Feature extraction and selection for defect classification of pulsed eddy current NDT
Tianlu Chen\textsuperscript{a,b}, Gui Yun Tian\textsuperscript{a,*}, Ali Sophian\textsuperscript{a}, Pei Wen Que\textsuperscript{b}

\textsuperscript{a} School of Computing and Engineering, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK
\textsuperscript{b} Department of Instrument Science and Engineering, Shanghai Jiaotong University, Shanghai 200240, China

\textbf{Article info}

Article history:
Received 12 July 2007
Received in revised form
12 November 2007
Accepted 18 February 2008
Available online 4 March 2008

Keywords:
Pulsed eddy current
Hilbert transform
Feature extraction
Feature selection
Defect classification

\textbf{Abstract}

Pulsed eddy current (PEC) is a new emerging nondestructive testing (NDT) technique using a broadband pulse excitation with rich frequency information and has wide application potentials. This technique mainly uses feature points and response signal shapes for defect detection and characterization, including peak point, frequency analysis, and statistical methods such as principal component analysis (PCA). This paper introduces the application of Hilbert transform to extract a new descending feature point and use the point as a cutoff point of sampling data for detection and feature estimation. The response signal is then divided by the conventional rising, peak, and the new descending points. Some shape features of the rising part and descending part are extracted. The characters of shape features are also discussed and compared. Various feature selection and integrations are proposed for defect classification. Experimental studies, including blind tests, show the validation of the new features and combination of selected features in defect classification. The robustness of the features and further work are also discussed.

\textcopyright 2008 Published by Elsevier Ltd.

1. Introduction

Pulsed eddy current (PEC) sensing is a new and emerging technique \cite{1,2} that has been particularly developed and devised for sub-surface crack measurements, crack reconstruction, and depth estimation \cite{3}. In contrast to conventional, multi-frequency or broadband eddy current techniques, where a single, multiple, or a broadband frequency sinusoidal field excitation is used, PEC techniques use pulsed excitation that is characterized by the richness of frequency contents. This is thought to be potential in bringing up information about the testing condition \cite{4–6}. Generally, the peak time and peak value are the main features used (the peak point P in Fig. 1). The former is related to the depth of the defect, whereas the latter corresponds to the size of the defect \cite{7–9}. Tian and Sophian \cite{4} have proposed and analyzed a new feature, termed as the rising point related to the propagation time of electromagnetic waves in metallic targets for defect classification. They also proposed a two-step framework for defect classification and quantification by using adopted features from principal component analysis (PCA) and wavelet analysis. However, a training data set for eigensignals or PCA basis is required, which is generated before the actual inspection is carried out \cite{10}. Nowadays, it has gained more and more attention in research and development of eddy current in nondestructive testing (NDT) from multilayer structures to on-line monitoring \cite{2,3}. These features are far from enough for fast and accurate defect classification and quantification in real time \cite{11}.

This paper applies the Hilbert transform to build up an analytic representation from a PEC differential response and then extracts a descending point feature. Based on the response segmentation by the rising, peak, and descending points, slope, derivative, and curvature of the rising part and the descending part are calculated and some shape features are extracted and selected for robust and fast identification of different defects.

The rest of the paper is organized as follows. Section 2 discusses the extraction of the descending point feature and the segmentation of the response. Some shape features of the rising and descending parts are proposed and explained. Section 3 introduces the experimental setup and discusses experimental results about response segmentation and feature extraction, performance comparison, selection, and integration for defect classification and finally Section 4 is on conclusions.

2. New fast feature extraction

PEC data analysis is mainly carried out in the time domain. The response signal with the system probe located on defect-free area is taken as a reference response. A differential response is obtained by subtracting the reference response from the testing
sample response, and this differential response is usually used for feature extraction as well as for defect classification and quantification. Features commonly used are the peak value and for feature extraction as well as for defect classification and sample response, and this differential response is usually used for feature extraction as well as for defect classification and quantification. Features commonly used are the peak value and

In this paper, another time domain feature: arrival time of the descending point (TD) is extracted by Hilbert transform and analytic representation. By using the three time domain points: rising point (R), peak point (P), and descending point (D), a PEC differential response can be divided into four parts. Some shape features focusing on the rising part and the descending part, are extracted. Fig. 1 shows a typical PEC reference signal, response signal, differential response, the three feature points (R, P, and D), and the rising and descending parts for shape features extraction. The signal pre-processing, feature extraction, and integration for defect characterization process can be summarized as in Fig. 2.

