Characterization of Defects Using Ultrasonic Arrays: A Dynamic Classifier Approach

Long Bai, Alexander Velichko, and Bruce W. Drinkwater

***Abstract*—In the field of nondestructive evaluation, accurate characterization of defects is required for the assessment of defect severity. Defect characterization is studied in this paper through the use of the ultrasonic scattering matrix, which can be extracted from the array measurements. Defects that have different shapes are classified into different defect classes, and this essentially allows us to distinguish between crack-like defects and volumetric voids. Principal component analysis (PCA) is used for feature extraction, and several representational principal component subsets are found through exhaustive searching in which quadratic discriminant analysis (QDA) and support vector machine (SVM) are used as the pattern classifiers. Instead of choosing a single optimal classifier, the best classifier is dynamically selected for different measurements by estimating the local classifier accuracy. The proposed approach is validated in simulation and experiments. In simulation, the depths (lengths of the minor axes) of 4441 out of 4636 test samples are measured accurately, and the measurement errors (with respect to the defect size) are below 10%. Arbitrarily shaped rough volumetric defects are identified as ellipses, which are reasonably good matches in shape to the original defects. Experimentally, six subwavelength scatterers are characterized and sized to within 0.14λ.**

I. Introduction

In the field of nondestructive evaluation, accurate characterization of defects is required because it provides us with quantitative information on the integrity of the structures being inspected [1]. When characterizing, the aim is to determine the key information about the defect, including size, orientation, and shape. The shape of a defect is important because different types of defects can have completely different effects on the mechanical performance of a structure and should be treated differently during structural integrity assessment. For example, cracks have a high severity because they have sharp edges which are stress concentrators and this leads to potentially rapid growth in structures [2]. Volumetric voids and inclusions do not have such severe stress concentrators and hence can often be regarded less hazardous. The current lack of quantitative characterization methods means that current practice is typically to assume a worst-case scenario, which is, for example, a crack oriented in the least favorable direction [3].

Manuscript received September 4, 2015; accepted September 20, 2015. This work was supported through the core research program within the UK Research Centre in NDE (RCNDE) funded by the Engineering and Physical Sciences Research Council (EPSRC; grant number EP/L022125/1).

The authors are with the Department of Mechanical Engineering, University of Bristol, Bristol BS8 1TR, UK (e-mail: lb13340@bristol.ac.uk).

DOI http://dx.doi.org/10.1109/TUFFC.2015.007334

The use of ultrasonic arrays [4] and advanced high-resolution imaging algorithms [5], [6] offers one way of characterizing defects. In particular, the size and orientation of crack-like defects were successfully determined by applying a certain threshold to the image pixel intensities [7]. Although such image-based characterization approaches work well for relatively large cracks, their performance degrades for small cracks (e.g., subwavelength cracks). To address this issue, it was shown that small crack-like defects could be characterized based on their scattering matrices either by measuring the half-width at half-maximum (HWHM) from the pulse–echo component [7], or by a systematic database searching approach [8]. For experimental data measured from a lack-of-fusion crack implanted within a stainless steel test sample, the sizing performance was improved by averaging the results obtained at multiple frequencies [9]. In this paper, we extend the defect database approach to include both the crack-like defects and volumetric voids.

With the development of pattern recognition and machine learning theory [10], [11], advanced pattern classification algorithms have been applied to data obtained from ultrasonic measurements [12]–[19]. To date, the measured signals were ultrasonic A-scans (time-domain amplitude plots), from which features were extracted using discrete wavelet transforms [12], [17], [18] and Hilbert and Fourier transforms [17]. Artificial neural networks (ANN) and the support vector machine (SVM) approach were among the popular pattern classifiers, and it was shown that SVM outperforms the well-trained ANN in terms of classification accuracy and computational efficiency [12]. For the SVM classifier, three different kernel functions were compared, and the radial basis function kernel was shown to have the best performance for classifying different types of defects (delamination-like flaws, porosity, and inclusions) which were located at different positions in the test specimen [12].

The scattering matrix contains all of the information about the defect that can be extracted during ultrasonic testing [7], and hence is the logical choice of data to form the starting point of any defect classification scheme. Feature extraction is often used to extract the key information from the data and hence allow the measurement to be more robust to uncertainties [20]. Here we employ principal component analysis (PCA) [21] for this task and use it to compress the scattering matrix into a smaller number of principal components. The goal is to classify defects into different defect classes which are related to the defect shape. For the classification task, quadratic discriminant

analysis (QDA) and SVM [10] are adopted as the pattern classifiers, and several principal component subsets are obtained through an exhaustive searching approach. Then, the optimal classifier is dynamically selected based on the estimation of the local classifier accuracy. We note that although here we use a specific combination of PCA and the QDA/SVM classifiers, the approach is general to other combinations of feature extraction and pattern classification algorithms.

This paper is organized as follows. The scattering matrix and the studied problem are introduced in Section II. Principal component analysis and the statistical pattern classification approaches used in this paper are briefly introduced in Section III. Section IV presents a step-by-step description of the proposed defect characterization strategy. The simulation and experimental results are discussed in Sections V and VI.

II. The Scattering Matrix and the Database

In two dimensions, the far-field scattering matrix is defined as [7]

$$S(\theta_1,\theta_2,\omega)=\frac{a_{\rm sc}(\omega)}{a_{\rm in}(\omega)}\sqrt{\frac{d_{\rm sc}}{\lambda}}\exp\left(-\frac{i\omega d_{\rm sc}}{c}\right), \qquad (1)$$

where θ_1 and θ_2 are the incident and scattering angles, $a_{\rm in}$ is the amplitude of the plane incident wave at the defect, $a_{\rm sc}$ is the amplitude of the scattered wave, $d_{\rm sc}$ is the distance from the defect where the amplitude $a_{\rm sc}$ is measured, λ is the wavelength, c is the speed of the ultrasonic wave, and ω is the angular frequency. Only the amplitude of the scattering matrix (of the longitudinal waves) is considered here, and the phase [i.e., the exponential term in (1)] is ignored hereafter. As can be seen, the scattering matrix describes how waves at any incident angle are scattered. For this reason, it encodes all the information obtained from a scatterer, assuming linear acoustics apply and measurements are made in the far field of the defect.

The primary objective of this paper is to classify defects into different defect classes which are related to the defect shape. This is important because it essentially allows us to distinguish between crack-like defects and volumetric voids of the same size. Crack-like defects are the worst-case scenario of all defects in terms of their effects on component life, and their characterization was studied previously [7], [8]. However, not all defects in structures are crack-like, and the method proposed in [8] failed if the defect cannot be *a priori* identified as crack-like (e.g., for some rough cracks).

