

Automatic Ferrite Content Measurement based on Image Analysis and Pattern Classification

Hafiz Muhammad Tanveer
Center for High Energy Physics (CHEP)
University of the Punjab
Lahore, Pakistan

Waleed Asif
Computer Science Department
Qarshi University
Lahore, Pakistan

Hafiz Muhammad Tahir Mustafa
Electrical Engineering Department, College of Engineering
& Emerging Technology (CEET), University of the Punjab
Lahore, Pakistan

Munir Ahmad
Institute of Nuclear Medicine and Oncology (INMOL)
Lahore, Pakistan

Muhammad Anjum Javed, Maqsood Ahmad
Center for High Energy Physics (CHEP)
University of the Punjab

Abstract—The existing manual point counting technique for ferrite content measurement is a difficult time consuming method which has limited accuracy due to limited human perception and error induced by points on boundaries of grid spacing. In this paper, we present a novel algorithm, based on image analysis and pattern classification, to evaluate the volume fraction of ferrite in microstructure containing ferrite and austenite. The prime focus of the proposed algorithm is to solve the problem of ferrite content measurement using automatic binary classification approach. Classification of image data into two distinct classes, using optimum threshold finding method, is the key idea behind the new algorithm. Automation of the process to measure the ferrite content and to speed up specimen's testing procedure is the main feature of the newly developed algorithm. Improved performance index by reducing error sources is reflected from obtained results and validated through the comparison with a well-known method of Ohtsu.

Keywords—Pattern classification; Decision threshold; Machine learning; Microstructure

I. INTRODUCTION

In the perspective of materials science and engineering, the microstructure (microscopic image) of specimen is the most considerable entity. The study of microstructure correlates the properties of material with microstructure. Quantitative measures of micrographs determine the specific characteristics of microstructures.

The measurement of different metallic components present in specimen, by employing microstructure to evaluate percentage composition, is an important aspect of quantitative metallography. Vital role of quantitative metallography is very important in the discipline of materials science and engineering. It provides the basis to develop an appropriate mathematical model by considering the relations between processes, mechanical properties and microstructures.

Considering these points, the vital role of quantitative metallography is very obvious in the discipline of materials science and engineering [1-3].

Ferrite content measurement is a significant parameter which determines the mechanical strength of material. Experimental techniques for ferrite content measurement are broadly classified as destructive techniques and non-destructive techniques [4]. Destructive techniques refer to the measurements from microstructures. On the other hand, non-destructive techniques are directly applicable to the material's specimen for ferrite content measurement. There is no need to acquire fine microstructure of specimen for ferrite content measurement procedure. In destructive techniques, measurements are taken from microstructures, while in non-destructive techniques specimen is directly employed for measurements [5]. Manual point counting and Image analysis are examples of destructive techniques while magnetic method (Magne-Gage instrument) and magnetic induction method (ferritescope or vibrating sample magnetometer, VSM) falls under the category of non-destructive techniques for ferrite content measurement[1;4].

Ferritescope is a device which measures the ferrite contents from material's specimen directly instead of microstructure. The principle of magnetic induction is applied to measure the ferrite contents by employing ferritescope[6]. Magnetic portion of specimen interacts with the magnetic field generated by coil and induced voltage measures the proportion of ferrite content in second coil. The correct identification of ferrite and non-ferrite structures is a particular advantage of magnetic induction technique. The magnetic permeability of steel plays an important role for the measurement of ferrite content because the amount of ferrite present in specimen tends to correlate with the permeability of steel. This permeability specifies the ferrite content of material being analyzed, when calibrated with standards having known ferrite

content. Digital read-out dial or calibrated dial provides the measure of ferrite content [4]. Ferritescope is very costly device, so in spite of having high accuracy, this method is not cost effective and rarely used.

Manual point counting method is based on stereological principle in which a square grid of particular dimensions is superimposed over the specimen's microstructure. This square grid is moved systematically through the specified number of different fields to cover specimen's surface area. This method gives an impartial statistical estimation of volume fraction corresponding to ferrite or any identifiable constituent in microstructure. There are few sources of error associated with manual point counting method such as points on grid boundaries and human perception limitation which are responsible for low accuracy of obtained results. The American Society for Testing Materials (ASTM E562) [7] standard describes the detailed procedure for the measurement of volume fraction of ferrite.

