# A TEXTURE ANALYSIS APPROACH FOR AUTOMATIC FLAW DETECTION IN PIPELINES

*S. V. Bharath Kumar and Sivaramanivas Ramaswamy*

Imaging Technologies Lab
General Electric - Global Research
JFWTC, Bangalore - 560086, INDIA
{*bharath.sv, ramaswamy.nivas*}*@geind.ge.com*

**ABSTRACT**

Oil and natural gas are normally transported through a vast network of pipelines, a major segment of which are seamless pipes. The manufacturing processes associated with the production of seamless pipes introduces an artifact known as seamless pipe noise (SPN), in the data obtained from magnetic flux leakage (MFL) inspection of these pipelines. SPN poses a major challenge in the flaw detection and characterization as it can overwhelm the flaw response and can therefore, mask flaw signature in MFL data.

In this paper, we present a texture analysis approach to automatically detect flaws while characterizing the textures of SPN and flaw response using gray level co-occurrence matrix (GLCM). The motivation to propose texture analysis approach is to investigate the statistical properties of these signals. The proposed method provides a recognition rate of $97.29\%$ for flaws which are deeper than $20\%$ of the wall thickness of the pipe, thus confirming the existence of different textural characteristics for SPN and flaw response.

## 1. INTRODUCTION

Oil and natural gas are transported in most countries through a vast network of buried and aboveground pipelines. In order to ensure the integrity of the system, the pipelines are periodically inspected for damage caused by corrosion and other factors using a device called a "pig". The pig, in brief, is a magnetizer-sensor assembly, which employs the magnetic flux leakage (MFL) technique [1] for assessing the condition of the pipe. A strong permanent magnet in the pig nearly saturates the pipe wall with magnetic flux and the sensors in the pig capture any flux leakage, which occur when the pig encounters an anomaly. The signal picked up by the sensors can be interpreted as a *magnetic image* of the condition of pipeline, which is appropriately sampled, stored and subsequently analyzed off-line by trained data analysts.

MFL non-destructive evaluation of pipelines typically generate voluminous data, whose manual inspection could be very time consuming and the performance is subject to skills, training, emotional status and level of fatigue of the analyst. This necessitates for automatic methods for analyzing MFL data in order to improve reliability and decrease turn around time between actual pigging and the delivery of inspection results. Sophisticated signal processing techniques are required for MFL data analysis in order to detect corrosion pits, which are as shallow as $5 - 10\%$ of the pipe wall thickness.

Seamless pipes, which typically form $25\%$ of the pipeline network, are commonly found in the collection and distribution ends of the network and hence are usually located close to populated areas. Since any failure can be potentially catastrophic, it is extremely important that any anomaly in these pipes be detected in a timely and accurate manner. Because of the helical nature involved in the manufacturing process of the pipe, the data from its MFL inspection contains an artifact known as seamless pipe noise (SPN). SPN poses a major challenge in the flaw detection and characterization as it can overwhelm the flaw response and can therefore, mask flaw signature in MFL data. Methods for minimizing the effect of SPN and improving the detectability of defect signals in MFL data are, therefore, required.

The detection of flaws immersed in SPN has been treated as a de-noising problem in [2, 3], where an adaptive noise cancellation scheme involving normalized LMS algorithm has been used to improve the signal-to-noise ratio of the MFL data from seamless pipes. The drawback of the scheme is that it fails in situations where the pipes are highly corroded or pipes with low signal-to-noise ratio between SPN and flaw response. It also assumes the noise cancellation filter as a linear time variant system, where the linearity property is not valid in most cases.

In this paper, we present a texture analysis approach for automatic flaw detection, which involves the computation of texture descriptors for SPN and flaw response using gray level co-occurrence matrix (GLCM) [4]. The proposed method is motivated by the desire to study the statistical properties of SPN and flaw response. Since the second-order statistics provide more information than the first-order statistics, and in turn, also captures the textural content of these signals, we investigated texture features using GLCM. The support vector machine (SVM) [5] is used as the classifier to automatically detect flaw response from SPN using the derived texture features We achieve a recognition rate of $97.29\%$ for flaws with depth greater than $20\%$ of the wall thickness of the pipe, thus confirming the existence of different textural characteristics for SPN and flaw signals.