In this paper, scattering matrices of three types of 2-D defects (namely, cracks, ellipses, and holes) are simulated and categorized into different defect classes. Scattering matrices of cracks and holes are simulated using the fast semi-analytical technique proposed in [22] and the analytic solution in [23], respectively. Scattering matrices of ellipses are obtained using a finite element local scattering (FELS) model [24]. Figs. 1(a)-1(d) show the example profiles of a crack, two ellipses, and a hole, respectively, with their corresponding scattering matrices. The scattering matrices shown in Fig. 1 are obtained using the measurement configuration shown in Fig. 2. The material is assumed to be aluminum, with Young's modulus = 69 GPa, Poisson's ratio = 0.334, and density = 2700 kg·m^{-3}. θ_1 and θ_2 are defined with respect to the defect normal (positive if measured clockwise), and the maximum amplitude in pulse–echo (i.e., where $\theta_1=\theta_2$) is found at $\theta_1=\theta_2=0^\circ$ (except for the hole whose pulse–echo amplitude is a constant). For the crack and flat ellipse (i.e., $b=0.2\lambda$) shown in Figs. 1(a) and 1(b), this specular normal-incidence reflection is also the maximum amplitude of the scattering matrix, whereas the tip-diffraction amplitude (corresponding to $\theta_1=\theta_2=\pm 90^\circ$ and related to the effective cross-sectional area of a defect) is significantly lower. On the other hand, for the more rounded ellipse and hole shown in Figs. 1(c) and 1(d), the maximum amplitude is found at $\theta_1=-\theta_2=\pm 90^\circ$, which means that the backscatter is greater than the specular reflections. The scattering matrix of a hole is only a function of $\theta_2-\theta_1$, which can be used to differentiate the holes from other defects. In general, the range of θ_1 and θ_2 is determined by both the positions of the array elements and that of the defect, hence the scattering matrix extraction described here is conducted after the preceding defect localization step [25]. It is reasonable to assume that as the angular range de-

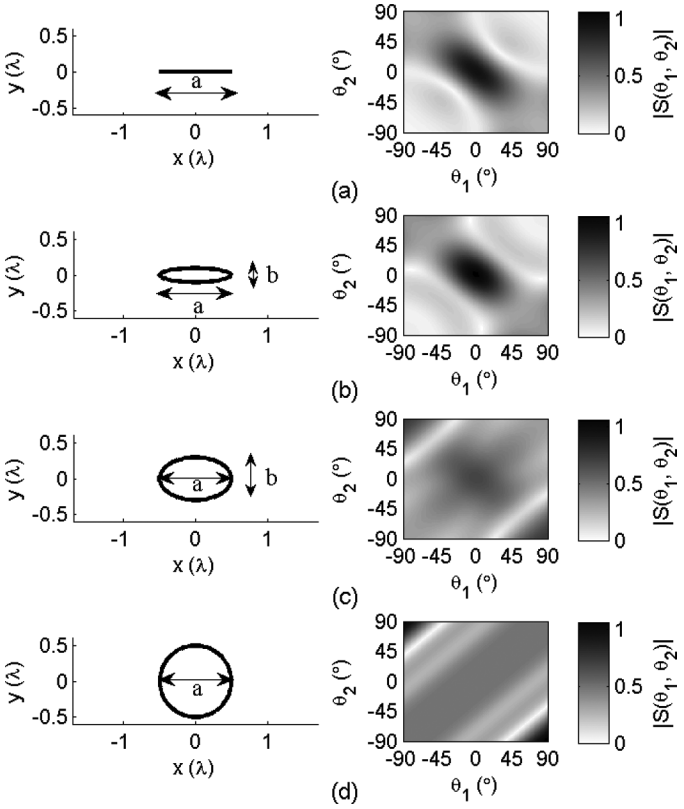

Fig. 1. Profiles of four defects and the corresponding scattering matrices. (a) a 1λ crack, (b) an ellipse with $a=1\lambda$, $b=0.2\lambda$, (c) an ellipse with $a=1\lambda$, $b=0.6\lambda$, and (d) a hole whose diameter is 1λ.

TABLE I. DEFECT CLASS DEFINITIONS.

Defect class	Defect type	Aspect ratio
1	Crack	0
2	Ellipse	0.1
3	Ellipse	0.2
4	Ellipse	0.3
5	Ellipse	0.4
6	Ellipse	0.5
7	Ellipse	0.6
8	Ellipse	0.7
9	Ellipse	0.8
10	Ellipse	0.9
11	Hole	1.0

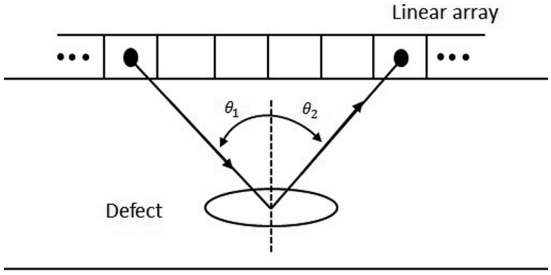

Fig. 2. Experimental configuration of the scattering matrix extraction measurement. θ_1, θ_2 represent the incident and scattering angles, and are defined with respect to the defect normal (positive if measured clockwise).

creases, the ability to characterize will also decrease. However, Fig. 1 represents an extreme case where the incident and scattering angles range from -90° to 90°. The problem is first studied in simulation based on this idealized configuration, and a more practical situation in which the experimentally measured scattering matrices have a much smaller angular coverage is studied in Section VI.

It can be seen from Fig. 1 that for a given defect size, the scattering matrix is related to the shape (i.e., aspect ratio) of a defect. Here, we define the aspect ratio of a defect as the ratio of the minor axis (b in Fig. 1) to the major axis (a in Fig. 1). Note that the severity of a defect is related to its aspect ratio [e.g., the aspect ratio of 0 (for cracks) and 1 (for holes) represent defects with highest and lowest severity, respectively.] Furthermore, defects with the same aspect ratio show similar scattering behavior (a remarkable example of this is the holes which are isotropic scatterers). For these reasons, the defects are classified into different classes (see Table I) according to their aspect ratios. For each defect class, scattering matrices of defects that have sizes from 0.2λ to 2λ (at 0.1λ intervals) are simulated and used as the training samples. Note that the range considered focuses on small defects because that is where we believe this approach has the most benefit. The training set contains 209 samples (19 samples in each class), and is used to train the pattern classifiers which will be introduced in the next section. At the same time, a separate test set, which consists of the scattering matrices of defects with sizes from 0.16λ to 2.04λ (at 0.02λ intervals), and aspect ratios from 0 to 1 (at 0.02 intervals) is prepared. These 4636 test samples are not used to train the classifiers, and are prepared to evaluate the classifier performance. The scattering matrices in the database (training and test sets) are simulated at the frequency $f = 2$ MHz, and hence the database includes the scattering matrices of all possible cracks and volumetric voids which have sizes between 0.63 mm and 6.3 mm. Fig. 3(a) shows the distribution of training and test samples in a small region of the two-dimensional problem space. As can be seen from Fig. 3(a), the class boundary between the defect class i and $i + 1$ can be formulated as r(aspect ratio) $= 0.1i - 0.05$, $i = 1, 2, \ldots, 10$, and the test samples are assigned the true class label y accordingly. However, other choices exist for the definition of the class boundaries. For example, if more conservative classification results are preferred, the class boundary between the defect class i and $i + 1$ can be defined as $r = 0.1i - 0.01$, leading to the underestimation of the aspect ratio in some cases. The scattering matrices in the database are obtained from the same measurement configuration as shown in Fig. 2, which implies that all the defects are modeled to be horizontal (major axis is parallel to the array), and the incident and scattering angles cover the range $[-90^\circ, 90^\circ]$. However, the proposed strategy in this paper also works for 2-D angled defects as long as the range of θ_1 and θ_2 in the database is kept the same as that of the measured data. If the defect has an orientation angle of α (with respect to the array, positive if measured clockwise), then the maximum amplitude in pulse–echo will be at $\theta_1 = \theta_2 = \alpha$, and the scattering matrices in the database should have incident and scattering angles limited to $[-90^\circ - \alpha, 90^\circ - \alpha]$ instead of $[-90^\circ, 90^\circ]$. In practice, this could be achieved by moving the measurement array to obtain the specular reflection (i.e., the maximum response). We note that further work is required to characterize 3-D defects such as branched cracks.