Image analyzer is a device capable of measuring ferrite content measurement from specimen's microstructure directly. The accuracy of desired results is strictly dependent on the acquired image quality. In the presence of bad contrast or analyzed features have a discontinuous outline in microstructure, the automatic measurements of image analyzer provide inaccurate results [8].

The microstructure under experiment contains the information of two metallic components. One of them is Austenite and other one is Ferrite [9]. The histogram of microstructure has two distinct regions. First region corresponds to Ferrite component and second region indicates the presence of Austenite. This pattern of histogram in microstructure attracts the attention to solve the problem of separating each metallic component by volume. Pattern classification based on Image data analysis is the possible approach towards efficient solution for ferrite content measurement. Image data analysis based on histogram pattern is the first step in classifier design. Optimum decision threshold between two regions separates one class of data from other [7;10-13].

The analysis and implementation of systems that are able to learn pattern from data is the subject of machine learning. Supervised machine learning and un-supervised machine learning are two primary machine learning techniques. In supervised learning, data analysis and prior knowledge of class label are used to design a classifier whereas finding pattern in data without any prior knowledge is the main focus of un-supervised learning [14;15]. Supervised learning leads to classification, whereas un-supervised learning ends up with clustering (similar groups in data). In variety of situations, linear and non-linear classifiers are employed to separate data into different classes. Linear classifier is most suitable option in situations where decision boundary may be marked as hyper-plane among different classes. In complex data distributions, inter mixing of data from different classes is quite random, and not possible to separate data into different classes by passing hyper-plane as decision threshold. Non-linear classifier deals such type of complex data distributions [7;15]. The Pattern classification deals with the identification

and separation of data into different classes. Feature selection or extraction plays an important role in determining the performance of classifier.

Otsu's method provides a global threshold value to convert the grayscale images into binary images. The existence of two distinct and separable categories of image pixels is an essential condition for this technique to work. As a consequence, the image histogram must follow bi-modal pattern. The threshold value corresponds to the optimal decision boundary for classification of image pixels. Image pixels having values below this threshold belongs to one class and pixels with values above threshold represent other class. These two classes describe the two distinct phases present in image. The percentage of each phase is evaluated quite easily after classification of pixels.

There exists no method in literature for automatic ferrite content measurement from microstructure by using image analysis and pattern classification. Up to our knowledge, present research fills this gap and provides an efficient solution for ferrite content measurement [13;15-17]. The GPF algorithm based on image processing and pattern classification has the ability to automatically measure the volume fraction of ferrite. This method will enhance the specimen's testing speed to a great extent. Machine learning nature of GPF algorithm in this paper has minimized the errors of manual point counting technique. The proposed technique will probably replace the tedious manual point counting method [17].

This paper is organized as follows. Section II explains the methodology and experimental details. The analysis and discussion on obtained results are included in section III. Finally, the section IV deals with the conclusion and possible future directions to extend the present research along with the potential application areas.

II. EXPERIMENTAL DETAILS

First of all, specimen is prepared by following some etching standard. The heat treatment and cooling method assigns a unique color to each metallic component in specimen. Microstructure (microscopic image) of specimen is taken with some suitable magnification index (200 or 500 times) to visualize inner details clearly. In the present study the specimen contains two metallic components (ferrite and austenite) and correspondingly two phases in microstructure. Each phase having distinct color and particular range of pixel intensities. The binary classifier is designed to separate ferrite content from specimen's microstructure. Analysis of image histogram suggests the linear classifier with optimal decision threshold to solve problem. MATLAB software is an efficient tool to program the functionality of linear classifier.

The process of evaluating the ferrite content measurement in a specimen having two phases is categorized in analysis and classification modules. The function $f(x, y)$ describes the discrete-time version of specimen's microstructure with spatial resolution of $(M \times N)$,

$$f(x, y) = \sum_{j=-N/2}^{N/2} \sum_{i=-M/2}^{M/2} f(i, j) \delta(x - i, y - j) \quad (1)$$

In the quantization of 2-D discrete time signal, the standard of 256 gray levels (0-255) is most appropriate to

assign intensity level of each image pixel. One byte (8-bits) is required to store the information of one pixel in gray scale image. For analysis purpose, RGB image is converted to grayscale image to find the range of pixel intensities corresponding to each phase.

