, (e) Classifying candidate regions into "particle" and "non-particle". (iii) Particle classification: (f) Calculating morphological feature, (g) Classifying char particles into high or low reactive using the ICCP decision tree. (h) Characterising char combustion reactivity.

2.1.1. Sample preparation

Coals samples from two Colombian regions are used to generate char particles: Valle (South West) and Antioquia (Central West), see Fig. 3. The proximate and the ultimate analysis are presented in Table 2. The ultimate analysis

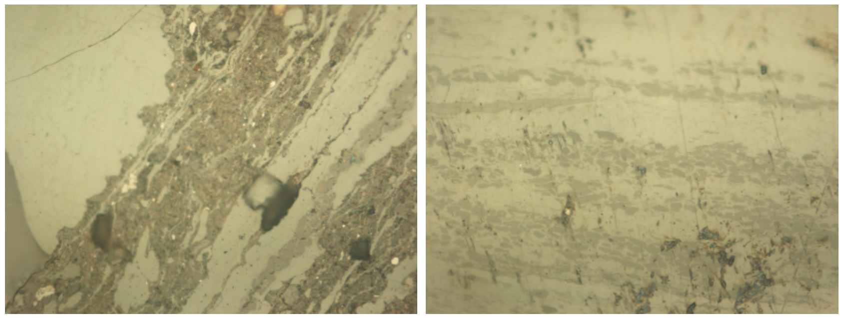
Table 1: Summary of texture descriptor abbreviations.

Symbol	Definition
LBP	Local Binary Pattern
HGM	Histogram of Gradient Magnitudes
HOG	Histogram of Oriented Gradients
ASM	Angular Second Moment
IDM	Inverse Difference Moment or Homogeneity
$SumAvg$	Sum Average
$SumEnt$	Sum of Entropy
$Cont$	Contrast
$Corr$	Correlation
Ent	Entropy
$DEnt$	Difference Entropy
$IMC1$	Information Measures of Correlation 1
$IMC2$	Information Measures of Correlation 2

is a chemical approach to characterise coals by determining the amounts of the principal chemical elements in a sample. The proximate analysis is a way to determine the thermal energy released when coal is burned and predict how coals will behave when handled and burned. In this work, the proximate analysis is performed using the standard ASTM D5142-9 [29].

In particular, Valle and Antioquia coals are bituminous with a high volatile content, as can be observed in Table 2. This kind of coals ignites easily and burns well to generate electricity in coal-fired power plants. However, if burned improperly it can produce excessive air pollution when, for instance, the operating conditions are not optimised.

Coal samples are milled using a milling ball equipment to particle sizes of -75μm and are used to produce chars in a drop tube furnace (Fig. 4a). At this size, gravity has minimal influence on particles, leading to better fluidisation inside a reactor [30, 31]. Coal samples and a nitrogen-oxygen mixture fed a



(a) Coal from Valle (b) Coal from Antioquia

Figure 3: Pulverised coal samples ($-30\mu m$).

Table 2: Proximate and ultimate analysis of coal samples.

Coal sample	Valle	Antioquia
Proximate Analysis (p/p.%, dry free)		
Ash	36.05	13.62
Volatile matter	28.87	48.05
Fixed carbon (by difference)	35.09	38.33
Higher Heating Value (BTU/lb)	7727	9488
Ultimate Analysis(p/p.%, dry ash free)		
Carbon	72.19	72.22
Hydrogen	5.45	5.14
Nitrogen	1.16	1.44
Sulphur	4.87	0.85
Oxygen (by difference)	16.32	20.34

furnace where devolatilisation took place. An amount of 1% v/v oxygen is used for facilitating tar oxidation and avoiding char particle condensation. Coal particle residence time, in the furnace, is 200 ms, at 900 °C with 10^4 °C/s heating velocity.

2.1.2. Image acquisition

Char-blocks are built using char, resin and liquid hardener and are polished with fine polishing clothes using suspensions of alumina at 0.5, 0.3 and 0.05 microns. A set of 200 char images of 1600×1200 pixels —that contain 1784 char particles— is acquired using a metallurgical microscope Eclipse LVD 100 Nikon at 50x magnification lens (Fig. 4b). This magnification corresponds to a scale of 0.8 μm/pixels.

A ground truth is built using a set of 1784 char particles —that are manually annotated by experts indicating particle location, for detection evaluation, and reactive group, for classification evaluation— in order to evaluate the performance of the automatic system (Fig. 4b).

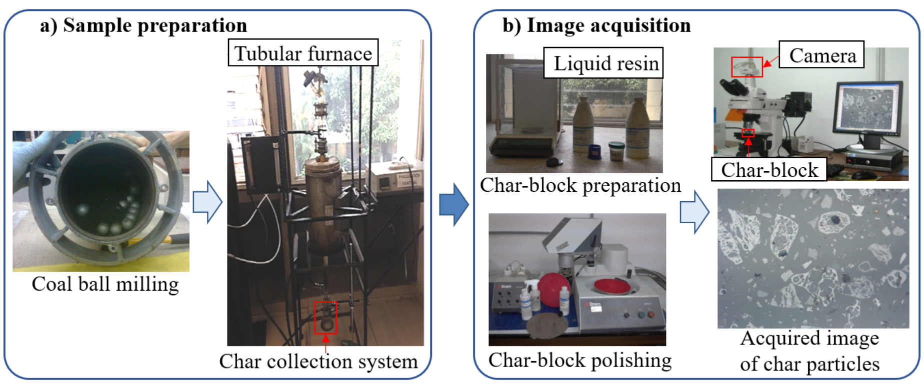


Figure 4: Char images acquisition process. (a) Preparation of char sample in a drop furnace. (b) Acquisition of char images from a polished char-block surface.