III. FEATURE EXTRACTION AND STATISTICAL PATTERN CLASSIFICATION ALGORITHMS

Here we present the detailed description of the feature extraction and the classification methods used in this paper. These are reasonably well known by the detection and classification research communities. However, we present them again for completeness, and our specific implementation can be found in Section IV.

A. Feature Extraction by Principal Component Analysis

Scattering matrices normally have the size of $Q \times Q$, where Q is the number of array elements, and it is desirable to reduce the dimensionality of the input pattern to classifiers to avoid the "curse of dimensionality" [26]. Principal component analysis (PCA) is a widely used approach for dimensionality reduction of a data set [21], [27]. The basic idea is to find a new set of orthogonal directions (called the principal components, or PCs) with which the original data can be represented more efficiently. As

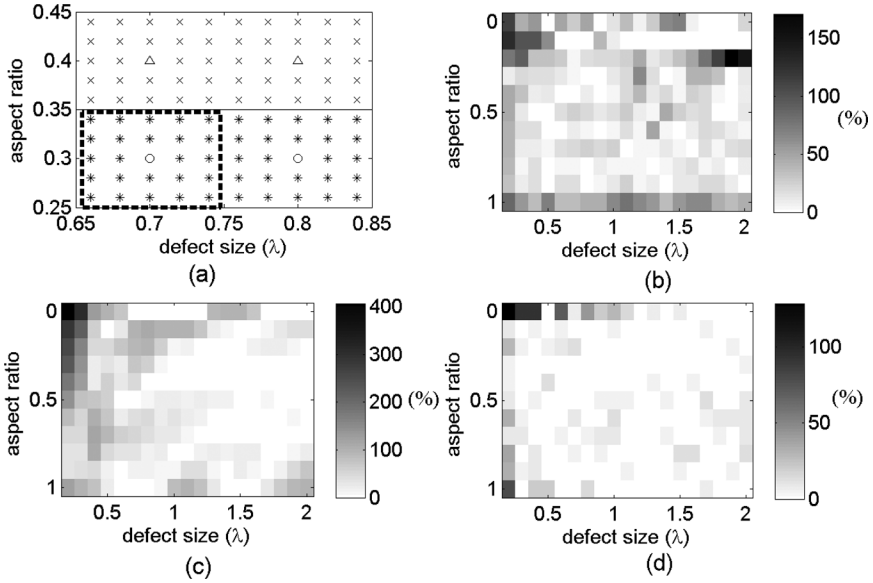

Fig. 3. Distribution of the training and test samples in the problem space, and the local weighted misclassification rates [WMRs, see (12)] of different individual classifiers. (a) Training (circles and triangles) and test (asterisks and crosses) samples of class 4 and 5, the ideal boundary between the two classes (solid line), and one example of the local region of a training sample (dashed box). (b) and (c) The local WMRs across the whole parameter space of the QDA classifier with 11 PCs and the SVM classifier with 7 PCs. (d) The local WMRs obtained with the proposed dynamic classifier selection approach. The overall WMRs in (b)–(d) are 21.48%, 37.90%, and 4.40%, respectively.

a result, projections of the data onto the first few PCs turn out to be sufficient to describe the same information included in the original high-dimensional data set. In addition, because the compressed data are resistant to high-dimensional noise, PCA has been successfully used as a denoising tool (e.g., for the noisy scattering matrices of crack-like defects [8]).

Consider the training set $\mathbf{S} \in \mathrm{R}^{Q^2 \times N}$, whose columns are the vectorized scattering matrices, and N is the number of the training samples. For the simulated scattering matrices in the database, Q is set to be 61, and θ_1 and θ_2 are sampled at 3° intervals within $[-90°, 90°]$. Suppose $\mathbf{S}$ is centered (the mean scattering matrix is subtracted from each column of $\mathbf{S}$), then the covariance matrix can be estimated by

$$\mathbf{C} = \frac{1}{N-1}\mathbf{S}\mathbf{S}^{\mathrm{T}}. \tag{2}$$

Eigenvalues of $\mathbf{C}$ can be calculated from the eigendecomposition

$$\mathbf{C} = \mathbf{E}\mathbf{D}\mathbf{E}^{\mathrm{T}}. \tag{3}$$

Most of the variation of the original training set is now captured by the first few PCs, which are the eigenvectors (columns of $\mathbf{E}$) corresponding to the largest eigenvalues (diagonal elements of $\mathbf{D}$). Let d_i be the eigenvalues (sorted in descending order); then, the variation explained by the first k PCs [21] is

$$v = \frac{\sum_{i=1}^{k} d_i}{\sum_{i=1}^{Q^2} d_i}. \tag{4}$$

In this paper, the number of PCs retained is selected so that v exceeds 99%. According to this criterion, the original Q^2 (i.e., $61^2 = 3721$) -dimensional training samples are typically compressed into 14-dimensional PC space, and these 14 PCs are used as features for the individual classifiers. We note that the number of PCs required will increase with defect size and frequency.

B. Quadratic Discriminant Analysis

After the application of PCA, the training set is categorized and represented by the training samples $\{\mathbf{x}_i, y_i\}$, $i = 1, 2, \ldots, N$. $\mathbf{x}_i \in \mathrm{R}^{14}$ is 14-dimensional representation of the training sample i in the PC space, and y_i is the corresponding class label ($1 \le y_i \le 11$). In statistical decision theory [10], Bayes classifiers work by classifying new data into the class which can minimize the expected classification cost:

$$\hat{y} = \underset{y=1,2,\ldots,M}{\operatorname{argmin}} \sum_{i=1}^{M} L(i,y) P(i|\mathbf{x}), \tag{5}$$

where $L(i, y)$ is the cost of classifying a sample whose true class label is i into class y and is set to be $|i - y|$ in this paper (i.e., the cost increases linearly with class error), $P(i|\,\mathbf{x})$ is the posterior probability of class i given $\mathbf{x}$, and M is the number of classes (11 in our classification problem).

Quadratic discriminant analysis (QDA) is based on the assumption that data in each class is generated from a multivariate Gaussian distribution, which is given as

$$P(\mathbf{x}|i) = \frac{1}{(2\pi)^{p/2}|\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu_i)^{\mathrm{T}}\Sigma_i^{-1}(\mathbf{x}-\mu_i)\right), \tag{6}$$

where i is the class label, $p = 14$ is the dimensionality of the input features, μ_i, Σ_i are the class mean and covariance matrix, and $|\Sigma_i|$ is the determinant of Σ_i. μ_i and Σ_i can be estimated from the training samples as

$$\begin{aligned} \mu_i &= \sum_{y_j=i} \mathbf{x}_j / N_i, \\ \Sigma_i &= \sum_{y_j=i} (\mathbf{x}_j - \mu_i)(\mathbf{x}_j - \mu_i)^{\mathrm{T}} / (N_i - 1), \end{aligned} \tag{7}$$

where N_i is the number of training samples that belong to class i. Then, $P(i|\mathbf{x})$ can be obtained from Bayes theorem [10] as

$$P(i|\mathbf{x}) = \frac{P(\mathbf{x}|i)P(i)}{\sum_{j=1}^{M} P(\mathbf{x}|j)P(j)}, \tag{8}$$

where the class prior probability is set to be equal for all the classes [i.e., $P(j) = 1/11$, $j = 1, 2, \ldots, 11$].