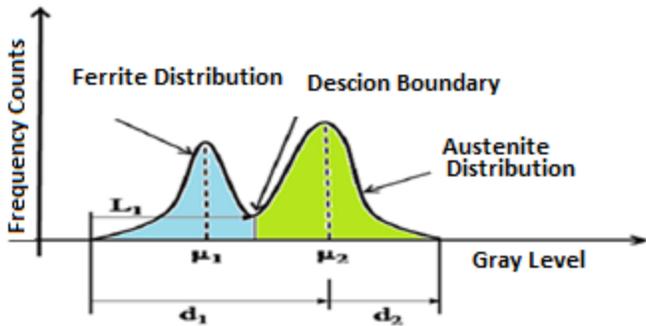


Fig. 1. Description of proposed algorithm with various parameter definitions

For instance, the distribution of data in Figure 1 describes two distinct regions before and after decision boundary. The region before decision boundary indicates the presence of ferrite contents and after that austenite part. Distribution of data in Figure 1 allows the application of Gaussian kernel to each component. Statistical relations for normal distribution are applicable to each component of distribution. Gray levels corresponding to two data peaks describe the mean locations (μ_1 and μ_2) of two Gaussian kernels. In existing situation, the decision boundary is approximately at $\mu_1 + \sigma_1$ from mean (μ_1) of first Gaussian kernel and at $\mu_2 - \sigma_2$ from data peak (μ_2) of second Gaussian kernel. The standard deviation σ from mean (μ) covers approximately 84.4% of area under distribution curve. Decision threshold is the intersection of two Gaussian kernels, the lowest pixel count with respective gray level between two optimized data peaks. This optimum threshold minimizes the classification error to a negligible value. The reason behind that lies in the fact that this optimized threshold is actually the starting point of second distribution and last point of first distribution. The proposed algorithm first determines the maximum data peak (maximum pixel count with respective gray level) of whole image.

$$y_1 = \Phi[g(f)], \quad 0 \leq f \leq 255 \quad (3)$$

In above equation, Φ is the function that provides the maximum pixel count of whole image histogram with corresponding gray level. Suppose y_1, f_1, n_1 and $N = (f_1 - n_1)$ are maximum pixel count, corresponding gray level, gray level at non-zero starting pixel count and total number of data samples (data peaks) between f_1 and n_1 in histogram respectively. The maximum data peak may fall in any distribution. In case, second distribution contains maximum data peak, the process of finding the data peak in first distribution is quite straight forward. When maximum data peak falls in first distribution, reversal in the order of distributions needs to be considered to make the process identical to above mentioned. Recursive window is applied

from data peak in second distribution by considering two consecutive data samples (peaks) in single step. The process continues, until non-zero starting point of first distribution reaches. This is an indication to the completion of first iteration.

Image histogram gives information of relative frequency of pixels at particular gray level. The mathematical formulation of histogram is represented by function $g(f)$.

$$g(f) = n_f, \quad 0 \leq f \leq 255 \quad (2)$$

Histogram pattern of microstructure shows two distinct and linearly separable regions. The existence of two distinct regions provides evidence to the two metallic components in specimen's microstructure. One region corresponds to the existence of ferrite and other to austenite. The primary function of linear classifier is to find optimal decision boundary. Ferrite class label is assigned to pixel intensity range below decision boundary. The range of pixel intensity above decision threshold is assigned austenite class labels. Analysis of microstructure gives an identification of different phases, and classification separates metallic components in terms of pixel count with specific intensity range. The scope of present research is to separate ferrite content by volume in microstructure having two phases based on pattern classification. The beauty of proposed algorithm is to transform the complex and lengthy procedure for measurement of metallic components to a simple arithmetic problem. Finding data peaks in each distribution and decision boundary in between provide the required solution of data classification problem. Minimum pixel count between two data peaks is the optimum decision threshold and the most important parameter for the classification of image data.