The paper is organized as follows. In Section 2, we present the method to compute first-order and second-order statistics of the given data and their related texture descriptors. We present the approach, experimental details and classification results of the proposed texture analysis method on MFL data from seamless pipes in Section 3. We conclude by pointing out to the different textural characteristics of SPN and flaw signals, which can be used to automatically detect flaws, even in highly corroded pipes.

## 2. STATISTICAL FEATURES FOR TEXTURE ANALYSIS

Although there is no strict definition of the image texture, it is easily perceived by humans and is believed to be a rich source of visual information about the nature and three-dimensional shape of physical objects. Generally speaking, textures are complex visual patterns composed of entities, or sub-patterns, that have characteristic brightness, color, slope, size, etc. Thus texture can be regarded as a similarity grouping in an image. The local sub-pattern properties give rise to the perceived lightness, uniformity, density, roughness, regularity, linearity, frequency, phase, directionality, coarseness, randomness, fineness, smoothness, granulation, etc., of the texture as a whole.

Different approaches to texture analysis have been investigated in the literature, for e.g. structural [6], statistical [6], model-based and transform-based. In this paper, we investigated statistical based approach of texture analysis using first-order and second-order statistics of SPN and flaw response, in order to identify good metrics that would automatically detect flaws in MFL data. Since, the MFL data has been defined as the magnetic image of the condition of the pipe, we use the terms "image" and "data" to mean MFL data.

### 2.1. First-order statistics

Let us assume an image of dimensions $N \times M$, to be a function $f(x, y)$ of two space variables $x$ and $y$, where $x = 0, 1, \cdots, N-1$ and $y = 0, 1, \cdots, M-1$. The function $f(x, y)$ can take discrete values $i = 0, 1, \cdots, G-1$, where $G$ is the total number of intensity levels in the image. The intensity-level histogram is a function showing (for each intensity level) the number of pixels in the whole image, which have this intensity, which is given as

$$h(i) = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \delta(f(x, y), i) \tag{1}$$

where $\delta(j, i)$ is the Kronecker delta function:

$$\delta(j, i) = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases} \tag{2}$$

The histogram of intensity levels is obviously a concise and simple summary of the statistical information contained in the image. Since, the calculation of the gray level histogram involves single pixels, it contains the first-order statistical information about the image (or its fragment). The probability mass function of occurrence of the intensity levels is obtained by dividing the values $h(i)$ by the total number of pixels in the image.

$$p(i) = \frac{h(i)}{MN}, \;\; i = 0, 1, \cdots, G-1 \tag{3}$$

The shape of the histogram provides many clues as to the character of the image. Different useful parameters (image features) can be calculated from the histogram to quantitatively describe the first-order statistical properties of the image, which are mentioned below.

The mean describes the average level of intensity of the image or texture being examined, whereas the variance describes the variation of intensity around the mean. The skewness is zero if the histogram is symmetrical about the mean, and is otherwise either positive or negative depending whether it has been skewed above or below the mean. Thus $\mu_3$ is an indication of symmetry. The kurtosis is a measure of flatness of the histogram. The entropy is a measure of histogram uniformity.

| | |
|---|---|
| *Mean:* | $\mu = \sum_{i=0}^{G-1} ip(i)$ |
| *Variance:* | $\sigma^2 = \sum_{i=0}^{G-1} (i-\mu)^2 p(i)$ |
| *Skewness:* | $\mu_3 = \sigma^{-3} \sum_{i=0}^{G-1} (i-\mu)^3 p(i)$ |
| *Kurtosis:* | $\mu_4 = \sigma^{-4} \sum_{i=0}^{G-1} (i-\mu)^4 p(i) - 3$ |
| *Energy:* | $E = \sum_{i=0}^{G-1} (p(i))^2$ |
| *Entropy:* | $H = -\sum_{i=0}^{G-1} p(i) \log_2(p(i))$ |