Although in practice the data distribution does not always follow the Gaussian distribution, QDA performs very well and is one of the most popular classifiers [11], because the simple quadratic decision boundaries are often reasonable choices for discriminating different classes in the feature space [10], [11].

C. The Support Vector Machine Classifier

The support vector machine (SVM) approach has been widely used in defect classification problems [12], [13], [28], [29], and was shown to have superior performance compared with classifiers based on ANNs [12]. For a binary classification problem where all the training samples belong to two classes ($y_i = 1$ or -1), the optimal separating hyperplane $f(x) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b = 0$ can be calculated from the following optimization problem [30]

$$\begin{aligned} &\min_{\mathbf{w},b,\xi_i} \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{N}\xi_i, \\ &\text{subject to } \xi_i \geq 0,\ y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \geq 1 - \xi_i,\ \forall i. \end{aligned} \tag{9}$$

In (9), C is the so-called cost parameter which defines the weights assigned to the training samples that are "not well inside their class boundary" [10]. The solution of this optimization problem can be obtained by solving its dual problem, and $\mathbf{w}$ can be expressed as [11]

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i. \tag{10}$$

In (10), only a small number of the coefficients α_i will be nonzero, and the corresponding training samples are called the support vectors [31]. This means that the decision boundary is determined by only those support vectors (the number can be much smaller than the size N of the training set). It is significantly different from the case of QDA, where the class mean and covariance matrix are estimated from all of the training samples.

For a new input $\mathbf{x}_t$, the class label output of the SVM classifier is given by $\mathrm{sgn}(\mathbf{w}^{\mathrm{T}}\mathbf{x}_t + b)$, and from (10), it is clear that the result is only related to the inner product of the features $\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_t$. Nonlinear decision boundaries can be easily formed when the original feature is mapped to some higher dimensional space through $\mathbf{x} \mapsto \phi(\mathbf{x})$. In this case, the decision boundary becomes $f(x) = \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}) + b = 0$, and the classification result is determined by the inner product $K_e(\mathbf{x}_i, \mathbf{x}_j) = \phi(x_i)^{\mathrm{T}}\phi(\mathbf{x}_j)$. K_e is called the kernel function, and its introduction has greatly improved the computational efficiency, because now the result can be obtained without explicitly calculating the mapping $\mathbf{x} \mapsto \phi(\mathbf{x})$ [32]. The radial basis function kernel was previously shown to have a superior performance to linear and polynomial kernels [12], and is used in this paper as the kernel function for SVM classification. It has an additional parameter γ, and is defined as

$$K_e(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma||\mathbf{x}_i - \mathbf{x}_j||^2), \quad \gamma > 0. \tag{11}$$

For the multi-class classification problem, the one-against-one [30] method was used in the SVM implementation.

IV. The Proposed Approach

After the training and test sets are prepared as described in Section II and the data are compressed by PCA, the proposed defect characterization approach can be conducted in four steps: PC subset selection, dynamic classifier selection, sizing of the defect and estimation of the class probability, and verification of the characterization result.

A. PC Subset Selection

To determine which specific combinations of the PCs are desirable for correct classification, feature selection (i.e., the wrapper method [33]) is conducted using QDA and SVM as predictors. For a given number of PCs, all possible PC combinations were evaluated according to a certain criterion, and the best PC combination was recorded. In this paper, the weighted misclassification rate (WMR) [34] is chosen as the criterion to minimize, and is defined as

$$\mathrm{WMR} = \frac{\sum_{i=1}^{N}|y_i - \hat{y}_i|}{N} \times 100\%, \tag{12}$$

where y_i and $\hat{y}_i$ are the true class label and predicted class label of the training sample i. Note that the WMR is calculated under the leave-one-out scheme here, which im-

TABLE II. OPTIMAL PC COMBINATIONS FOR QDA AND SVM CLASSIFIERS.

Number of PCs	Index of selected PCs (QDA)	Index of selected PCs (SVM)
2	1, 2	1, 2
3	1, 2, 5	1, 2, 12
4	1, 2, 4, 5	1, 2, 3, 12
5	1, 2, 3, 4, 5	1, 2, 3, 5, 12
6	1, 2, 3, 4, 5, 8	1, 2, 3, 4, 5, 12
7	1, 2, 3, 4, 5, 8, 12	1, 2, 3, 4, 6, 9, 12
8	1, 2, 3, 4, 5, 7, 10, 12	1, 2, 3, 4, 6, 10, 12, 14
9	1, 2, 4, 5, 6, 7, 8, 10, 12	1, 2, 3, 4, 5, 6, 10, 12, 14
10	1, 2, 4, 5, 7, 9, 11, 12, 13, 14	1, 2, 3, 4, 5, 6, 7, 8, 11, 12
11	1, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14	1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 14
12	1, 2, 3, 4, 5, 6, 7, 9, 11, 12, 13, 14	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13
13	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14
14	All PCs	All PCs

plies that when predicting the training sample i, it should be temporarily removed from the training set which is used to build the classifiers.

In the training phase of the SVM classifier, the parameters C [see (9)] and γ [see (11)] must be determined, and this is done by the grid search approach suggested in [30]. The 5-fold cross-validation WMR [see (12)] was calculated for different values of C and γ, and the optimal parameter values were found as those associated with the lowest WMR.

Table II lists the optimal PC combinations for QDA and SVM classifiers when the number of PCs is $2, 3, ..., 14$, and the corresponding WMR is shown in Fig. 4(a). Fig. 4(b) shows the distribution of the training samples in the optimal 2-dimensional PC space (found to be comprised of PC 1 and PC 2 for both QDA and SVM classifiers). From Fig. 4(a), it can be seen that the size of the best PC subset is 11 for the QDA classifier, and the WMR remains low when the number of PCs is larger than 9. For the SVM classifier, the WMR reaches minimum when 7 PCs are used but remains small for larger number of PCs. It can also be seen that the optimal QDA classifier performs better than the optimal SVM classifier, but as will be explained in the next subsection, this is mainly caused by a small number of the training samples that led to large misclassification errors. Note that consideration of the WMR values may allow some combinations of PCs to be discarded before dynamic classification.