The number of data samples reduces to one half after each iteration, and this procedure ends up with two optimized data peaks with corresponding gray levels of both distribution parts. One of these optimized data peaks refers to the mean of first Gaussian kernel and other to the second Gaussian kernel. Decision threshold provides the pixel intensity range for ferrite content measurement. On the basis of this information, volume fraction of ferrite is determined by comparing the pixel counts in ferrite region to the total number of pixels.

$$g_i^m(q_j)|_{j=1}^i = g^{m-1}(f_p) * H_{u \rightarrow u-1}|_{u=f_1-2k}^i \quad (4)$$

(* Operator defines recursive window for the selection of greater sample value in the comparison of every two samples till the completion of iteration).

$$\frac{N}{2} \leq i \leq 2, \quad n_1 \leq f_p \leq f_1, \quad 0 \leq k \leq i, \quad 1 \leq m \leq n \quad (5)$$

The vector q_j contains gray levels at ascending order index, after the completion of each iteration, the length of this vector reduces to one half. Window function $H_{u \rightarrow u-1}$ compares two data samples in one step and selects the highest of two. The result will be the reduction of data peaks to one half for next iteration. Algorithm flow chart is shown in Figure 2.

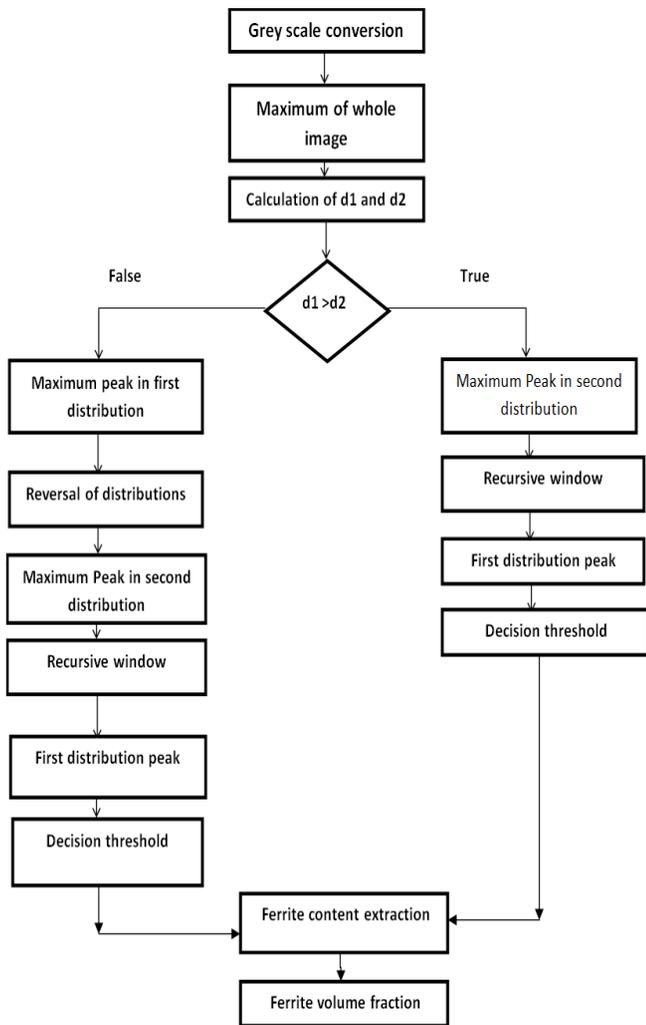


Fig. 2. Algorithm flow chart of the proposed algorithm to calculate ferrite volume fraction in a microstructure

We select four microstructures (Sample 1, 2, 3 and 4) having variations in histogram pattern. First microstructure (Sample 1) contains histogram in which both regions of distribution are very similar to Gaussian kernel. The histogram pattern in second microstructure (Sample 2) indicates Gaussian trend in first distribution part, but second distribution part doesn't obey Gaussian pattern in second half. In third microstructure (Sample 3), the situation is approximately same as described in Sample 2. Worst scenario is being presented having lot of irregularities in both distribution parts in fourth and last microstructure (Sample 4). Figure 3 shows four samples (microstructure images) used to measure ferrite volume ratio.