2.1. Descending feature point extraction and response segmentation

In signal processing, an analytic signal, or analytic representation $S_o(t)$ of a real-valued PEC differential response $S(t)$ is defined by [12]

$$S_o(t) = S(t) + jH[S(t)]$$

(1)

$$H[S(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{S(t)}{t-t'} dt'$$

(2)

$H[S(t)]$ is the Hilbert transform of $S(t)$ and $j$ is the imaginary unit. The analytic representation of a differential response facilitates mathematical manipulations and as discussed below, it makes certain attributes of $S(t)$ more accessible. Fig. 3 shows a differential response (the real part of an analytic signal), its Hilbert transform (the imaginary part of an analytic signal), and the modulus ($\sqrt{H^2[S(t)] + S^2(t)}$) of the analytic signal. There is a trough point in the modulus of the analytic signal which is taken as the descending point of the corresponding differential response. The arrival time of this point is taken as the new descending point feature (TD).

In mathematics, the Hilbert transform of signal $S(t)$ is obtained by convolving signal $S(t)$ with $1/j\omega$, whose frequency response is $-j \text{sgn}(\omega)$, where $\text{sgn}(\omega)$ is the signum function and $\omega$ is the angular frequency [12]. The basic idea of the analytic representation is that negative frequency components of the Fourier transform of a real-valued function are superfluous, due to spectral symmetry [13]. This can be expressed in terms of the Fourier transforms of $S(t)$ and $S_o(t)$, respectively, denoted by $S(\omega)$ and $S_o(\omega)$:

$$S_o(\omega) = F[S(t) + jH[S(t)]]$$

$$= S(\omega) + j[-j \text{sgn}(\omega)S(\omega)]$$

$$= \begin{cases} 2S(\omega), & \omega > 0 \\ S(0), & \omega = 0 \\ 0, & \omega < 0 \end{cases}$$

(3)

Therefore, the analytic signal $S_o(t)$ comprises only the nonnegative frequency components of $S(t)$ with physical significance and some hidden attributes are easily to access.

Actually, the descending point feature TD denotes the time when $\sqrt{H^2[S(t)] + S^2(t)}$ is the smallest and the response changes slightly after this point, with little useful information. Similarly, the response is almost constant before the rising point proposed by Tian and Sophian [4]. Since the differential response indicates the difference between the response and the reference signal, the rising and descending points are the starting and ending points of the difference. Therefore, the signal segments between these two points contain most information for defect evaluation. The descending points can be used for a criterion of decision on sampling numbers for real-time digital signal analysis and defect characterization. In other words, the descending point can be used as the cutoff point in real-time detection.
Table 1
Shape features and their explanation

<table>
<thead>
<tr>
<th>Feature name</th>
<th>Description</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS</td>
<td>Slope of the line between the starting and peak points of the rising part</td>
<td>Linear changing speed of the rising part</td>
</tr>
<tr>
<td>DS</td>
<td>Slope of the line between the peak and ending points of the descending part</td>
<td>Linear changing speed of the descending part</td>
</tr>
<tr>
<td>TPDER</td>
<td>Peak time of the rising part first derivatives</td>
<td>Time when the maximum change occurs of the rising part</td>
</tr>
<tr>
<td>TTDER</td>
<td>Trough time of the descending part first derivatives</td>
<td>Time when the maximum change occurs of the descending part</td>
</tr>
<tr>
<td>RCUR</td>
<td>Mean of the rising part curvatures</td>
<td>Averaging curvature of the rising part</td>
</tr>
<tr>
<td>DCUR</td>
<td>Mean of the descending part curvatures</td>
<td>Averaging curvature of the descending part</td>
</tr>
</tbody>
</table>

2.2. Shape feature extraction

As discussed above, the response segments between the rising and descending points can be named as two parts by the peak point: the rising part between R point and P point and the descending part between P point and D point. These two parts are investigated in detail and three shape features are extracted respectively (listed in Table 1) after noise removing e.g. Gaussian (low-pass) filtering.

The slope of a part, which is the ratio of the change of amplitude to the change of time between the starting and ending points of the part, indicates its changing rate. RS (RS = (\(V_p - V_k\))/(\(T_{p} - T_{k}\)) and DS (DS = (\(V_p - V_k\))/(\(T_{p} - T_{k}\)), representing the slope of the rising and descending part, respectively, are taken as shape features.