2.2. Particle detection

Although, human experts can easily identify particles, automatic identification becomes a difficult task due to factors such as (i) broken particle walls caused by changes in char particle structure during combustion, (ii) unfocused image regions because of poor char-block polishing during blocks preparation and (iii) low contrast between the resin (background) and char particles. Particle detection is one of the main stages during an automatic char sample classification since it may affect the accuracy of char morphology estimation. For particle detection, we used the three-step process described next (Fig. 2c-e).

2.2.1. Candidate regions selection

Candidate regions are selected by a four-fold process (Fig. 5). First, given a grey scale image, I, the Triangle method [16] is used to convert it into a binary image, in which '1' indicates particles. This method assumes a bimodal distribution, which char images satisfy, as shown in Fig. 5b. Two peaks can be observed in the histogram, the most prominent peak corresponds to the background pixels, and the smallest peak corresponds to the particle pixels. Briefly, the Triangle method draws a line between the maximum value of the histogram and the lowest value larger than zero. The threshold is set to the value that maximises the distance between the histogram and the line. Second, the Sobel operator is applied to I in order to obtain edges. Third, images from the two previous steps are combined, using a set union operation, to refine candidate regions. Fourth, connected white regions with an area less than 1000 pixels are discarded since they may correspond to isolated fragments, which are not of interest in the analysis, according to expert's criteria. The threshold value, 1000 pixels, was experimental tune. The flood-fill algorithm [32] is used to obtain connected region.

Finally, a set of bounding boxes $R = \{r_1, r_2, ..., r_i, ..., r_n\}$ is generated around connected regions as they may correspond to locations of char particles (candidate regions).

2.2.2. Candidate regions representation

Once candidate regions, potentially containing particles, are generated, the content can be described using several methods [33]. We adopted texture features, Haralick [34], LBP [35], HGM, and HOG [36] methods to represent the content of candidate regions (Table 1). The selected features may reveal complex patterns —such as brightness, slope and size, among others—, which discriminate regions containing particles. In particular, given a candidate region r_i,

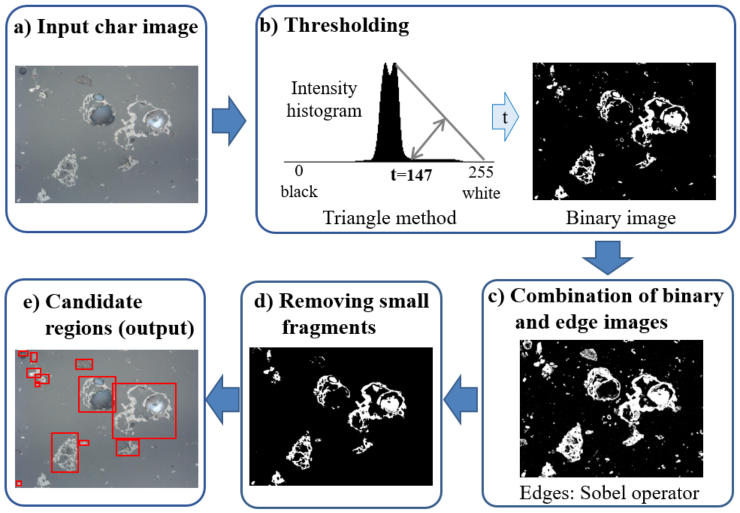


Figure 5: Illustration of candidate regions selection from a particles image.

a feature vector based on texture information is obtained —as illustrated in Fig. 6— to represent a region.

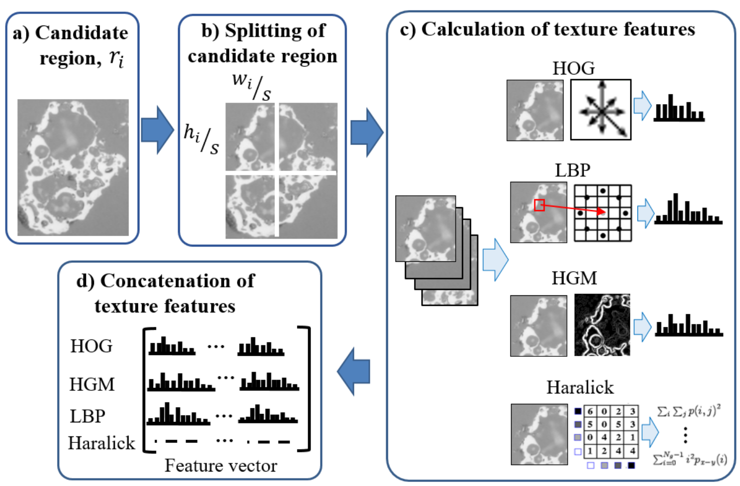


Figure 6: Illustration of texture features extraction from a candidate region.