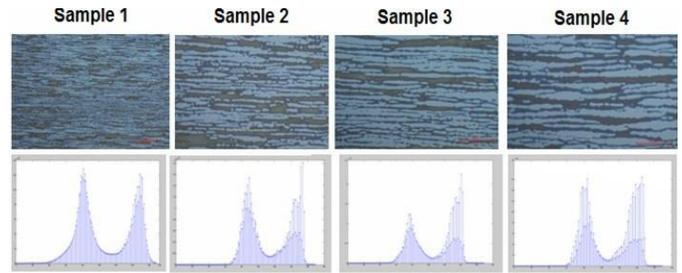


Fig. 3. Four different microstructure samples used to evaluate performance of various algorithms for the measurement of ferrite volume fraction. The samples were selected based on the spread of histogram degrading from an almost Gaussian curve

III. RESULTS AND DISCUSSIONS

The chi-square goodness-of-fit (GOF) test is employed to know that data samples came from a population with specific distribution. Information of model (distribution) for which decision is required and sufficient number of data samples are the limitations of chi-square GOF test. In this case, there is a need to test that data samples follow Gaussian distribution. This hypothesis is tested by applying chi-square GOF test, in which chi-square statistic for the given data samples is compared against some critical value. The degree of freedom and significance level (0.05 default) are important factors to decide some critical value. In case of hypothesis acceptance, the Gaussian kernel is applied on the distribution of data and all the standard mathematical relations for normal distribution are valid here.

The chi-square statistic value is less than critical value ($X_{statistic}^2 < X_{critical}^2$) for different microstructure images used in the experimental procedure. The values of degree of freedom and level of significance are 6 and 0.1 respectively. This test validates the GOF for Gaussian distribution. Gaussian kernel is applied to approximate the distribution of data corresponding to both phases present in microstructure. The GPF algorithm works accurately in this situation and gives reliable results. These results include mean, variance, standard deviation associated to the distribution of each phase and most important one is the optimal decision boundary for data classification.

For manual point counting (MPC) procedure, 30 equal sized fields were placed on each sample image in order to measure ferrite content to cover the whole image and pixels were counted for their gray level classification manually. Figure 4 presents data for Sample 1 for all 30 fields evaluated. Table 1 bears values for various parameters for each sample evaluated using MPC algorithm. In all four samples ferrite volume fraction evaluated as percentage of the ferrite content present are very close to each other with relatively low error.

TABLE I. FERRITE VOLUME FRACTION MEASUREMENT USING MPC WITH 30 FIELDS PER MICROSTRUCTURE

Sample No.	Total Fields Counted	Phase Vol. Fraction (%)	Standard Deviation	95% CI
1	30	40.1	7.7	2.9
2	30	47.1	7.6	2.9
3	30	44.5	10.1	3.8
4	30	44.5	8.8	3.3

TABLE II. DECISION BOUNDARY CALCULATION BY SPC AND GPF ALGORITHMS

Sample No.	μ_1	σ_1	SPC ($\mu_1 + \sigma_1$)	GPF (lowest pixel count)
1	82	30.1	112.1	116
2	90	17.7	107.7	116
3	79	21.5	100.5	109
4	85	18.6	103.6	108

TABLE III. FERRITE VOLUME FRACTION MEASUREMENT USING PROPOSED ALGORITHM

Sample No.	Pixel Intensity Range	Ferrite Volume Fraction
1	13-189	56.59
2	56-169	53.03
3	36-157	44.82
4	45-160	48.53

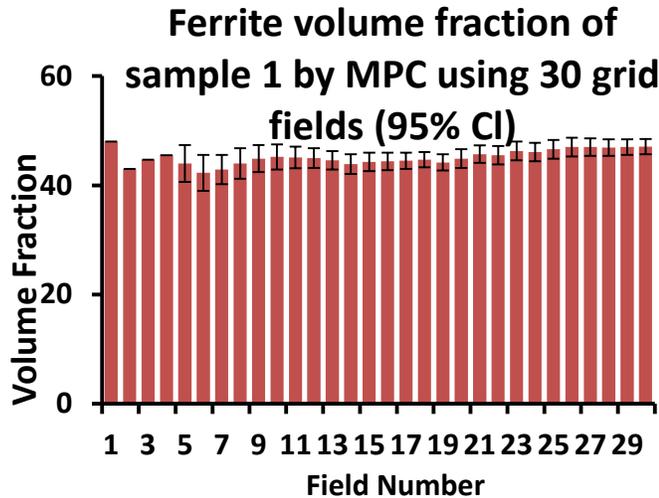


Fig. 4. Calculation of ferrite volume fraction of sample no. 1 by manual point counting using 30 fields per microstructure with relative error bars