The first derivatives D(n) of the de-noised differential response Y(n) are calculated according to Eq. (4), which indicates how quickly the magnitude changes during the time interval of every two sample points. The maximum point of the rising part and the minimum point of the descending part indicate the time of largest change. Therefore, the arrival times of peak point (TPDER) and trough point (TTDER) are also taken as shape features.

\[
D(n) = \frac{dY}{dt} = \frac{Y(n) - Y(n-1)}{\Delta t}. \tag{4}
\]

In addition, average curvature \(R = \frac{\Delta \phi}{\Delta S}\), a mathematical parameter indicating the curving extent of a curve is the ratio of the change of the tangent angle \(\Delta \phi\) to the change of arc length \(\Delta S\) between the starting and ending point of a curve. It is useful in expressing the shape trend of a curve.

On the basis of first and second derivatives, the curvatures of each sample point k(n) can be calculated using Eq. (5), where Y\(^\prime\) and Y\(^\prime\prime\) are the first and second derivatives of Y(n), respectively [14]. These curvatures indicate how close to circular every part of the curve actually is. Considering the different sample points of each part for different responses, the means of curvatures (RCUR and DCUR for the rising and descending part, respectively) are taken as shape features indicating the curving extent of the region.

\[
k(n) = \frac{d\phi}{ds} = \frac{|Y^\prime\prime(n)|}{(1 + |Y^\prime(n)|^2)^{3/2}}. \tag{5}
\]

3. Experimental setup and test results

In NDT, it is desired to detect, classify, and quantify defects that may occur in metal structures. In such structures, defects mainly occur as surface cracks, sub-surface cracks, and hidden corrosion. For this requirement, we have developed a PEC system based on Hall-effect device. The excitation coil has 40 turns with inner and outer diameters of 17 and 26 mm, respectively.

We have also applied two aluminum samples for evaluation. In NDT, the typical crack width is smaller than 0.2 mm, which is difficult to calibrate at the scale. To evaluate the proposed method, 13 defects with distinguished features were produced for the study. One is to test thickness variation (thickness ranges from 1 to 10 mm, with 1 mm step), which can be used for metal loss simulation, and the other for surface and sub-surface defect detection and quantification (four surface and four sub-surface slots with different depths). The samples can be seen in Fig. 4a. For sub-surface crack simulation, we place the probe next to the surface of the sample shown in Fig. 4b, and for surface crack, we place it next to the other side. All the responses are sampled and recorded by a 12-bit analog to digital converter (ADC) card, while the sampling rate is 1 MHz.

Thirteen defects of specimen 1 and 2 are detected using our PEC system (Fig. 4c). Thirteen responses are collected where responses 1–5 are metal loss from 1 to 5 mm, responses 6–9 are surface defects, and responses 10–13 are sub-surface ones. The response detected at 10 mm thickness is taken as the reference signal with the assumption that the normal, defect-free Al slab has a depth of 10 mm. Thirteen differential responses can be obtained by subtracting the reference signal from other base signals.

3.1. Rising and descending point feature extraction

Hilbert transform is applied on the above 13 differential responses and corresponding analytic signals are built up. Fig. 5 shows the typical analytic signal modulus of metal loss, surface, and sub-surface defects. The trough point is the descending point and the arrival time of this point is the new feature TD.

The rising point features (TR) of each response are also extracted according to [4]. Fig. 6 shows the feature TR and TD of PEC differential responses obtained from the 13 defects. X-axis here is the defect number (1–5 are metal loss, 6–9 are surface defects, and 10–13 are sub-surface defects). In agreement with [4], the feature TRs of metal loss defects (Fig. 6a) vary with thickness. The larger the thickness is, the larger TR is. The TRs of sub-surface defects vary slightly with defect depths. All TRs of surface defects are the same and are smaller than that of the other two types. This feature is probably the best one to separate the metal loss defects from the other types. On the other hand, the feature TDs (Fig. 6b) of all the three types of defects vary with defect depths and the TDs of surface defects are also smaller than that of the other two types. This feature can separate the surface defects from the other two types easily. However, the two features cannot identify metal loss and sub-surface defects.