### 2.2. Second-order statistics

The second-order statistics involve context by considering the past pixel, unlike the case in first-order statistics. Hence, the order in which the data or image is scanned becomes important. The second-order histogram is defined as the gray level co-occurrence matrix $h_{d,\theta}(i, j)$. The joint probability mass function, $p_{d,\theta}(i, j)$ of two pixels that are distance $d$ apart along a given direction $\theta$ having particular (co-occurring) values $i$ and $j$ is computed by dividing $h_{d,\theta}(i, j)$ by the total number of neighboring pixels $R(d, \theta)$ in the image. Two forms of co-occurrence matrix exist $-$ one symmetric where pairs separated by $d$ and $-d$ for a given direction $\theta$ are counted, and other not symmetric where only pairs separated by distance $d$ are counted. Formally, given the image $f(x, y)$ with a set of $G$ discrete intensity levels, the matrix $h_{d,\theta}(i, j)$ is defined such that its $(i, j)^{th}$ entry is equal to the number of times that $f(x_1, y_1) = i$ and $f(x_2, y_2) = j$, where $(x_2, y_2) = (x_1, y_1) + (d\cos\theta, d\sin\theta)$. This yields a $G \times G$ matrix for each distance $d$ and orientation $\theta$. Different features that describe the second-order properties of the image can be computed from the joint probability mass function, $p_{d,\theta}(i, j)$, which are mentioned below.

| | |
|---|---|
| *Energy:* | $\sum_{i,j=0}^{G-1} (p_{d,\theta}(i, j))^2$ |
| *Correlation:* | $\frac{\sum_{i,j=0}^{G-1} ij p_{d,\theta}(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$ |
| *Contrast:* | $\sum_{i,j=0}^{G-1} (i-j)^2 p_{d,\theta}(i, j)$ |
| *Absolute value:* | $\sum_{i,j=0}^{G-1} \lvert i-j \rvert p_{d,\theta}(i, j)$ |
| *Inverse difference:* | $\sum_{i,j=0}^{G-1} \frac{p_{d,\theta}(i,j)}{1+(i-j)^2}$ |
| *Homogeneity:* | $\sum_{i,j=0}^{G-1} \frac{p_{d,\theta}(i,j)}{1+\lvert i-j \rvert}$ |
| *Entropy:* | $-\sum_{i,j=0}^{G-1} p_{d,\theta}(i, j) \log_2(p_{d,\theta}(i, j))$ |
| *Maximum probability:* | $\max_{i,j} p_{d,\theta}(i, j)$ |

$\mu_x$, $\mu_y$, $\sigma_x^2$ and $\sigma_y^2$ are defined as follows. $\mu_x = \sum_{i=0}^{G-1} ip_{d,\theta}^x(i)$, $\mu_y = \sum_{j=0}^{G-1} jp_{d,\theta}^y(j)$, $\sigma_x^2 = \sum_{i=0}^{G-1} (i-\mu_x)^2 p_{d,\theta}^x(i)$ and $\sigma_y^2 =$

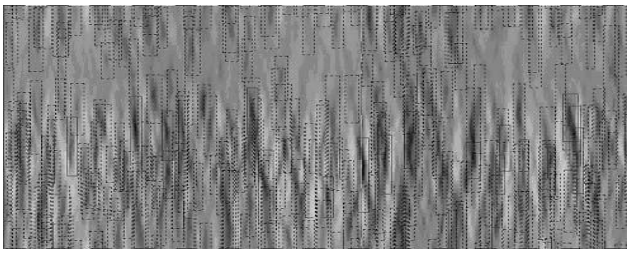

**Fig. 1**. Raw MFL data from $10''$ dia. seamless pipe, in which the flaw (solid box) and SPN (dotted box) regions are marked by the analyst.

$\sum_{j=0}^{G-1}(j-\mu_y)^2 p_{d,\theta}^y(j)$. $p_{d,\theta}^x(i)$ and $p_{d,\theta}^y(j)$ are the marginal densities computed as $p_{d,\theta}^x(i) = \sum_{j=0}^{G-1} p_{d,\theta}(i,j)$ and $p_{d,\theta}^y(j) = \sum_{i=0}^{G-1} p_{d,\theta}(i,j)$.