Firstly, r_i is split into s^2 patches of size $\frac{w_i}{s} \times \frac{h_i}{s}$, where $s > 0$, and w_i and h_i correspond to the width and the height of a proposal region r_i. The value of the splitting parameter s is tuned experimentally.

Secondly, texture and edge features are obtained per patch. Texture is captured by Haralick and LBP. Haralick features [34] provide a set of metrics calculated over co-occurrence matrices representing spatial relations among pixels values, in order to identify patterns. We used a step size of one pixel at four angles of $0°$, $45°$, $90°$ and $135°$. The final Haralick value corresponds to the average over all angles. LBP [35] searches for binary texture patterns at each pixel considering a circular region, then a histogram is calculated to summarise the LBP values. LBP is computed using a radius of two and considering eight neighbours. Edge information is quantified by calculating HGM and HOG. HGM is obtained using the Sobel operator to generate a histogram of gradient magnitudes. HOG [36] creates histograms by counting frequencies of gradient angles; in this work, eight angles are considered.

Finally, a feature vector for a candidate region is formed by concatenating the LBP, HGM, HOG and the Haralick values obtained per patch. The concatenation of features allows to get a richer representation of a candidate region.

2.2.3. Candidate regions classification

Texture feature vectors —composed by Haralick, LBP, HGM, and HOG— represent candidate regions and are used to classify candidate regions into "particle" and "non-particle". Notice that the "non-particle" class includes regions with partial and multiple particles. In this work we used the Support Vector Machine (SVM) classifier [28]. SVM learnt a classification model from class annotated feature vectors, by choosing the best kernel transformation that linearly separates new examples. Intuitively, a good separation between classes is achieved by a kernel transformation that has the largest margin to a decision boundary. In general, the larger the margin, the lower the generalisation error of a classifier. The trade-off between maximising the margin distance and minimising the training error controlled by a regularisation parameter C. Regions classified as "particle" are used next to determine char reactivity.

2.3. Particle classification

Char particles are classified into two reactivity groups —high and low— using the three-step process described next (Fig. 2f-h).

2.3.1. Morphological features extraction

Morphological features are required to classify char particles and the calculation of those features heavily depends on the correct char particle detection. Four morphological features are used in the ICCP decision tree to describe a particle content —percentage of unfused material, porosity, sphericity, wall thickness— along with, auxiliary variables —such as area and number of pores— that are used in an intermediate step. Features and auxiliary variables are computed as follows [12].

1. Area: the number of white pixels in a binary image obtained by the Triangle method (Fig. 7b).
2. Percentage of unfused material: the ratio between unfused material and particle area. The unfused material corresponds to the brightest grey intensities in char images —values between 250 and 255— (Fig. 7c).
3. Number of pores: the number of voids identified in a char particle (Fig. 7d).
4. Porosity: the ratio between pores area or voids and particle area.
5. Sphericity: the ratio between the minimum and the maximum Feret diameters.The minimum and maximum Feret diameter correspond respectively to the shortest and the longest distance between any two parallel tangents at a particle (Fig. 7e).
6. Wall thickness: The second quartile (the median) of wall thickness distribution is used as the wall thickness measure. Wall thickness distribution is calculated using line transects in three steps. First, a particle image is converted into binary. Second, lines transects are drew from the image centre in all directions. At each line, the distance of two intersected points at the particle edges is computed as a measure of thickness. Third, the histogram of wall thickness is calculated (Fig. 7f).

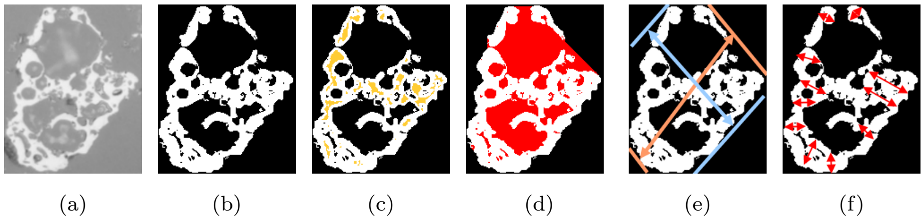


Figure 7: Illustration of morphological char features. (a) Char particle image in gray scale; (b) Area of particle, in white; (c) Unfused material, in yellow; (d) Pores identified in red; (e) The maximum and minimum Feret diameters; (f) Line transects used for calculating wall thickness.

2.3.2. *The ICCP decision tree*

Particle classification is performed by the ICCP decision tree [3], which was specially built for classifying char morphology based on experts knowledge (Fig. 1). The classifier uses morphological features calculated previously. In particular, particles are classified into one of the two groups, "high reactive" or "low reactive". High reactive group corresponds to morphologies of thin-walled, high porosity and large superficial area, such as Crassisphere, Teniusphere, Tenuinetwork and Crassisnetwork. Low reactive group refers to morphologies of thick-walled, low porosity and small superficial area such as Mixed Porous, Mixed Dense, Solid and Inertoid.