3.2. Shape feature extraction

Differential response is divided into four parts using point R and D as well as the peak point P. The corresponding shape features for the rising and descending parts are extracted and investigated, as illustrated in the upper two sub-figures of Fig. 7. Fig. 8 shows the slope features (RS and DS) of different defects of the 13 PEC differential responses. The performances of these two features are similar (approximately symmetrical about X-axis) and their correlation coefficient is −0.94. It can be seen from the figures that the surface defects (d6–d9) always vary contrarily to the other two types and their feature values are all larger or smaller than that of the other two types. Therefore,
surface defects can be identified easily from the other two types by these two features, while feature RS is more effective in classification since surface defect response rises very quickly. However, again, metal loss and sub-surface defects are difficult to be classified.

Gaussian filter or the empirical mode decomposition (EMD) method is used to remove noise and first derivatives are calculated out [15]. Fig. 7 shows the de-noised rising and descending parts of a 1 mm metal loss defect and its first derivatives. The arrival time of the peak point and trough point of the first derivatives is the feature TPDER and TTDER, respectively.

Fig. 9 shows the TPDERs (for the rising part) and TTDERs (for the descending part) of the 13 PEC differential responses. Noticeably, the performances of these two features are also similar and their correlation coefficient is 0.9816. For metal loss defects (d1–d5), they increase with the defect number, while for surface (d6–d9) and sub-surface defects (d10–d13), they both vary slightly. In addition, the features of surface defects are all smaller than that of the features from the other two types of defects. Therefore, these two features can separate metal loss from the rest of defects, which are complementary to previous features. However, the sensitivities for different surface and sub-surface defects are low.

Fig. 10 shows the feature RCUR and DCUR of differential responses obtained from the 13 defects. As expected, the performances of these two features are also similar, with a correlation coefficient of 0.9295. For surface defects, RCURs are all larger than 0.01, while DCURs are all larger than 1.7E–03. Therefore, surface defects can be separated easily from the other two types of defects by the features. Furthermore, the RCUR and DCUR features have good sensitivities for same types of defects.
Six shape features (three for the rising part: RS, TPDER, and RCUR; three for the descending part: DS, TTDER, and DCUR) of pulse eddy current differential responses are extracted. All these features can separate surface defect from the other two types in different variations and some of them can separate metal loss from the rest of defect types. The behavior of features at descending parts is similar to that at rising parts in terms of feature behaviors for defect classification and characterization. Considering the redundancy of information, it can be concluded that rising parts alone may be enough for defect classification and characterization, which is important for dynamic, real-time monitoring without capturing response data at descending parts.

3.3. Defect classification by feature selection

As reported in [10], the defect points in conventional peak time (PT)–peak value (PV) coordinate are dispersive. Most sub-surface and metal loss defects have close values and so are difficult to classify. However, by feature integration, classification results will be improved.

As discussed in Sections 3.1 and 3.2, TR, TPDER, TTDER, PT, and PV can separate metal loss from the other two types to different extents, while TD, PT, PV, and shape features can separate surface defects from the rest. Therefore, improved classification results can be obtained when combining more features, especially those with different characters or from different signal parts together properly. A good combination should satisfy two basic criteria: (1) defects of same class should go together and (2) the distance between each class is large enough. We tried all the possible combinations and found out some good ones. One is PV–TR–RCUR (Fig. 11a), integrating the information of the peak value, the rising time, and a shape feature of the rising part. The distance between classes is large enough and the layout of the elements from same class is well regulated according to the depth or thickness of the specimen. Another good combination is PV–TR–RS (Fig. 11b), integrating the information of shape and point features of the rising part. This combination has not only improved classification performance but it also has higher speed than other feature combinations. All the three features can be obtained at the moment peak point appears and no transforms or derivatives are needed (light computation burden).

The differential responses in PV–TR–DS coordinate are shown in Fig. 11c, which is similar to Fig. 11b. This will also help us to conclude that the characters of features from descending parts are similar to features from rising parts.

3.4. Blind test

In order to evaluate the identification robustness and accuracy of the proposed method and features (Fig. 12), additional three specimens with through-wall holes, vertical slots with different depths, and angle slots are added. PEC differential responses are obtained by an excitation coil and a giant magnetoresistance (GMR) sensor, another kind of widely used magnetic sensor (bigger and more sensitive). In addition to the traditional 2–5 mm diameters of the applied inductive sensors used in NDT nowadays, we also test sensor array probe with a big excitation coil [3]. The coil has 14 mm inner diameter, 16 mm outer diameter, and 1 mm height. The sampling frequency is also 1 MHz. Responses of the...
above 13 defects and new added ones are analyzed under the assumption that all the defect types are unknown. The fast 3D combination mentioned above is used to identify which type each defect belongs to.

