Automatic Morphological Categorisation of Carbon Black Nano-aggregates

Juan López-de-Uralde[†], Iraide Ruiz[†], Igor Santos[†], Agustín Zubillaga[†], Pablo G. Bringas[†], Ana Okariz[‡], and Teresa Guraya[‡]

[†]S^3Lab, University of Deusto, Bilbao, Spain
{jlopezdeuralde,iraide.ruiz,isantos}@deusto.es
{agustin.zubillaga,pablo.garcia.bringas}@deusto.es

[‡]Universidad del País Vasco UPV/EHU
Bilbao, Spain
{ana.okariz,teresa.guraya}@ehu.es

Abstract. Nano-technology is the study of matter behaviour on atomic and molecular scale (i.e. nano-scale). In particular, carbon black is a nano-material generally used for the reinforcement of rubber compounds. Nevertheless, the exact reason behind its success in this concrete domain remains unknown. Characterisation of rubber nano-aggregates aims to answer this question. The morphology of the nano-aggregate takes an important part in the final result of the compound. Several approaches have been taken to classify them. In this paper we propose the first automatic machine-learning-based nano-aggregate morphology categorisation system. This method extracts several geometric features in order to train machine-learning classifiers, forming a constellation of expert knowledge that enables us to foresee the exact morphology of a nano-aggregate. Furthermore, we compare the obtained results and show that Decision Trees outperform the rest of the counterparts for morphology categorisation.

Key words: aggregate morphology classifying, image processing, machine-learning, carbon black

1 Introduction

Matter behaviour on nano-scale is subject to quantum mechanics where microscopic and macroscopic theories are no longer applicable [1]. On this scale, nano-technology is the science that studies the comportment of the matter. This science has experienced a great development in the last years. In fact, they are considered to be the *basis for the next industrial revolution* since they have been applied to different areas such as energy, health care, chemical industry and material production [2]. Therefore, these processes are leading material manufacturers to a new generation of nano-material based products [2].

In the particular case of rubber compounds, reinforced materials with nano-particles, such as carbon black, are of great interest to the material industry.

Concretely, the latter modifies the mechanical and electrical properties of the former [3]. Although this process has been used in industrial production of rubber reinforced with carbon black [4] for the last years, the internal mechanisms that make that happen are not completely known.

In this way, there have been several studies about the morphology and microstructure of carbonaceous particles, such as the ones produced by diesel combustion [5]. This engine-emitted particles were studied with the purpose of assessing their climate impact. Likewise, the waste-water treatment includes similar steps to the ones needed for carbon black characterization: microscopic image processing, object segmentation, morphological characterisation and fractal analysis [6]. Similarly, with CAT (Computerized Axial Tomography) scans the same procedure has been applied to evaluate the rank of a tumour [7]. Still, these methods are performed in a semi-automatic or manual way with the consequent time and resource consumption.

Against this background, we present the first automatic machine-learning-based nano-aggregate morphology categorisation method. This method, based upon geometrical and fractal features is able to train several machine-learning algorithms in order to correctly determine the morphology of these aggregates. Specifically, we contribute to the state of the art in two main ways. Firstly, it consists in automatically segmenting and characterising the carbon black aggregates within an image. This technique makes the geometrical characterisation of carbon black and other nano-particles easy and fast. Secondly, a machine learning based classifier sorts carbon black aggregates according to their morphology.

2 Carbon Black

As we mentioned before, one of the principal carbon black applications is the reinforcement of rubber. This process creates a material with notably increased tensile strength and better tear and abrasion resistance (i.e. the capacity of a material to withstand different forces). These changes are conditioned by molecular, chemical and rheological attributes of the elastomer, on the filler characteristics and on the mixing process and technology [8]. In addition, carbon black primary particles seem to be spherical, blended together forming aggregates[9]. Following the *Van der Waals* forces, aggregates connect forming agglomerates [9]. Fig. 1 shows a graphic representation of the size of particles, aggregates and agglomerates.

Furthermore, the structure of carbon black particles ranges from crystalline to amorphous materials. Crystallite flat surfaces and amorphous carbon surfaces are less energetic areas, whereas crystallite edges are the most energetic ones [10].