2.3.3. *Char reactivity characterisation*

Particles in a char sample are characterised by relative frequencies of the two reactivity groups: high and low. The char group with the highest frequency indicates the reactivity of a char sample.

3. Experimental results and discussion

In our study, a set of 200 images, that contain 1784 char particles, are used to evaluate particles and particles classification (Section 2). The program to analyse char particles was developed using Python and C++ programming languages.

Training conditions of SVM classifiers, which are used to detect particles, along with the performance and the evaluation of particles classification are described in the following sections.

3.1. Feature selection for SVM classifiers in particle detection

SVM classification models were built using 80% of the dataset, 160 images that contain 1476 particles, which are annotated as particles to form a ground truth. Since there is not negative examples, we introduce the Intersection over the Union of the objects (IoU_{obj}) to generate negative examples. IoU_{obj} in a region is used to determine whether a candidate region r_i matches with a particle from the ground truth, gt_i. The expression for IoU_{obj} is defined as follows:

$$IoU_{obj}(r_i, gt_i) = \frac{area_{obj}(r_i \cap gt_i)}{area_{obj}(r_i \cup gt_i)}, \tag{1}$$

where $area_{obj}$ is area of objects in a candidate region. Area is calculated using a convex hull [37].

The training set was composed by 1476 examples of the class "particle" and 1628 examples of the class "non-particle". Candidate regions with $IoU_{obj} \leq 0.5$ were considered as "non-particle" examples.

SVM classifiers were learnt using a linear kernel with a regularisation parameter $C = 1$. A 5-fold cross-validation with a sub-sampling strategy was used to optimise the model parameters since the training sets were imbalanced (more examples in "non-particle" class examples than in "particle" class). In the cross-validation strategy, the training dataset was randomly split into five subsets of equal size; one subset was used for evaluation and the remaining ones for training the model. This process was repeated five times (the folds), i.e. a subset was used for evaluation. The model is selected as the one with the highest accuracy (Acc) measure during the cross-validation. Acc is the percentage of candidate regions correctly classified as "particle" and "non-particle".

Thirteen features were evaluated for describing proposal regions: HGM, HOG, LBP and the ten Haralick features: ASM, Cont, Corr, IDM, SumAvg, SumEnt, Ent, DEnt, IMC1 and IMC2. In addition, splitting factor values,

$s = \{1, 2, 3, 4\}$, were considered for extracting features, as was described in Section 2.2.2. The number of bins, h, to construct a histogram was chosen based on the following rules:

- The Freedman-Diaconis rule, $h = 2\frac{IQR}{\sqrt[3]{m}}$,
- The Scott rule, $h = 3.5\frac{STD}{\sqrt[3]{m}}$,
- Fix bin sizes $h = 20$, $h = 50$ and $h = 100$,

where IQR is the interquartile range, STD is the standard deviation and m is the amount of data. Outlier values were not considered during the construction of histograms. The Tukey method [38] was used to identify the range for outlier values.

Individual texture features were chosen based on Acc values and Area Under the ROC Curve values (AUC) [39]. The AUC values corresponds to the probability that a classifier ranked a randomly chosen "particle" example higher than a "non-particle" one, which indicates how well a feature can distinguish among classes. Table 3 and Table 4 present the Acc and the AUC values obtained using LBP, HGM and HOG features. Feature histograms were calculated with a whole training set using five rules to determine the number of bins. The classification performance measures, the Acc and the AUC, do not exhibit significant differences regarding the different number of bins. Regarding the splitting factor to obtain feature vectors, the best accuracy values were yielded for a factor $s = 4$. It is observed that describe proposal regions using local patches ($s > 1$) instead of considering a whole region ($s = 1$) led to a better particles detection since a better representation is obtained.

Table 5 and Table 6 show the Acc and the AUC values obtained using Haralick features. Similar to the performance observed using HOG, LBP and HGM features, the Acc and the AUC computed using Haralick features increase when large values of splitting factor were used. In particular, a splitting factor of $s = 4$ was used to obtain feature vectors.

Individual features with Acc and AUC values above 70 and $s = 4$ —in Ta-

Table 3: Average Acc values calculated using HOG, LBP and HGM features by different splitting factors and methods to estimate the number of bins. The best performance values are highlighted in bold.