Figs. 13 and 14 show the blind test results by the new sensor. The responses connected by solid lines are detected from 1–4 mm metal loss, 2, 4, 6, and 8 mm surface, and 1, 2, 3, and 4 mm sub-surface defects. Fig. 13 illustrates the robustness of the descending points. Fig. 14 shows the classification result by using the PV–TR–RS feature combination. The result is similar to, but better than, Fig. 11a since the feature TR of sub-surface responses detected by the new sensor is more sensitive to defect depth. The other five response points are newly added defects, including a vertical slot with 20 mm depth and 2 mm width, a through-wall hole with 3 mm diameters and 2.5 mm depth, an angle slot (45°), a 5 mm sub-surface defect, and a 2.5 mm metal loss (the blank space of specimen 4). They can all be classified into correct types easily.

3.5. Liftoff and feature robustness

An important factor limiting the PEC differential responses interpretation is the liftoff (the probe-to-specimen distance) influence [16]. When there is liftoff, the magnitude of the response increases and the rising point of the response appears earlier since the liftoff influence has the same effect as surface defects [4]. Sometimes, especially for metal loss and sub-surface
defects, two peaks appear, where the earlier portion corresponds to liftoff and the posterior portion indicates thickness or depth of defects [17]. Fig. 15 shows the 3 mm metal loss differential response with 0, 0.1, and 0.2 mm liftoff values. Obviously, the combination of liftoff and metal loss causes two peaks in the differential response.

Fig. 16 shows the performance of feature TD under different liftoff values (defects 1–4 are metal loss, 5–8 are surface, and 9–12 are sub-surface defects). For metal loss and sub-surface defects, the larger the liftoff is, the smaller the TD is. For surface defects, the larger the liftoff is, the slower TD increases with the depth of the defect. Although this feature varies with liftoff, it is easy to separate surface defects from the rest. More influences of liftoff will not be discussed in this paper. However, for quantitative NDE, normalization pre-process or invariant features are also required to remove the affects of liftoff [8,10,16].

4. Conclusions

A new time domain feature, termed as arrival time of descending point (TD) of PEC differential response, is extracted by Hilbert transform and analytic representation. The unique descending point can be used for determination of sample data numbers (cutoff point) for defect characterization. A response can be divided into four parts by the rising, descending, and peak points, while the rising and descending parts contain most useful information.

Six shape features (three for the rising part: RS, TPDER, and RCUR; three for the descending part: DS, TTDER, and DCUR) are extracted and explained. The performances of these features are compared and illustrated. Shape features from the descending part have similar behaviors as that of the rising part. For fast detection and classification, data after
Fig. 11. Defects in different feature combination coordinates. (a) PV–TR–RCUR coordinate. (b) PV–TR–RS coordinate. (c) PV–TR–DS coordinate.

Fig. 12. Newly added specimens for blind test. (a) Specimen 3: crack detection. (b) Specimen 4: Through-wall hole and metal loss detection. (c) Specimen 5: crack detection.

Fig. 13. Blind test of feature TD.
the peak point may not require data analysis and defect classification.

Different features reflect different aspects of information about the testing condition. The performance of 3D defect classification by combining different kinds of features, particularly response shape feature and point features, has proven to be much better than the conventional method using only peak features. The 'good' and 'fast' combinations are recommended and the performance similarity of rising and descending part features is validated. The effectiveness and robustness of features for defect classification, natural defects in particular, are validated further by experimental tests.

After response segmentation, different parts can be analyzed separately using different approaches or techniques. In the future, the method will be testified by real applications and the work will be extended to interpretation and quantitative analysis of different defect e.g. lift-off plus sub-surface defect, metal loss plus sub-surface defects [17,18] and multiple layers with air gap problem, where hidden corrosion is presented [19]. Two local peak points may be occurred in PEC signals, which we have observed in our robust experimental study in the Section 3.5 and other report [20]. Integration of time and spectrum features for the multiple peaks will be integrated to challenges these problems of ‘real’ specimen.

Acknowledgments

The authors would like to thank EPSRC, UK, and the Exchange Scholar Grant from the Shanghai Jiao Tong University China for funding the research work.

References