Angular second moment or energy measures the textural uniformity of the image. Absolute value and contrast measure the amount of local variation in the image. Inverse difference and homogeneity measure the smoothness of the data. Entropy measures the disorder of an image and maximum probability results in the pixel pair that is most predominant in the data.

## 3. TEXTURE ANALYSIS OF MAGNETIC FLUX LEAKAGE DATA

As aforementioned, the underlying idea in adaptive noise cancellation approach [2, 3] is to mitigate the sources of corruption (i.e. SPN) in MFL data obtained from seamless pipes. Because of the drawbacks involved in this technique, which are mentioned in Section 1, we investigate the characteristics of SPN and flaw response in statistical sense with the intention of capturing their textural content. Figure 1 shows the raw MFL data in which the flaws and SPN are marked by the analyst. It is obvious from Figure 1 that both the flaw signals and the background (SPN) look alike and noise cancellation schemes would fail. Though, we don't visually perceive any textures in both these regions, we do perceive some roughness content in the flaw region than in the regions where SPN alone is present. This is one of the motivations to investigate the textural content of the data in both of these regions.

In our study to capture the textural contents of SPN and flaw signal, MFL data from 30 highly corroded seamless pipes of varying diameters and varying pipe wall thickness is considered. Depending on the depth of the flaws, we considered two cases for automatic flaw detection.

1. flaw signals which are deeper than $20\%$ of the pipe wall thickness. In this case, the SPN and flaw signals with depth $<= 20\%$ of the pipe wall thickness would confuse the classifier making it to classify them as flaw signals and hence, we term them to be false positives.
2. flaw regions which are deeper than $10\%$ of the pipe wall thickness. For the similar reasons mentioned above, SPN and flaw signals with depth $<= 10\%$ of the pipe wall thickness are considered as false positives.

The flaws of larger depth are easier to detect than the flaws of smaller depth. Figure 1 shows a section of seamless pipe under consideration with flaw and SPN regions marked. The solid boxes represent the flaw signals and the black dotted boxes represent SPN. These boxes are marked by analysts, which are considered as the ground truth data. It can be seen that both SPN and flaw signals look similar and it calls an experienced analyst to distinguish flaw from SPN. Hence, an automated algorithm for flaw detection should not miss any flaw and should provide very low false calls.

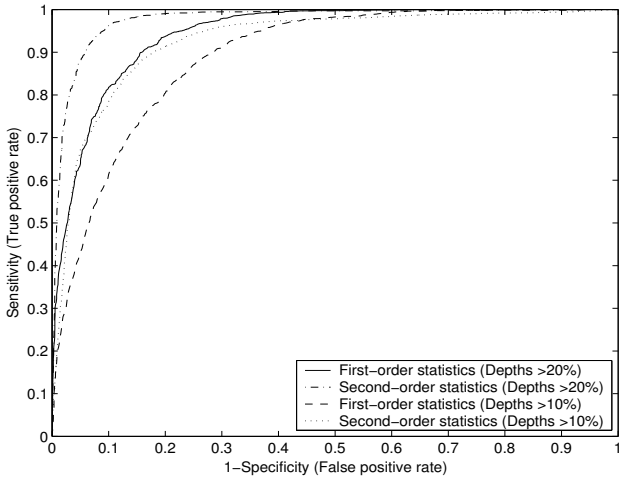


**Fig. 2**. Receiver operating characteristics for flaws of depth $> 10\%$ and $> 20\%$ using first-order and second-order statistical features.

A $6 \times 6$ ROI (region of interest) is considered around the center of the marked regions to extract the aforementioned first-order and second-order statistical features. In the MFL data from 30 different seamless pipes, we had 6681 ROIs for flaw and 17417 ROIs for SPN, for the case of flaw depths greater than $20\%$ of the pipe wall thickness. For the other case, where the flaw depths are greater than $10\%$ of the pipe wall thickness, we had 10586 ROIs for flaw and 13512 ROIs for SPN. Hence, for automatic detection of flaw signals, we have sufficient data to characterize both SPN and flaw signals. Before computing the statistical features, we normalize the ROIs so that they have same mean and standard deviation. The first-order and second-order features (with $d = 1$, $\theta = 0^o, 45^o, 90^o, 135^o$), a total of $6 + 8 \times 4 = 38$, are extracted for the normalized ROIs. Since we are not sure about the orientation of textures in these regions, we computed texture features for all four orientations. The mean and standard deviation are not considered for the classification task as they are same for all the ROIs.