Commonly, the aggregates can be divided into four different types of morphologies [11] (shown on Fig. 2). To this end, an estimation for discerning between the four categories is to calculate the *aggregate length/width* ratio and aggregate irregularity, however, this method is not an exact classification and includes a difficult value to measure: *irregularity* [11]:

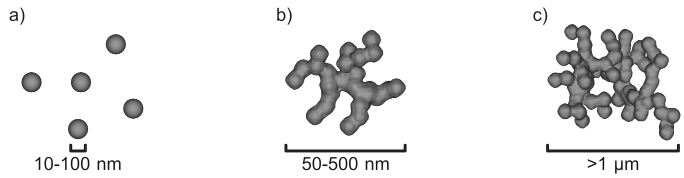

Fig. 1. Carbon black: a)particles; b)aggregate; c)agglomerate

- **Spheroidal:** Aggregates with a L/W ratio lower than 1.5 can be classified as spheroidal.
- **Ellipsoidal:** Aggregates with a L/W ratio between 2 and 3.5 can be classified as ellipsoidal.
- **Linear:** Linear ones have a L/W ratio greater than 3.5 and have low irregularity due to having elongated chains with few branches.
- **Branched:** Branched aggregates have also a L/W ratio greater than 3.5 but are highly irregular as a result of having more branches.

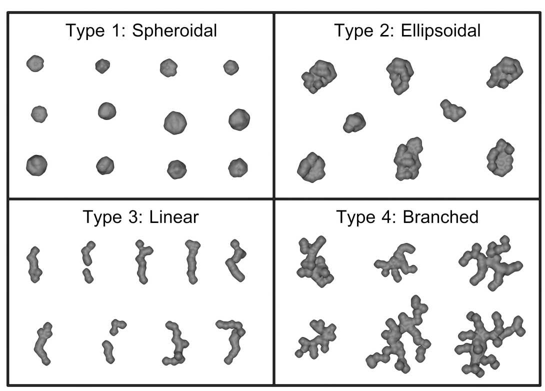

Fig. 2. Morphological categories for carbon black aggregates

Moreover, fractal dimension can also describe the aggregate structure [12]. Specifically, a fractal is a morphology that can be split into small copies of the whole [13]. To this end, Kaye [14] was the first one to apply fractal analysis to carbon black aggregates. He determined a perimeter fractal based upon the perimeter-area relationship of Mandelbrot [15] (shown on equation 1):

$$P \sim A^{Dp/2} \tag{1}$$

where P is defined as the projected aggregate perimeter, A as the projected area and D_p as the perimeter fractal. The greater the irregularity, the greater the D_p, however, highly acicular particles with a smooth perimeter may also give a high perimeter fractal [4].

Considering the scale of carbon black aggregates, *electron microscopes* are needed to analyse them. There are several types of microscope techniques based on the use of a particle beam of electrons such as *Transmission Electron Microscopy* (TEM) and *Scanning Electron Microscopy* (SEM).

3 Image Feature Extraction

Specifically, the aim of this treatment is to segment the aggregates and to extract several geometric features from them. Not only basic ones, such as area or perimeter, are considered but also more complex ones like perimeter fractal. Thereby, *machine-learning* classifiers will determine the morphology of unclassified aggregates using these features. Our algorithm follows the operations required by the *Standard Test Method for Carbon Black* [16] for analysing images captured by electron microscopes: background/noise elimination, thresholding, erosion and dilation.

In order to conduct the binarization, we start applying a Gaussian smoother [17], a 2-D convolution operator used to remove detail and noise. Second, we estimate a threshold for aggregate-background discrimination using Otsu's method [18]. We adjust this threshold to be more adequate for SEM images. We generate a binary image considering that pixels with value below the threshold correspond to background and pixels above it are part of the aggregate area.

Although we accomplish a smoothing process for noise reduction in the binarization phase, undesired elements may still be present in the image. These elements can be easily confused with the desired aggregates, thus, it is mandatory to eliminate them. To this end, we begin deleting minor areas and we continue filling holes inside aggregates. Moreover we improve the edge quality by dilating and eroding it with a disk shape morphological structuring element and we end deleting incomplete aggregates touching the edge of the image. Besides, we identify aggregates segmenting from the image the regions that surpass an specified area.

Based on the output image from the previous phase, we extract some geometric features, the ones marked with an '*' are the required ones according to the Standard Test Method for Carbon Black [16]. These parameters are measured in *nm* and when necessary estimated using stereological principles (i.e. the three-dimensional interpretation of two-dimensionally observed objects). To start with, the common ones are: perimeter*, area*, area-perimeter ratio, equivalent diameter, aggregate and particle volume, axis ratio, number of particles per aggregate, occlusion factor, absorption and circularity. In the second place are the parameters that require an explanation:

- **Feret diameters*:** A Feret diameter is defined as the distance between two tangents on opposite sides of the particle profile that are parallel to some fixed direction. So as to obtain valuable information related to the form of the particle, we extract 16 Ferets [16] separated by 11.5 degrees choosing the biggest (*Major Feret*), the smallest (*Minor Feret*) and the perpendicular one to the biggest.