B. Dynamic Classifier Selection

After the PC subset selection step, 26 PC subsets are obtained (13 for QDA and 13 for SVM, see Table II). Here, instead of selecting one optimal PC subset, all of the PC subsets are used as input to a dynamic classifier selection system [35]. This is based on the hypothesis that different classifiers could have different regions of expertise, and even the globally optimal classifier could perform poorly in some of the local regions. In this paper, the local region of a training sample is defined as the region containing the training sample itself and its nearest neighbors in the test set [see for example, the region represented by dashed box in Fig. 3(a)]. The size of the local region is 5×5 for the training samples of classes 2 through 10, and for the training samples of classes 1 and 11, their local regions have the size of 3×5 (test samples exist on only one side of the training sample). Then, the WMR of the local region can be calculated using the test samples within the region. Figs. 3(b) and 3(c) show the WMRs of such defined local regions across the whole parameter space, calculated from the best QDA classifier (using 11 PCs as features) and the best SVM classifier (using 7 PCs as features), respectively. From Figs. 3(b) and 3(c), it can be seen that the SVM classifier has the larger value of the overall WMR mainly because of the local regions around the 0.2λ crack (note that because of its definition, the WMR can exceed 100%). The idea of the dynamic classifier selection was

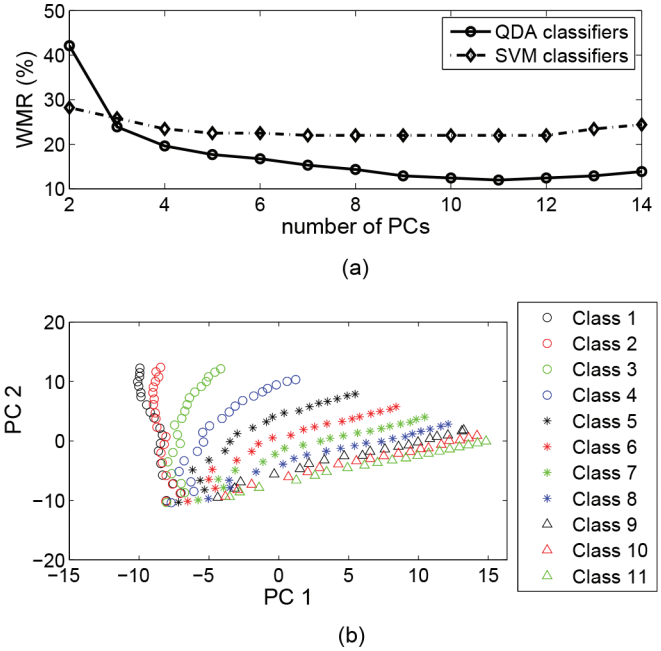

Fig. 4. Result of the PC subset selection: (a) the optimal WMR [see (12)] for different numbers of PCs, and (b) the distribution of the training samples in the optimal 2-dimensional PC space.

proposed in [35], and adopted in this paper to deal with classifier inconsistency and to choose the best classifier whose output is likely to have the highest accuracy.

Unlike the training samples, for which the local region can be determined as described earlier, given a measurement $\mathbf{S}_t$, its local region is not known *a priori*, and thus is defined here to be consist of its K nearest neighbors in the test set. The structural similarity (SSIM) index [36] is adopted as the similarity metric, and is defined by

$$\mathrm{SSIM}(\mathbf{S}_1,\mathbf{S}_2) = \mathrm{corr}(\mathbf{S}_1,\mathbf{S}_2)\frac{2\overline{\mathbf{S}}_1\overline{\mathbf{S}}_2}{\overline{\mathbf{S}}_1^2+\overline{\mathbf{S}}_2^2}\frac{2\sigma(\mathbf{S}_1)\sigma(\mathbf{S}_2)}{\sigma(\mathbf{S}_1)^2+\sigma(\mathbf{S}_2)^2}, \tag{13}$$

where corr represents the correlation coefficient, $\overline{\mathbf{S}}$ and $\sigma(\mathbf{S})$ are the mean and standard deviation of $\mathbf{S}$. The SSIM index takes into account both the amplitude and shape information of the scattering matrix, and the nearest neighbors are those that have the highest values of the SSIM index. The default size, K, of the local region is set to be 25 (this means that the size of the local region is the same for unknown data and for a training sample, except that the local region of a training sample contains the training sample itself).

Suppose the measurement $\mathbf{S}_t$ is classified by some classifier as class i, the local classifier accuracy can be estimated from the K nearest neighbors by [35]

$$a_i = \frac{n_i+1}{n_d+M}\times 100\%, \tag{14}$$

where n_d is the number of the test samples within the local region of $\mathbf{S}_t$ that are classified by the same classifier as class i, and n_i counts the number of the correct classifications among the n_d test samples. Laplace smoothing [37] is adopted in (14), and M is the number of the defect classes. Eq. (14) is used to estimate the local accuracy of each individual classifier, and the one with the highest local accuracy is selected as the optimal classifier for the characterization of $\mathbf{S}_t$.

C. Sizing of the Defect and Estimation of the Class Probability

After the optimal classifier is selected for $\mathbf{S}_t$ and it is assigned some class label i, the size of the defect is estimated by averaging the sizes of the n_i samples in the local region [see (14)] that have the same class label i and are correctly classified by the optimal classifier. Finally, the probability of this decision is given by (14), and the probability that the defect belongs to class j ($j \neq i$) is given by

$$a_j = \frac{n_j+1}{n_d+M}\times 100\%, \tag{15}$$

where n_j represents the number of the test samples which have the true class label of j among the n_d samples in the local region that are classified as class i by the optimal classifier.

D. Verification of the Characterization Result

Because of the generalization ability of the individual classifiers, which is one of the desirable properties of a good classifier, some defects that are significantly different in nature than those included in the database could potentially be incorrectly characterized by the proposed approach. Here we propose the following two verification conditions, and the characterization result can be safely accepted if both of them are met.

- Condition 1: The maximum amplitude of the measurement $\mathbf{S}_t$ should fall in the same range as those in the database.
- Condition 2: $\mathbf{S}_t$ should be similar (near) enough to the scattering matrices in the database. In other words, the SSIM index [see (13)] of $\mathbf{S}_t$ and its nearest neighbor should exceed a certain threshold (e.g., 0.8).

According to Condition 1, attempts to characterize defects larger (or smaller) than those in the database can be rejected. Condition 2 can be used to determine if the defect belongs to some unknown defect type (e.g., is porosity or an inclusion).

E. An Example Case of an Elliptical Defect

The working procedure of the proposed defect characterization approach is summarized here by applying it to an example case, where the measurement $\mathbf{S}_t$ [shown in Fig. 5(a)] is the scattering matrix of an ellipse (simulated using the FELS model [24]) whose size and aspect ratio are 1.42 mm (0.45λ when $f = 2$ MHz) and 0.67, respectively. The desired classification output for $\mathbf{S}_t$ is class 8 (corresponding to an aspect ratio of 0.7).