For statistical point counting (SPC) algorithm, Gaussian kernel is applied on each distribution of histogram to obtain parameters (μ , σ). Decision boundary is approximately at the spread of σ_1 from mean μ_1 of first distribution. Ferrite content measurement is provided by considering pixel count for the range of gray levels below decision boundary. Classification boundary by applying Gaussian kernel is close to the result of proposed algorithm. This closeness in results provides mathematical support of Gaussian kernel to GPF algorithm.

Table 2 shows the parameters (μ_1 , σ_1) of first distribution part representing ferrite content region, and decision boundary values evaluated based on statistical estimation and proposed GPF algorithm. Pixel intensity range for ferrite content measurement in various microstructures is provided in Table 3.

The ferrite volume fraction by considering optimized decision boundary based on proposed algorithm for different microstructures are shown in Table 2. Several microstructures having highly symmetrical to worst possible scenario have been considered to check the validity of GPF algorithm. In the presence of irregularities, where second distribution doesn't obey Gaussian pattern, the GPF algorithm works well and provide quite satisfactory results.

The results obtained by applying GPF algorithm are approximately same to the one achieved by applying Gaussian kernel to each component of distribution. Sources of errors like points on grid boundaries and limited human perception have been resolved in the considerations of GPF algorithm. Thereby, results provided by GPF algorithm are more accurate and reliable than manual point counting method. The distance between first non-zero pixel count with corresponding gray level and optimal decision boundary L_1 indicates the range of pixel intensities for ferrite content measurement. Total pixel count for pixel intensity range L_1 indicates the presence of ferrite content measurement in microstructure having two phases. Application of Gaussian kernel on both distribution parts provides approximately the same result. The closeness of results obtained by applying Gaussian kernel and GPF algorithm provides mathematical support to measurement process.

If we carefully observe the Table1, we can clearly see that Table1 shows the results by manually applying the grid, field by field on the microstructures. Measurements of thirty fields are recorded, the volume fraction of ferrite in each field with 95% confidence interval and relative accuracy has been considered to average out the result. In microstructure (sample 3), ferrite volume fraction by applying manual point counting and GPF algorithm are very close to each other. On the other hand, there is a difference of approximately 16% in case of microstructure (sample 1). The reason behind this big difference lies in the fact that acquired image doesn't fulfill the required essential conditions for analysis. This difference will reduce to an acceptable level in the presence of conditions such as color contrast between different phases, clear grain boundaries and noise free focus. In sample 4 and sample 2, the difference in volume fraction is approximately 4% and 6% respectively.

The comparison between Otsu's method and proposed GPF algorithm is shown in table 4. Threshold values and percentage of ferrite content measurement for different microstructures are obtained by applying these two algorithms. The results obtained by employing two different techniques are very close to each other. Also the results obtained by GPF algorithm are very close to MPC results. This indicates the validity and reliability of results provided on the basis of GPF algorithm.

TABLE IV. COMPARISON BETWEEN OTSU'S METHOD AND GPF ALGORITHM

Sample No.	Otsu's Method		GPF Algorithm	
	Threshold	Ferrite Vol. Fraction (%)	Threshold	Ferrite Vol. Fraction (%)
1	113	57.65	116	56.59
2	118	62.65	116	53.03
3	110	49.25	109	44.82
4	110	56.36	108	48.53

The approach of GPF algorithm presented in this paper considers the whole image, pixel by pixel for the classification of all the data into two classes by deciding the optimized decision boundary. The results obtained in two techniques are shown in Table 3, but GPF algorithm adds the factor of accuracy by removing doubts and approximations associated with manual point counting measurement technique.

IV. CONCLUSION

Automation of ferrite content measurement based on image analysis and pattern classification is much faster and more accurate in comparison to the conventional manual point counting technique. The selection of optimal decision boundary minimizes classification errors and adds factor of accuracy to acquired results in relatively less time. Limitations associated with image analyzer and error sources regarding human perception in manual point counting have been resolved in machine learning algorithm for pattern classification. The crux of this research is to increase the efficiency of manual point counting for ferrite content measurement process by introducing the automatic computerized method to perform the same task in a very short span of time.