Feature (Rule for histogram bins)	Splitting factor, $s = 1$		Splitting factor, $s = 2$		Splitting factor, $s = 3$		Splitting factor, $s = 4$	
	Fea. len	Acc ± I.C.	Fea. len	Acc ± I.C.	Fea. len	Acc ± I.C.	Fea. len	Acc ± I.C.
LBP(Freedman)	445	0.727 ± 0.011	1780	0.762 ± 0.004	4005	0.801 ± 0.010	7120	**0.823 ±0.012**
LBP(Scott)	334	0.726 ± 0.008	1336	0.758 ± 0.023	3006	0.800 ± 0.012	5344	0.820 ± 0.010
LBP(Fixed to 100)	100	0.724 ± 0.010	400	0.755 ± 0.020	900	0.800 ± 0.013	1600	0.819 ± 0.009
LBP(Fixed to 50)	50	0.723 ± 0.009	200	0.752 ± 0.018	450	0.803 ± 0.012	800	**0.818 ± 0.012**
LBP(Fixed to 20)	20	0.701 ± 0.012	80	0.741 ± 0.014	180	0.789 ± 0.013	320	0.808 ± 0.012
HGM(Freedman)	767	0.679 ± 0.015	3068	0.716 ± 0.015	6903	0.751 ± 0.003	12272	0.774 ± 0.008
HGM(Scott)	462	0.685 ± 0.020	1848	0.721 ± 0.019	4158	0.752 ± 0.018	7392	0.771 ± 0.015
HGM(Fixed to 100)	100	0.719± 0.015	400	0.736 ± 0.015	900	0.775 ± 0.011	1600	0.790 ± 0.015
HGM(Fixed to 50)	50	0.730± 0.019	200	0.739 ± 0.014	450	0.784 ± 0.013	800	0.798 ± 0.014
HGM(Fixed to 20)	20	0.735± 0.013	80	0.739 ± 0.019	180	0.793 ± 0.014	320	**0.806 ± 0.012**
HOG	8	0.544 ± 0.029	32	0.625 ± 0.032	72	0.681 ± 0.012	128	**0.712 ± 0.013**

Table 4: Average AUC values calculated using HOG, LBP and HGM features by different splitting factors and methods to estimate the number of bins. The best performance values are highlighted in bold.

Feature (Rule for histogram bins)	Splitting factor, $s = 1$		Splitting factor, $s = 2$		Splitting factor, $s = 3$		Splitting factor, $s = 4$	
	Fea. len	AUC	Fea. len	AUC	Fea. len	AUC	Fea. len	AUC
LBP(Freedman)	445	0.833	1780	0.839	4005	0.885	7120	**0.910**
LBP(Scott)	334	0.775	1336	0.803	3006	0.875	5344	0.904
LBP(Fixed to 100)	100	0.775	400	0.803	900	0.874	1600	0.903
LBP(Fixed to 50)	50	0.770	200	0.798	450	0.872	800	0.902
LBP(Fixed to 20)	20	0.739	80	0.773	180	0.859	320	0.888
HGM(Freedman)	767	0.756	3068	0.762	6903	0.824	12272	0.840
HGM(Scott)	462	0.705	1848	0.736	4158	0.792	7392	0.820
HGM(Fixed to 100)	100	0.731	400	0.750	900	0.817	1600	0.843
HGM(Fixed to 50)	50	0.741	200	0.754	450	0.829	800	0.851
HGM(Fixed to 20)	20	0.748	80	0.757	180	0.837	320	**0.853**
HOG	8	0.421	32	0.729	72	0.764	128	**0.785**

ble 3, Table 4, Table 5 and Table 6— were chosen to select the optimal subset of features by applying a forward wrapper approach [40]. In this way, irrelevant and redundant features are removed since those features do not contribute or may decrease the performance of a classifier. The forward wrapper is described as follows: Starting from individual features, feature subsets are created by adding a new feature, as long as the new feature increases the accuracy performance of the feature subset, at each iteration. SVM classifiers were trained for

Table 5: Average Acc values calculated for Haralick features using different splitting factors. The best performance value by feature in bold.

Feature	Splitting factor, $s = 1$		Splitting factor, $s = 2$		Splitting factor, $s = 3$		Splitting factor, $s = 4$	
	Fea. len	Acc ± I.C.	Fea. len	Acc ± I.C.	Fea. len	Acc ± I.C.	Fea. len	Acc ± I.C.
ASM	1	0.501 ± 0.001	4	**0.501 ± 0.001**	9	**0.501 ± 0.001**	16	0.499 ± 0.001
Cont	1	0.499 ± 0.003	4	0.523 ± 0.017	9	0.648 ± 0.025	16	**0.684 ± 0.025**
Corr	1	0.564 ± 0.008	4	0.644 ± 0.012	9	0.693 ± 0.014	16	**0.699 ± 0.016**
IDM	1	0.649 ± 0.020	4	0.655 ± 0.019	9	0.726 ± 0.019	16	**0.746 ± 0.019**
SumAvg	1	0.501 ± 0.027	4	0.531 ± 0.014	9	0.634 ± 0.024	16	**0.659 ± 0.027**
SumEnt	1	0.582 ± 0.016	4	0.562 ± 0.016	9	0.709 ± 0.024	16	**0.731 ± 0.021**
Ent	1	0.556 ± 0.019	4	0.568 ± 0.026	9	0.729 ± 0.020	16	**0.741 ± 0.031**
DEnt	1	0.531 ± 0.030	4	0.561 ± 0.023	9	0.687 ± 0.024	16	**0.708 ± 0.022**
IMC1	1	0.615 ± 0.017	4	0.600 ± 0.031	9	0.717 ± 0.017	16	**0.769 ± 0.021**
IMC2	1	0.497 ± 0.005	4	0.487 ± 0.006	9	0.591 ± 0.018	16	**0.712 ± 0.028**

Table 6: Average AUC values calculated for Haralick features using different splitting factors. The best performance value by feature in bold.