- Step 1: the first step includes the preparation of the training and test sets as introduced in Section II, feature extraction by PCA, and PC subset selection as described in Section IV-A. This learning step can be conducted off-line, and the resulting PC subsets which are used by QDA and SVM classifiers as input features (see Table II) can be used to classify any incoming data that has the same angular coverage as the scattering matrices in the database.
- Step 2: for each of the individual classifiers, calculate the classification output and estimate its local accuracy [see (14)]. Figs. 5(b) and 5(c) show the local region of $\mathbf{S}_t$ [using the SSIM index as the similarity metric, see (13)]. The test samples in Fig. 5(b) are classified by the QDA classifier with 3 PCs, and $\mathbf{S}_t$ is classified by the same classifier as class 7. Hence, the accuracy of this decision is estimated from the test samples that are classified as class 7 [marked as asterisks in Fig. 5(b)]. It can be seen that there are 18 such test samples, among which only 8 have the true class label of 7 (those having the aspect ratios of 0.62 and 0.64). According to (14), the local accuracy of this classi-

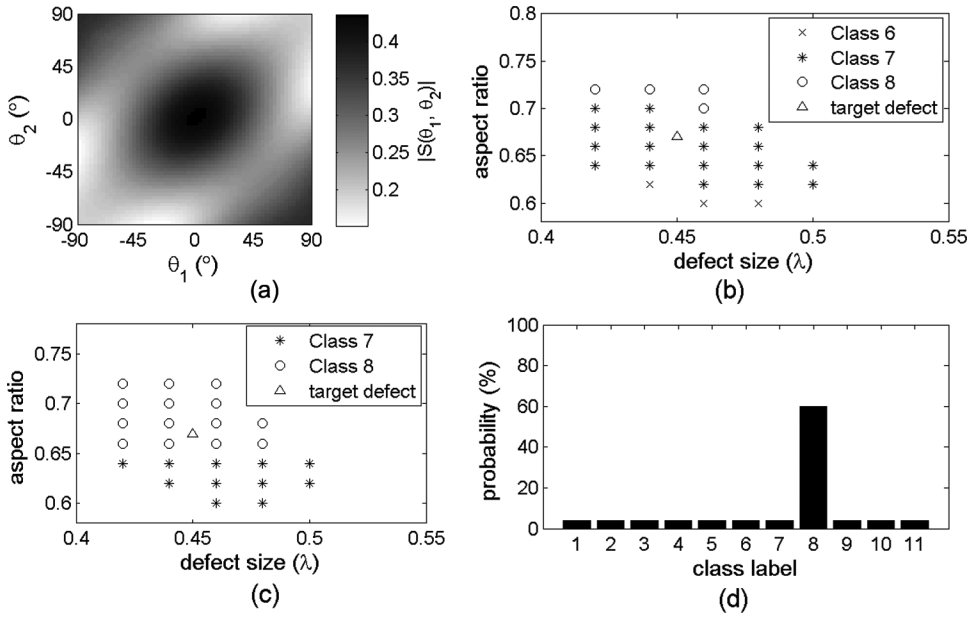

Fig. 5. Characterization of an ellipse of size 0.45λ and aspect ratio of 0.67: (a) the simulated scattering matrix of the ellipse, (b) the local region (K nearest neighbors in the test set), where the test samples are classified by the QDA classifier with 3 PCs, (c) the local region where the test samples are classified by the SVM classifier with 2 PCs, and (d) the distribution of the class probability obtained with (14) and (15).

fier is given as $(8 + 1)/(18 + 11) \times 100\% = 31.03\%$. On the other hand, the test samples in Fig. 5(c) are classified by the SVM classifier with 2 PCs, and $\mathbf{S}_t$ is classified by the same classifier as class 8. Because there are 14 test samples that are classified as class 8 [marked as circles in Fig. 5(c)] and all of them are correctly classified (their true class label is also 8 as they have aspect ratios between 0.66 and 0.72), the estimated accuracy of the classifier is $(14 + 1)/(14 + 11) \times 100\% = 60\%$. It is clear that the latter classifier has the higher accuracy between the two considered classifiers, and it turns out that it is indeed the optimal classifier for the classification of $\mathbf{S}_t$. As a result, the classification output of the dynamic classifier selection system for the measurement $\mathbf{S}_t$ is class 8.

- Step 3: the size of the defect can be obtained from the correctly classified (by the optimal classifier) nearest neighbors [marked as circles in Fig. 5(c)]—the average size of those test samples is 0.4457λ; therefore, this is the sizing result of the proposed approach for $\mathbf{S}_t$. Because none of the 14 test samples have a true class label other than 8, n_j [see (15)] is 0 for any $j \neq 8$, and the probability that $\mathbf{S}_t$ belongs to class j ($\neq 8$) is $(0 + 1)/(14 + 11) \times 100\% = 4\%$. The distribution of the class probabilities is shown in Fig. 5(d).
- Step 4: The peak amplitude of the scattering matrices in the database ranges from 0.22 (for the 0.16λ crack) to 2.17 (for the 2.04λ hole). Because the peak amplitude of the measurement $\mathbf{S}_t$ is 0.44, condition 1 is met. The SSIM index of $\mathbf{S}_t$ and its nearest neighbor in the database is 0.9932. Because this is above the predefined threshold of 0.8, condition 2 is also met, and the characterization result is accepted.

V. Other Validation Cases

Fig. 3(d) shows the local WMRs calculated in the same manner as in Figs. 3(b) and 3(c), but obtained with the proposed dynamic classifier selection approach. It is clear that compared with the results shown in Figs. 3(b) and 3(c), the classification result has been greatly improved by the proposed approach. The overall WMR is 4.40%, and 4441 out of the 4636 test samples have been correctly classified. In this section, defects that were not included in the database are studied to further evaluate the performance of the proposed approach. The studied defects are a rough crack and two volumetric voids which have rough profiles, as well as a porosity defect and one realization of random noise (as a non-defect case). The Kirchhoff approximation is often used to model the scattering matrices of defects having rough profiles and has a low computational cost [38]. However, it was shown that the accuracy of the Kirchhoff model varies with the surface roughness parameters and incident/scattering angles [38], [39]. The scattering matrices studied in this section were simulated using the FELS model [24] at $f = 2$ MHz, and as before, the material is assumed to be aluminum.

A. Rough Cracks

It was shown in [8] that cracks which have severe roughness could not be identified as a crack-like defect. This problem can be considered further using the proposed defect characterization approach because now we have extended the database to include both crack-like defects (class 1) and volumetric voids (classes 2–11). Fig. 6(a) shows the profile of a rough crack (solid line), of which the

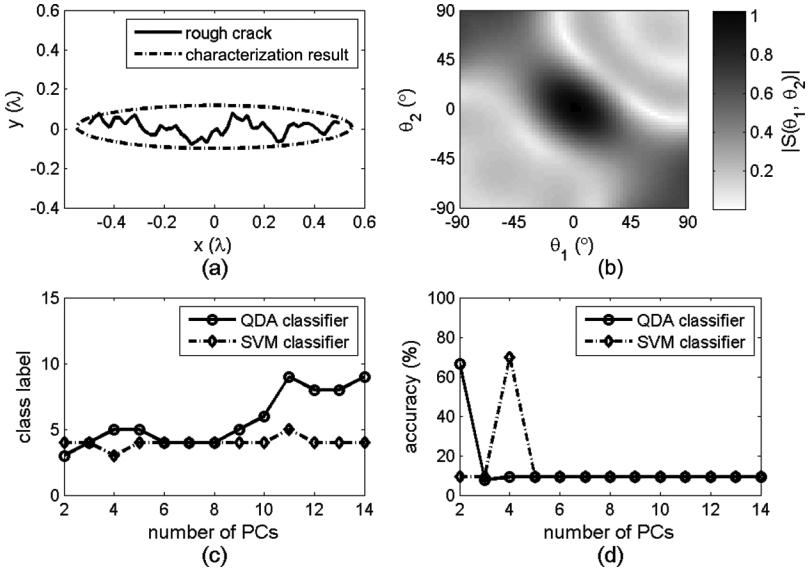

Fig. 6. Characterization of a rough crack: (a) the profile of the rough crack (solid line) and the characterization result obtained with the proposed approach (dashed line), (b) the scattering matrix of the rough crack, (c) the class label output by the individual classifiers, and (d) the estimated local accuracy of the individual classifiers.