The present research deals with the separation of two phases in a microstructure image. Support Vector Machine (SVM) is a well-known binary classifier, the comparison of GPF algorithm with SVM may be one of the possible future directions to extend the present research activity. There is another dimension to proceed with the application of GPF algorithm i.e. generalization of GPF algorithm to work with microstructure images having more than two phases is another promising future direction. There is a definite need to analyze in great detail; the data present in the microstructure images with more than two phases. In this way important information of hidden pattern in the data can also be acquired for analysis.

This research may prove beneficial for the modern metallurgical industry all around the world in terms of accuracy and time saving. Material quality is related to the percentage of ferrite content in the material; GPF algorithm is able to automatically measure the ferrite content in the material more accurately in a minimum time and with much less chances of errors. Instead of a simple histogram analysis, segmentation may also be used for error reduction analysis. Computational efficiency with mathematical simplicity of analysis is another area to work in future.

The application area of this newly developed algorithm is not limited to metallurgy and material sciences. This method for binary classification will work reasonably well in the field of medical diagnosis and as a test method in factories to qualify a sample as pass or fail. This method may also be applicable to check an item for a specific qualitative property using the microstructure image of that particular item.

REFERENCES

- [1] R.T.DeHoff, E. H. Aigeltinger, Experimental Quantitative Microscopy with Special Applications to Sintering. In: J. Hirschhorn, K. Roll, editors. Advanced Experimental Techniques in Powder Metallurgy, Springer US; 1970, 81-137.
- [2] E. Underwood, Quantitative Stereology for Microstructural Analysis. In: J. McCall, W. Mueller, editors. Microstructural Analysis, Springer US; 1973, 35-66.
- [3] Practical Applications of Quantitative Metallography ASTM, 1983.
- [4] J. Niagaj, Au. Mazur, Review of methods for measurement of ferrite content in high alloyed steels and their welded joints. Welding International 2012;28:345-353.
- [5] F.D. Huang, L. Froyen, Quantitative analysis of microstructure in metals with computer assistance. Non-destructive Testing and Ultrasonics 2001;6:
- [6] GCd. Souza, JM. Pardal, SSM. Tavares, MdPC. Fonseca, JLF. Martins, EPd. Moura, I. CardoteFilho. Evaluation of phases proportions in welded joints of duplex stainless steel by non-destructive testing. Soldag.insp. 2013; 18:158-168.
- [7] L. Devroye. Automatic Pattern Recognition: A Study of the Probability of Error. IEEE Transactions on Pattern Analysis and Machine Intelligence 1988; 10:530-543.
- [8] JE. Hilliard, JW. Cahn. An Evaluation of Procedures in Quantitative Metallography for Volume-Fraction Analysis. In: The Selected Works of W. John. Cahn, John Wiley & Sons, Inc.; 1998, 63-64.
- [9] AW. Brewer, RL. Moment. Techniques and Standards for Measuring Ferrite in Austenitic Stainless Steel Welds. Welding Journal 1976;55:159s-164s.
- [10] JC. Russ. The Image Processing Handbook, Sixth Edition Taylor & Francis, 2011.
- [11] P. Osmond, H. Petitgand, JR. Philippe. Use of image processing in iron and steel making industry. Rev.Metall. 2000;2:187-196.
- [12] F. Keinosuke. Introduction to statistical pattern recognition (2nd ed.) Academic Press Professional, Inc., 1990.
- [13] PA. Deuyver, J. Kittler. Pattern Recognition: A Statistical Approach Prentice-Hall, 1982.
- [14] L. Chulhee, DA. Landgrebe. Feature extraction based on decision boundaries. Pattern Analysis and Machine Intelligence, IEEE Transactions on 1993;15:388-400.
- [15] L. Kanal. Patterns in pattern recognition: 1968-1974. Information Theory, IEEE Transactions on 1974;20:697-722.
- [16] G. Nagy. State of the art in pattern recognition. Proceedings of the IEEE 1968; 56:836-863.
- [17] S. Salzberg. Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. Mach Learn 1994; 16:235-240.