Feature	Splitting factor, $s = 1$		Splitting factor, $s = 2$		Splitting factor, $s = 3$		Splitting factor, $s = 4$	
	Fea. len	AUC	Fea. len	AUC	Fea. len	AUC	Fea. len	AUC
ASM	1	0.498	4	**0.500**	9	0.498	16	0.273
Cont	1	0.605	4	0.500	9	0.707	16	**0.748**
Corr	1	0.602	4	0.699	9	0.737	16	**0.772**
IDM	1	0.633	4	0.639	9	0.758	16	**0.790**
SumAvg	1	0.452	4	0.544	9	0.714	16	**0.749**
SumEnt	1	0.565	4	0.530	9	0.805	16	**0.813**
Ent	1	0.525	4	0.590	9	**0.821**	16	0.815
DEnt	1	0.483	4	0.506	9	0.738	16	**0.780**
IMC1	1	0.613	4	0.604	9	0.758	16	**0.828**
IMC2	1	0.508	4	0.510	9	0.640	16	**0.769**

a subset of features using a 5-fold cross validation. Finally, the set of features chosen for representing the content of candidate regions are: LBP, computed with histograms fixed to 50 bins, HGM, obtained with histograms fixed to 20 bins, HOG, Corr, SumEnt and IMC2. The SVM classifier achieved an Acc of 0.880 ± 0.011 and an AUC of 0.955.

3.2. Evaluation of particles detection

The correct detection of char particles was evaluated using 20% of the dataset, 40 images which contained 308 particles. Candidate regions were generated by the method described in Section 2.2. We chose the recall and the precision as evaluation measures [39]. The recall is the fraction of particles correctly detected among regions identified as particles, it can be seen as the probability of detection. It is calculated using a confusion matrix as true positive divided by true positive plus false negative. The precision is the fraction of particles correctly detected and is calculated using a confusion matrix as true positive divided by true positive plus false positive. Table 7 shows the recall and the precision values obtained at at two stages: (i) using candidate regions and (ii) using detected particles. Recall and precision values were calculated considering $IoU_{obj} = \{0.9, 0.8, 0.7, 0.6, 0.5\}$ to determine if a candidate region matches a ground truth.

Recall and precision values change depending on the IoU_{obj} used to evaluate positive matches. In general, higher values of IoU_{obj} led to lower values of recall and precision, since a better localisation of candidate regions was expected. The initial set of candidate regions identified by combining regions and edges scored detection rates between 0.79 at $IoU_{obj} = 0.9$ and 0.90 at $IoU_{obj} = 0.5$. However, the precision was low (maximum precision of 0.36 at $IoU_{obj} = 0.5$) due to the large amount of candidate regions containing isolated fragments, as well as blur and scratched regions caused by a poor char-block polishing. After the SVM classifier was used to select candidate regions that contain particles, the recall yield values between 0.51 at $IoU_{obj} = 0.9$ and 0.58 at $IoU_{obj} = 0.5$ with a maximum drop of 0.32 in comparison to the recall obtained for the initial set of candidate regions. The precision values were between 0.63 at $IoU_{obj} = 0.9$ and 0.72 at $IoU_{obj} = 0.5$ with a maximum improvement of 0.31 in comparison to the initial set, showing that the SVM classifier used to refine candidate regions detection, in most cases, allows to select regions including char particles. Fig. 8 presents some particle detection results.

The experimental evaluations are conducted using a laptop with a processor Intel Core i7 and 4GB of RAM, and the implementation is done using Python, with the scikit-learn and scikit-image libraries, as programming language. Regarding the processing time, the training of a SVM model takes approximately

2.5 hours. Once trained, models are able to detect particles in less than 32 seconds per a whole image, and an individual char particle is detected in approximately 6 seconds. The processing time is calculated using the testing set —40 char images that contain 308 particles.

Table 7: Precision and recall values obtained the particle detection process.

Particle detection stage	Total of regions	Regions per image	Precision values by IoU_{obj}					Recall values by IoU_{obj}				
			0.9	0.8	0.7	0.6	0.5	0.9	0.8	0.7	0.6	0.5
Candidate regions	681	17	0.36	0.37	0.39	0.41	0.41	0.79	0.83	0.87	0.90	0.90
Particles selected by the SVM	250	6	0.63	0.64	0.68	0.71	0.72	0.51	0.52	0.55	0.58	0.58

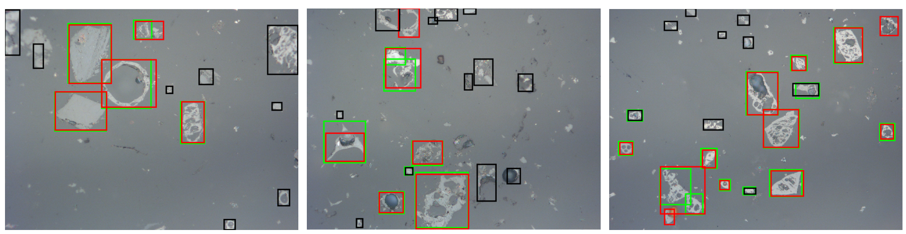


Figure 8: Illustration of "particle" and "non-particle". Particles manually annotated are in green squares (ground truth), particles automatically detected are in red squares and non-particles automatically detected are in black squares.