scattering matrix is shown in Fig. 6(b). The 26 individual classifiers output conflicting classification results as shown in Fig. 6(c), and the output class labels range from 3 (corresponding to the aspect ratio of 0.2, see Table I for the class definition) to 9 (corresponding to the aspect ratio of 0.8). However, from Fig. 6(d), it can be seen that two classifiers have significantly higher accuracy than the others (the QDA classifier with 2 PCs and the SVM classifier with 4 PCs), and they have the same output of class label 3. The SVM classifier with 4 PCs has the higher estimated local accuracy than the QDA classifier with 2 PCs, and is therefore the optimal classifier for this problem. Based on this result, the rough crack shown in Fig. 6(a) (solid line) is characterized as an ellipse for which the aspect ratio is 0.2. The size of the ellipse is determined by averaging the correctly classified test samples in the local region of the defect as described in Section IV-C, and the characterization result is plotted by dashed line in Fig. 6(a). It can be seen that the rough crack is characterized as an ellipse that covers the same spatial extent as the rough crack. Although it is not fully satisfactory to characterize the rough crack as an ellipse, this result is reasonable considering both the specular normal-incidence reflection amplitude and the tip diffraction amplitude of the rough crack are nearer to the scattering matrix of the ellipse [shown in Fig. 1(b)] than that of a smooth crack [shown in Fig. 1(a)].

B. Volumetric Voids

Figs. 7(a) and 7(b) show the profiles of two arbitrary volumetric defects (solid lines) and their characterization results (dashed lines). For the defect shown in Fig. 7(a), the class label output is 5 by the optimal classifier (i.e., 1 QDA classifier and 12 SVM classifiers are found to be the optimal classifier for the problem, and their classification result and estimated local accuracy is consistent). There are 15 test samples in its local region that are classified by the optimal classifier as class 5, and all of them are correctly classified. Hence, n_d in (14) and (15) is 15, n_i is 15 for $i = 5$ [see (14)], and n_j is 0 for $j \neq 5$ [see (15)]. As a result, the probability of the classification result shown in Fig. 7(a) is 61.54%, and the probability that the defect belongs to any other class is 3.85%. Similarly, for the defect shown in Fig. 7(b), it is classified as class 8 by the optimal classifier (1 QDA classifier and 5 SVM classifiers which have the same class output and the local classifier accuracy). The probability of the characterization result shown in Fig. 7(b) is 71.43% [$n_d = n_i = 24$ ($i = 8$) in (14)], and the probability that the defect belongs to any other class is 2.86% [$n_d = 24$, $n_j = 0$ ($j \neq 8$) in (15)]. Note that the absolute values of the probability of the characterization results seem to be not "confident" enough, but this is purely caused by the add-one smoothing [see (14)] and can be adjusted by adopting different values of the smoothing parameter. Nevertheless, both of the characterization results are reasonably good matches in shape to the original arbitrary shaped defects, and the sizes and depths of the defects are measured accurately.

C. Non-Defect Case (Noise) and Porosity

The SSIM index of the scattering matrices of the rough crack [shown in Fig. 6(a)] and volumetric defects [shown in Figs. 7(a) and 7(b)] and their nearest neighbors are 0.9296, 0.9967 and 0.9276, respectively, hence all of the characterization results shown above can be accepted according to acceptance condition 2. On the other hand, if the SSIM index of the measurement $\mathbf{S}_t$ and its nearest neighbor is below the suggested threshold of 0.8, this is an

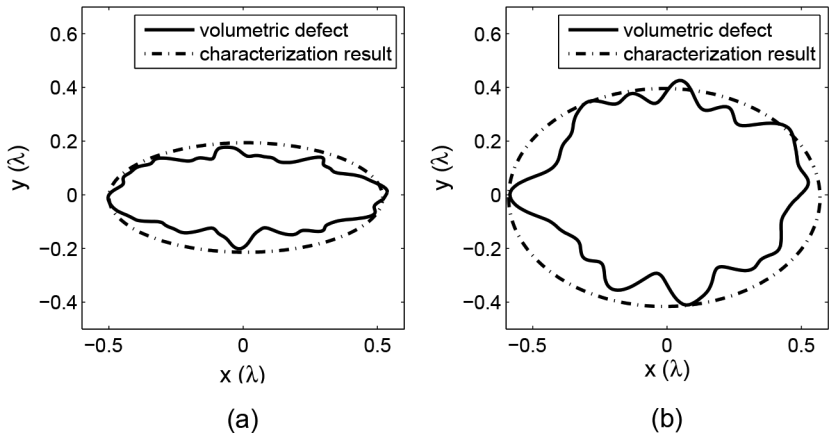

Fig. 7. Characterization of volumetric defects: (a) and (b) the profiles of two arbitrary volumetric defects (solid lines) and their characterization results (dashed lines).

indication of unknown defect type. Examples of this are given in Fig. 8. Fig. 8(a) shows the amplitude of random zero-mean Gaussian noise which has the same standard deviation as the scattering matrix of a 1.58-mm (0.5λ) crack. In Fig. 8(b), 11 holes (the diameter is 0.32 mm, or 0.1λ) are randomly distributed in a 3.15 × 2.21 mm ($1\lambda \times 0.7\lambda$) region. This is to simulate the case of a porosity defect, whose scattering matrix is shown in Fig. 8(c). The SSIM index of the random noise and the porosity defect and their nearest neighbors are 0.3565 and 0.6472, which are significantly lower than the suggested threshold of 0.8. Hence, for both of the two cases considered here, their characterization results should be rejected as not falling within the database.

VI. Experimental Results

The proposed defect characterization approach described in Section IV was also validated through experiments. The test specimen shown in Fig. 9 includes six machined defects that have different shapes. A 2-MHz, 64-element linear array (element pitch: 1.50 mm) was used to perform the measurements, and the scattering matrices of each individual defects were extracted using the inverse imaging algorithm proposed in [40] to eliminate interference from nearby defects. All the defects have the same size of 2.5 mm (0.8λ at $f = 2$ MHz), and detailed size and shape information is given in Table III. Note that this size was chosen because, first, it was small enough to represent a challenging characterization case study and, second, it