- **Major and minor axis length:** Scalars specifying the length of the major and minor axis of the ellipse that have the same normalized second central moments as the region.
- **Centroid:** The center of mass of the region. It is formed by 2 values (x and y coordinates) normalized to the size of the *bounding box*, defined as the smallest rectangle containing the region.
- **Convex area:** The area of the *convex hull*, which is the smallest convex polygon that can contain the region.
- **Eccentricity:** The eccentricity is the ratio of the distance between the foci of the ellipse (i.e. the two points from which the distance to every point of the ellipse is constant) and its major axis length, taking values between 0 and 1.
- **Length-width ratio:** This ratio is computed with the maximum Feret and with the perpendicular Feret to the latter. Thereby, this commonly used parameter [11] is normalized.
- **Maximum Feret - minimum Feret ratio:** This ratio is similar to the previous one. However, considering that length and width are always orthogonal, it gives some extra information.
- **Area - convex area ratio:** Relation between the real area of the aggregate and the area of the *convex hull*. The smaller the ratio, the bigger the irregularity of the aggregate.
- **Extent:** Defined as the real area divided by the area of the *bounding box*.
- **Perimeter Fractal:** Determined by $P \sim A^{Dp/2}$ as explained in section 2.
- **Aggregation factor:** Defined by $13.092(\frac{P^2}{A})^{-0.92}$ where P is the perimeter and A is the area. If lower than 0.4 then it is equal to 0.4.

Finally, we generate a training vector $\boldsymbol{v} = (v_1, v_2, ..v_{13})$ per aggregate containing all these characteristics. Concretely, each position v_n in the vector represents a geometric feature and has up to 6 decimals. The collection of vectors forms the *corpus* I, which provides the learning dataset for the classification system.

4 Experimental Evaluation

Initially, we obtained several images with three electron microscopes on different magnification scales. Thirteen images with 2 SEM microscopes, Hitachi S-3400N and Hitachi S4800, and eleven with a Transmission Electron Microscope, the Philips EM208S. After performing a preliminary evaluation of the aggregate segmenting process, we chose the second Scanning Electron Microscope (SEM).

In this way, we collected 102 images of carbon black aggregates with a Hitachi S-4800 Scanning Electron Microscope. Images were captured at 30000x magnification with an average of 3 aggregates per image resulting in 266 correctly segmented aggregates that have formed the case of study.

We segmented all the aggregates from the images, then we labelled them and finally, we generated a Comma-Separated Values (CSV) file with all the characteristics and finally we performed machine learning studies to classify the aggregates.

In these experiments, we extracted 26 variables from each aggregate. The dataset was not balanced for the four existing classes due to scarce data. Specifically, 9 aggregates were of type spheroidal, 86 ellipsoidal, 51 linear and 120 branched. To address both problems (scarce and unbalanced data) we applied Synthetic Minority Over-sampling TEchnique (SMOTE) [19], which is a combination of over-sampling the less populated classes and under-sampling the more populated ones. Nevertheless, the over-sampling is performed by creating synthetic minority class examples. In this way, instances were still unique and classes became more balanced.

More accurately, we conducted the next methodology in order to test the suitability of each machine-learning algorithm:

- **SMOTE:** We built a dataset that contains the result of applying SMOTE to the original dataset in order to compare the results of the machine-learning classifiers with and without this technique.
- **Cross validation:** This method is generally applied in machine-learning evaluation [20]. In our experiments, we performed a K-fold cross validation with $k = 10$. In this way, our dataset is 10 times split into 10 different sets of learning (90 % of the total dataset) and testing (10 % of the total data).
- **Learning the model:** For each fold, we accomplished the learning step of each algorithm using different parameters or learning algorithms depending on the specific model. In particular, we used the following models:
  - *Bayesian networks (BN):* With regards to Bayesian networks we utilize different structural learning algorithms: K2 [21], Hill Climber [22] and Tree Augmented Naïve (TAN) [23]. Moreover, we also performed experiments with a Naïve Bayes Classifier [20].
  - *Support Vector Machines (SVM):* We performed experiments with a polynomial kernel [24], a normalized polynomial Kernel [25] and Pearson VII function-based universal kernel [26].
  - *K-nearest neighbour (KNN):* We performed experiments with $k = 1$, $k = 5$, $k = 10$, $k = 15$, $k = 20$ and $k = 25$.
  - *Decision Trees (DT):* We performed experiments with J48(the *Weka* [27] implementation of the *C4.5* algorithm [28]) and Random Forest [29], an ensemble of randomly constructed decision trees.
- **Testing the model:** We evaluated the percent of correctly classified instances and the *area under the ROC curve* (AUC) that establishes the relation between false negatives and false positives [30].