3.3. Evaluation of particles classification

The particle classification obtained automatically for char samples of Antioquia and Valle was compared against the manual classification. In each case, a relative frequency was calculated by group "high reactive" or "low reactive". The automatic classification was carried out over the set of candidate regions classified as "particle" using the ICCP decision tree. The evaluation was conducted on two sets of char images: (i) 160 images that contain 1476 particles, which were used to train the SVM that select the regions containing particles; and (ii) 40 char images that contain 308 particles, that were used during the testing process of the SVM. Table 8 shows the relative frequencies for the two sets of char images. In the training process, a maximum classification error of

13% compared to manual classification was observed, while in testing, the error increased to 24%. This error was a result of fragmented particles that may appear as individual particles which affects global features (e.g porosity cannot be accurately measured when walls are broken).

Despite this classification error, the group assigned to the analysed char sample corresponds to the manual classification —using the ICCP decision tree errors are admitted up to 30%, since reactivity groups are assigned to a char sample based on the mode. All in all, the system can be applied to characterise the char reactivity preliminary.

The experimental evaluations are conducted using a laptop with a processor Intel Core i7 and 4GB of RAM, and C++ as programming language. Regarding the processing time, the ICCP decision tree produces results in less than 0.5 seconds per char particle. The processing time is calculated using the testing set —40 char images that contained 308 particles.

Table 8: Char particles manually and automatically classified.

Char group	Training sets				Testing sets			
	Coal from Valle		Coal from Antioquia		Coal from Valle		Coal from Antioquia	
	Manual	Auto	Manual	Auto	Manual	Auto	Manual	Auto
High reactive	0.50	0.50	0.29	0.16	0.44	0.36	0.37	0.13
Low reactive	0.50	0.50	0.71	0.84	0.56	0.64	0.63	0.87

4. Conclusions

Pulverise coal combustion produces residuals as air pollutants —particulate matter (PM), carbon dioxide (CO_2), sulphur oxides (SOx) and nitrogen oxides (NOx)— affecting the environment and also unburned coals representing economical losses. Residuals production may be reduced by optimally setting the combustion parameter, resident time and temperature. The combustion parameters depend on coal characteristics; high reactive coals burn faster and require lower temperature than low reactive coals. In this paper, we present an automatic system for coal characterisation by classifying char particles into high and low reactive.

The automatic system is two-fold: char particle detection and char particle classification. A classifier using texture features does the former. The latter uses the ICCP decision tree. Although, char images are characterise by low contrast, ill-defined edges and lack of colour, the best set of texture features was selected using the forward wrapper method. Those features are able to represent the content of candidate regions that are used for detecting particles. The ICCP decision tree is based on morphological features, which are commonly used by experts. Those morphological features are calculated using image processing techniques and represent correctly char structures. Experimental evaluations indicate that the proposed automatic system yielded results similar to manual analysis. As future work, strategies to detect as a whole particle fragmented particles will be evaluated.

Acknowledgement

The scientific work was supported by COLCIENCIAS, Scholarship "Estudios de Doctorado en Colombia 2013 (Doctoral Studies in Colombia 2013)". Also we would like to thank to "Carvajal Pulpa y Papel SAS" for providing the coal samples, and Edward Garcia and Victor Sanabria for preparing char-blocks and acquiring images.

References

[1] A. Rojas, J. Barraza, Caracterización morfológica del carbonizado de carbones pulverizados: determinación experimental, Revista Facultad de Ingeniería Universidad de Antioquia 43 (2008) 42–58.

[2] S. Kzgut, M. Bilen, I. Toroğlu, K. Barş, Size-related evaluation of unburned carbon, Combustion Science and Technology 188 (3) (2016) 439–450.

[3] E. Lester, D. Alvarez, A. Borrego, B. Valentim, D. Flores, D. Clift, P. Rosenberg, B. Kwiecinska, R. Barranco, H. Petersen, M. Mastalerz,

K. Milenkova, C. Panaitescu, M. Marques, A. Thompson, D. Watts, S. Hanson, G. Predeanu, M. Misz, T. Wu, The procedure used to develop a coal char classification–Commission III Combustion Working Group of the International Committee for Coal and Organic Petrology, International Journal of Coal Geology 81 (4) (2010) 333–342.

[4] J. G. Bailey, A. Tate, C. F. K. Diessel, T. F. Wall, A char morphology system with applications to coal combustion, Fuel 69 (2) (1990) 225–239.

[5] T. Wu, E. Lester, M. Cloke, Advanced automated char image analysis techniques, Energy & Fuels 20 (3) (2006) 1211–1219.

[6] Alpana, S. Mohapatra, Machine learning approach for automated coal characterization using scanned electron microscopic images, Computers in Industry 75 (2016) 35–45.

[7] Z. Zhang, J. Yang, Y. Wang, D. Dou, W. Xia, Ash content prediction of coarse coal by image analysis and ga-svm, Powder Technology 268 (2014) 429–435.