A support vector machine (SVM) [5] based classifier is used to judge the performance of the above computed features. The data set of 24098 ROIs is randomized and $70\%$ of the data set is used for training the classifier. The remaining $30\%$ is used in testing. Care is taken that the train and test data sets are independent and the classifier has not seen any of the test data while training. Figure 2 shows the receiver operating characteristic (ROC) curves for cases 1 and 2, where flaw signals of depth $> 10\%$ and $> 20\%$ are considered. It can be seen from Figure 2 that GLCM statistics perform better than first-order statistics, which is obvious because these statistics are based on second-order distribution of data. The second-order statistics provide more information than the first-order statistics and hence the better performance. The ROC curves shown in Figure 2 are computed on the data set, which has significantly more sample data points for SPN than flaw

| Features | Depth $> 20\%$ | | Depth $> 10\%$ | |
|---|---|---|---|---|
| | $A_a$ | $A_s$ | $A_a$ | $A_s$ |
| First-order | 89.17 | 63.33 | 81.18 | 77.36 |
| Second-order | 97.29 | 96.96 | 92.67 | 92.55 |

**Table 1**. Overall classification accuracy (%) for flaw depths $> 10\%$ and $> 20\%$ using first-order and second-order statistical features. $A_a$ and $A_s$ represent the accuracies with actual data and sub-sampled data respectively.

signal. Since the data is skewed, the experiment is conducted by sub-sampling the data points corresponding to SPN and making it equal to that of flaw response for both cases of 1 and 2.

Table 1 shows the area under ROC curves, which gives the overall accuracy of the above-mentioned features using actual and sub-sampled data for both cases of 1 and 2. Overall, the second-order features provide $97.29\%$ and $92.67\%$ recognition rates for flaw depths greater than $20\%$ and $10\%$ of the wall thickness of the pipe respectively. The high discriminability performance provided by texture-based features confirms the existence of different textural characteristics for SPN and flaw response and hence, justifies the usage of textural content as a measure for automatic flaw detection in MFL data.

## 4. DISCUSSION & CONCLUSION

We have proposed a texture analysis approach for automatic flaw detection in magnetic flux leakage data from seamless pipes. The method involves characterizing the textural content of seamless pipe noise and flaw signal for automatic flaw detection using texture classification. The proposed method provides a flaw detection rate of $97.29\%$ for flaw depths greater than $20\%$ of the wall thickness of the pipe, which confirms the existence of different textural characteristics for SPN and flaw signal. This algorithm can be incorporated in automatic analysis tools designed for classification and characterization of MFL data, because of its computational efficiency and data independence capabilities.

## 5. REFERENCES


[1] T. A. Bubenik et al., "Magnetic Flux Leakage (MFL) Technology for Natural Gas Pipeline Inspection," Report Number GRI-91/0367, Gas Research Institute, Battelle, November 1992.

[2] M. Afzal, S. Udpa, L. Udpa, and W. Lord, "Rejection of Seamless Pipe Noise in MFL Data obtained from Gas Pipeline Inspection," in *Review of Progress in Quantitative Nondestructive Evaluation*. 2000, vol. 19B, pp. 1217–1225, Plenum, New York.

[3] M. Afzal, R. Polikar, L. Udpa, and S. Udpa, "Adaptive Noise Cancellation Schemes for Magnetic Flux Leakage Signals obtained from Gas Pipeline Inspection," in *Proc. IEEE ICASSP*, Utah, USA, 2001, pp. 3389–3392.

[4] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classi£cation," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, November 1973.

[5] C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[6] R. M. Haralick, "Statistical and Structural Approaches to Texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.