Table 1 shows the obtained results in terms of accuracy percent. In this way, regarding the results without the use of SMOTE, most of the classifiers obtained only medium results, with the exception of Naïve Bayes method, which was the worst, with results lower than 50 %. Otherwise, when SMOTE technique was applied, every classifier improved its accuracy in a significant manner. Specially, Naïve Bayes increased its accuracy in more than 20 %. Furthermore, Random

Table 1. Results of the machine-learning classifiers with regards to accuracy (%).

Machine-learning Model	Original Dataset	With SMOTE	
DT: J48	69.04	79.77	✓
DT: RandomForest with 1000 trees	73.40	83.61	✓
SVM: Polynomial Kernel	68.21	78.27	✓
SVM: Normalized Polynomial Kernel	67.30	75.68	✓
SVM: Pearson VII universal kernel	68.48	80.24	✓
KNN K=1	63.58	77.30	✓
KNN K=5	66.13	78.23	✓
KNN K=10	64.75	76.01	✓
KNN K=15	66.52	76.57	✓
KNN K=20	68.09	76.57	✓
KNN K=25	68.39	75.92	✓
Naïve Bayes	48.99	70.32	✓
BN: K2	56.37	77.33	✓
BN: Hill Climber	56.37	77.33	✓
BN: TAN	68.60	79.03	✓

✓, x, − statistically significant improvement, degradation or non significant change

Table 2. Results of the machine-learning classifiers with regards to AUC.

Machine-learning Model	Original Dataset	With SMOTE	
DT: J48	0.76	0.81	−
DT: RandomForest with 1000 trees	0.89	0.94	−
SVM: Polynomial Kernel	0.82	0.91	✓
SVM: Normalized Polynomial Kernel	0.81	0.90	✓
SVM: Pearson VII universal kernel	0.81	0.90	✓
KNN K=1	0.70	0.71	−
KNN K=5	0.82	0.88	−
KNN K=10	0.83	0.90	−
KNN K=15	0.85	0.92	✓
KNN K=20	0.86	0.93	✓
KNN K=25	0.86	0.93	✓
Naïve Bayes	0.81	0.90	✓
BN: K2	0.85	0.92	✓
BN: Hill Climber	0.85	0.92	✓
BN: TAN	0.84	0.91	✓

✓, x, − statistically significant improvement, degradation or non significant change

Forest, a type of Decision Tree, outperformed the rest of the classifiers with an accuracy of 83.61 %.

Nevertheless, focusing only on accuracy may be misleading and, therefore, we performed an analysis of the AUC. To this extent, Table 2 shows the results in terms of AUC. As occurred with accuracy, when SMOTE is omitted from the methodology the results are quite modest. Naïve Bayes was also the worst this time with an AUC of 0.81. Notwithstanding, we observed the same improvement using SMOTE, increasing the AUC of every classifier. Random Forest was also the best classifier in terms of AUC with a value of 0.94.

Summarizing, by means of machine learning algorithms we were able to accomplish aggregate morphology classification. Besides, with the help of synthetic re-sampling more data was produced and the four classes became more balanced. Thereby, we overcame the imbalance problem without merging the dataset, an inappropriate option due to the size of our dataset.

5 Conclusions and Future Work

Since nano-particles are able to modify the mechanical and electrical properties of materials [3], manufacturers have been led to a new generation of nano material-based production. Moreover, depending on the aggregate type [11] and the mixing process the obtained product varies [8].

In this paper, we have proposed the first automatic machine-learning-based nano-aggregate morphology categorisation method. This technique correctly determined the morphology of nano-aggregates, based on the use of geometrical and fractal characteristics as features for the training of several machine-learning classifiers. Furthermore, the empirical validation showed that this method is capable of classifying the morphology of aggregates with an accuracy of over 80%.

Future work will compare results based on original samples with the present results obtained with SMOTE re-sampling [19]. To this end, we will acquire more SEM images in order to generate a larger training dataset. In addition, we are planning to improve the image-processing algorithm so as to work with TEM images. On the other hand, we will focus on developing a 3-dimensional tool in order to accomplish *skeletonization* and 3D modelling of the aggregates.

Acknowledgements

We thank Mikel Salazar for the carbon black 3D simulation images created for this paper. In addition, we thank the Basque Government for their support with the project UEGV09/C19. Technical and human support provided by SGIker, specifically the EM laboratory of the Analytical Microscopy and High Resolution in Biomedicine General Service, is gratefully acknowledged. Last but not least we are grateful to Maria Carmen Huarte for her invaluable corrections.