[8] C. Igathinathane, U. Ulusoy, Machine vision methods based particle size distribution of ball- and gyro-milled lignite and hard coal, Powder Technology 297 (2016) 71–80.

[9] E. Lester, M. Cloke, M. Allen, Char characterization using image analysis techniques, Energy & Fuels 10 (3) (1996) 696–703.

[10] D. Alvarez, A. Borrego, R. Menéndez, Unbiased methods for the morphological description of char structures, Fuel 76 (13) (1997) 1241–1248.

[11] M. Cloke, T. Wu, R. Barranco, E. Lester, Char characterisation and its application in a coal burnout model, Fuel 82 (1517) (2003) 1989–2000.

[12] D. Chaves, E. García, M. Trujillo, J. M. Barraza, Char morphology from coal blends using images analysis, in: World Conference on Carbon, CARBON, 2013, pp. 1–5.

[13] M. Cloke, E. Lester, Characterization of coals for combustion using petrographic analysis: a review, Fuel 73 (3) (1994) 315–320.

[14] T. W. Ridler, S. Calvard, Picture thresholding using an iterative selection method, IEEE Transactions on Systems, Man, and Cybernetics 8 (8) (1978) 630–632.

[15] J. Reyes, Identificación de descriptores en imágenes de carbonizado, Master's thesis, Universidad del Valle. School of Computer and Systems Engineering (2009).

[16] G. W. Zack, W. E. Rogers, S. A. Latt, Automatic measurement of sister chromatid exchange frequency., Journal of Histochemistry & Cytochemistry 25 (7) (1977) 741–53.

[17] D. Chaves, S. Saikia, L. Fernández-Robles, E. Alegre, M. Trujillo, Una revisión sistemática de métodos para localizar automáticamente objetos en imágenes, Revista Iberoamericana de Automática e Informática Industrial RIAI 15 (3) (2018) 231–242.

[18] P. Viola, M. J. Jones, Robust real-time face detection, International Journal of Computer Vision 57 (2) (2004) 137–154.

[19] B. Alexe, T. Deselaers, V. Ferrari, What is an object?, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 73–80.

[20] B. Alexe, T. Deselaers, V. Ferrari, Measuring the objectness of image windows, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (11) (2012) 2189–2202.

[21] C. L. Zitnick, P. Dollár, Edge boxes: Locating object proposals from edges, in: ECCV, European Conference on Computer Vision, 2014, pp. 391–405.

[22] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, A. W. M. Smeulders, Selective search for object recognition, International Journal of Computer Vision 104 (2) (2013) 154–171.

[23] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[24] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '14, 2014, pp. 580–587.

[25] R. Girshick, Fast R-CNN, in: 2015 IEEE International Conference on Computer Vision, ICCV, 2015, pp. 1440–1448.

[26] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (6) (2017) 1137–1149.

[27] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, SSD: Single shot multibox detector, in: Computer Vision – ECCV 2016, 2016, pp. 21–37.

[28] B. E. Boser, I. M. Guyon, V. N. Vapnik, A training algorithm for optimal margin classifiers, in: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT '92, 1992, pp. 144–152.

[29] ASTM D5142-09, Standard test methods for proximate analysis of the analysis sample of coal and coke by instrumental procedures (withdrawn 2010), Standard, ASTM International, West Conshohocken, PA (2009).

[30] A. Rojas, Estudio cinético de la combustión del "char" de carbón pulverizado, Ph.D. thesis, Universidad del Valle. School of Chemical Engineering (2005).

[31] D. Vargas, D. Chaves, M. Trujillo, J. Piñeres, J. Barraza, Beneficiated coals' char morphology, Ingeniería e Investigación 33 (1) (2013) 13–17.

[32] S. Burtsev, Y. Kuzmin, An efficient flood-filling algorithm, Computers & Graphics 17 (5) (1993) 549–561.

[33] J. Cervantes, J. Taltempa, F. García-Lamont, J. S. R. Castilla, A. Y. Rendon, L. D. Jalili, Análisis comparativo de las técnicas utilizadas en un sistema de reconocimiento de hojas de planta, Revista Iberoamericana de Automática e Informática Industrial RIAI 14 (1) (2017) 104–114.

[34] R. M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man, and Cybernetics SMC-3 (6) (1973) 610–621.

[35] M. Pietikäinen, T. Ojala, Z. Xu, Rotation-invariant texture classification using feature distributions, Pattern Recognition 33 (2000) 43–52.

[36] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1 of CVPR '05, 2005, pp. 886–893.

[37] C. B. Barber, D. P. Dobkin, H. Huhdanpaa, The quickhull algorithm for convex hulls, ACM Transactions on Mathematical Software 22 (4) (1996) 469–483.

[38] J. W. Tukey, Exploratory Data Analysis, Behavioral Science: Quantitative Methods, Addison-Wesley, Reading, Mass., 1977.

[39] D. M. Powers, Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation, Journal of Machine Learning Technologies 2 (1) (2011) 37–63.

[40] R. Kohavi, G. H. John, Wrappers for feature subset selection, Artificial Intelligence 97 (1) (1997) 273–324.