References

1. Roco, M., Bainbridge, W.: Societal implications of nanoscience and nanotechnology. Kluwer Academic Pub (2001)
2. Kiparissides, C., Clausen, B., Boehm, L., Wilkins, T., Kellermayer, M., Baraton, M., Hossain, K.: NMP expert advisory group (EAG) position paper on future RTD activities of NMP for the period 2010 - 2015. Technical report
3. Mather, P., Thomas, K.: Carbon black/high density polyethylene conducting composite materials: Part I Structural modification of a carbon black by gasification in carbon dioxide and the effect on the electrical and mechanical properties of the composite. Journal of materials science **32**(2) (1997) 401–407
4. Donnet, J., Bansal, R., Wang, M.: Carbon black: science and technology. CRC (1993)
5. Soewono, A.: Morphology and microstructure of diesel particulates. Master's thesis
6. Amaral, A.: Image analysis in biotechnological processes: applications to wastewater treatment. PhD thesis, Universidadde Do Minho (2003)
7. Al-Kadi, O.: Tumour Grading and Discrimination based on Class Assignment and Quantitative Texture Analysis Techniques. PhD thesis, University of Sussex (2009)

8. Fröhlich, J., Niedermeier, W., Luginsland, H.: The effect of filler–filler and filler–elastomer interaction on rubber reinforcement. Composites Part A **36**(4) (2005) 449–460
9. Donnet, J.: Black and white fillers and tire compound. Rubber chemistry and technology **71**(3) (1998) 323–341
10. De, S., Naskar, K., White, J.: Rubber Technologist's Handbook, Volume 2. Smithers Rapra (2008)
11. Herd, C., McDonald, G., Hess, W.: Morphology of carbon-black aggregates: fractal versus euclidean geometry. Rubber chemistry and technology **65**(1) (1992) 107–129
12. Meakin, P.: Formation of fractal clusters and networks by irreversible diffusion-limited aggregation. Physical Review Letters **51**(13) (1983) 1119–1122
13. Mandelbrot, B.: The fractal geometry of nature. W.H. Freeman, San Francisco (1982)
14. Kaye, B.: Fractal description of fineparticle systems. Particle Characterization in Technology: Morphological analysis (1984) 81
15. Mandelbrot, B.: Form, chance, and Dimension. Chance and Dimension. Freeman, San Francisco (1977) 1–234
16. for Testing, A.S., Materials: ASTM D3849-02 - Standard Test Method for Carbon Black - Morphological Characterization of Carbon Black Using Electron Microscopy (2002) Testing method.
17. Pajares, G., de la Cruz, J.M.: Visión por Computador. Ra-Ma Publishers (2007)
18. Otsu, N.: A threshold selection method from gray-level histograms. Automatica **11** (1975) 285–296
19. Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W.: SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research **16**(3) (2002) 321–357
20. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press (1995)
21. Cooper, G.F., Herskovits, E.: A bayesian method for constructing bayesian belief networks from databases. In: Proceedings of the 7^{th} conference on Uncertainty in artificial intelligence. (1991)
22. Russell, S.J., Norvig: Artificial Intelligence: A Modern Approach (Second Edition). Prentice Hall (2003)
23. Geiger, D., Goldszmidt, M., Provan, G., Langley, P., Smyth, P.: Bayesian network classifiers. In: Machine Learning. (1997) 131–163
24. Amari, S., Wu, S.: Improving support vector machine classifiers by modifying kernel functions. Neural Networks **12**(6) (1999) 783–789
25. Maji, S., Berg, A., Malik, J.: Classification using intersection kernel support vector machines is efficient. In: Proc. CVPR. Volume 1. (2008) 4
26. Üstün, B., Melssen, W., Buydens, L.: Visualisation and interpretation of support vector regression models. Analytica chimica acta **595**(1-2) (2007) 299–309
27. Garner, S.: Weka: The Waikato environment for knowledge analysis. In: Proceedings of the New Zealand Computer Science Research Students Conference. (1995) 57–64
28. Quinlan, J.: C4. 5 programs for machine learning. Morgan Kaufmann Publishers (1993)
29. Breiman, L.: Random forests. Machine learning **45**(1) (2001) 5–32
30. Singh, Y., Kaur, A., Malhotra, R.: Comparative analysis of regression and machine learning methods for predicting fault proneness models. International Journal of Computer Applications in Technology **35**(2) (2009) 183–193