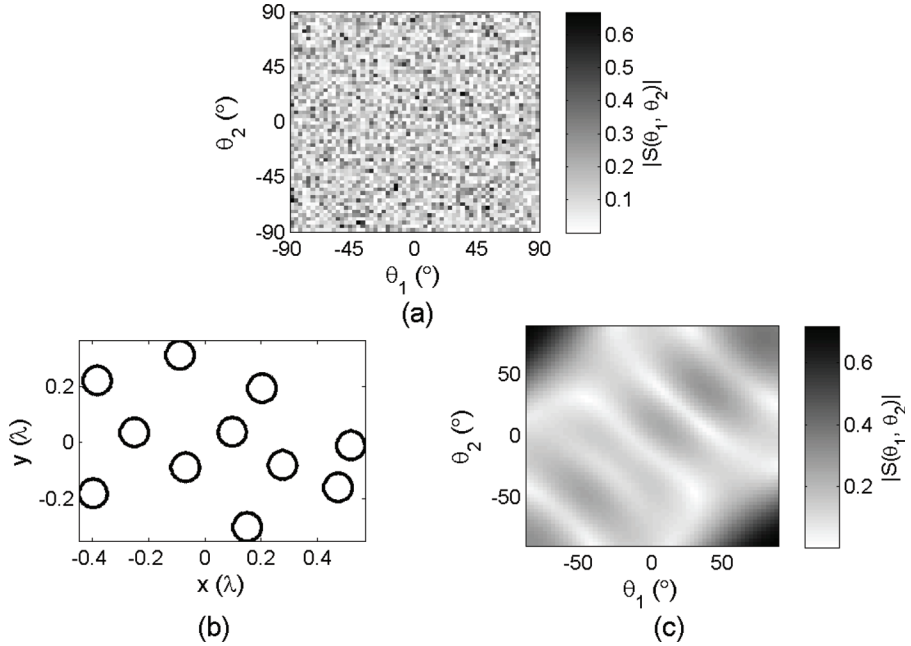

Fig. 8. The scattering matrix of a non-defect case and porosity: (a) the amplitude of random zero-mean Gaussian noise which has the same standard deviation as the scattering matrix of a 0.5λ crack, (b) the random distribution of eleven 0.1λ holes in a $1\lambda \times 0.7\lambda$ region, and (c) the corresponding scattering matrix.

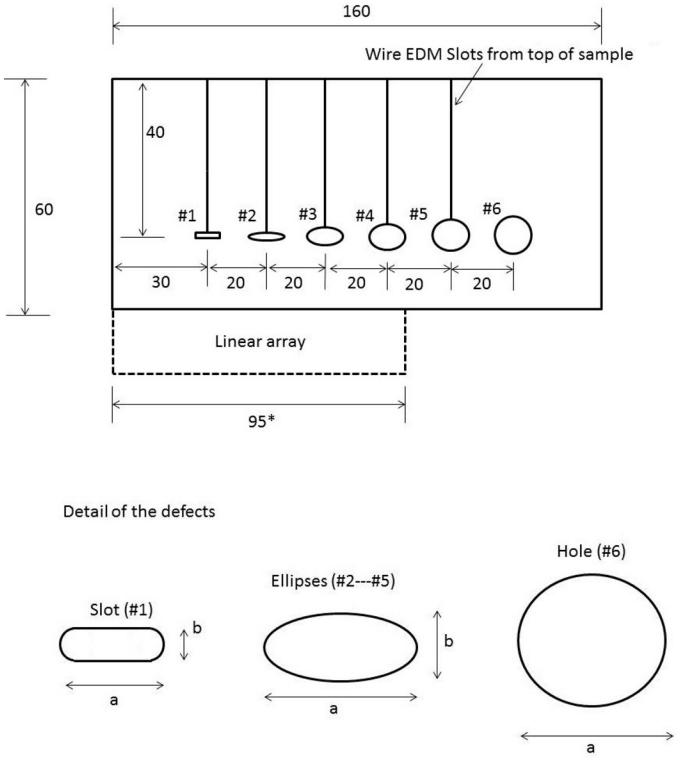

Fig. 9. (Dimensions are in millimeters.) Test specimen geometry and array position for the measurement of defects 1 through 3. The array moves to the right-hand edge of the test specimen for the measurement of defects 4 through 6. The size a of defects 1 through 6 is 2.5 mm, and the minor dimension b of the defects is given in Table III. (*This refers to the distance from the center of the first array element to the center of the last array element.).

was large enough to allow the manufacture of a range of defect shapes.

Figs. 10(a)–10(f) show the experimentally extracted scattering matrices of defects 1–6, respectively. To normalize the absolute amplitude of the scattering matrices extracted experimentally, the scattering matrices shown in Fig. 10 were scaled by a constant gain which makes the mean pulse–echo amplitude of defect 6 (hole) be equal to the simulated value. The incident and scattering angles of defects 1–5 range from $-35°$ to $35°$. For defect 6, because the pulse–echo amplitude appears to be symmetric about $\theta = -10°$, where the maximum amplitude of the scattering matrix is found, the incident and scattering angles extracted cover the range $[-45°, 25°]$. As explained in Section II, for the case of defect 6, it could be regarded having an orientation angle of $\alpha = -10°$ (so the actual angular coverage which is measured relative to the defect normal is still $[-35°, 35°]$), and this orientation angle will not have any significant meaning if the defect could be classified as a side-drilled hole by the proposed approach. Note that although the extracted scattering matrices of each of the defects have slightly different angular ranges given the array-defect configuration shown in Fig. 9, the angular ranges are kept the same here so that the scattering matrices of defects 1–6 can be characterized from the same database.

TABLE III. DETAILS OF THE DEFECTS IN THE TEST SPECIMEN.

Defect	a (mm)	b (mm)	Type
1	2.5	0.4	Slot
2	2.5	0.5	Ellipse
3	2.5	1.0	Ellipse
4	2.5	1.5	Ellipse
5	2.5	2.0	Ellipse
6	2.5	—	Hole

TABLE IV. DEFECT CLASS DEFINITION FOR THE EXPERIMENTAL DATA.

Defect class	Aspect ratio
1	0
2	0.1–0.2
3	0.3–0.4
4	0.5–0.6
5	0.7–0.8
6	0.9–1.0

Because the incident and scattering angles of the experimental scattering matrices are limited to $[-35°, 35°]$, the scattering matrices in the database should have the same angular coverage. Compared with the full angular coverage case, now the database contains less information, which also implies the discriminative ability of the individual classifiers could drop. By performing PCA on the new training set, it was found that only the first 4 PCs are needed to explain 99% of the total variation. After the PC subset selection step, 3 QDA classifiers and 3 SVM classifiers are obtained, which form the input to the dynamic classifier selection system for defect characterization proposed in this paper. Compared with the result shown in Fig. 3(d), both the overall and local accuracy are shown to decrease because of the reduced information content of the scattering matrices. The overall WMR becomes 22.43%, and this is significantly higher than the full angular coverage case (4.40%). This can affect the final result in two ways. First, the individual classifier accuracy decreases so it is more likely that we obtain a misclassified result. Second, low values of the local classifier accuracy prevent the correctly classified results from being selected. Based on these considerations, we reduced the number of the defect classes for the experimental data, so that the classification accuracy can remain high at the cost of the reduced resolution in the defect aspect ratio. The new definition of defect classes is given in Table IV. Now, as before, the same training and test samples are used for training the individual classifiers and the estimation of the local classifier accuracy, except that the incident and scattering angles are limited to $[-35°, 35°]$. Based on this new setup, an overall WMR of 9.69% is obtained by the dynamic classifier selection approach, whereas the globally best individual classifier (SVM classifier with 4 PCs) gives an overall WMR of 15.79%. Because there are fewer individual classifiers considered in the dynamic classifier selection system compared with the full angular coverage case, the performance improvement in terms of the overall WMR is less significant than that shown in Fig. 3. However, as will be discussed later, the globally best

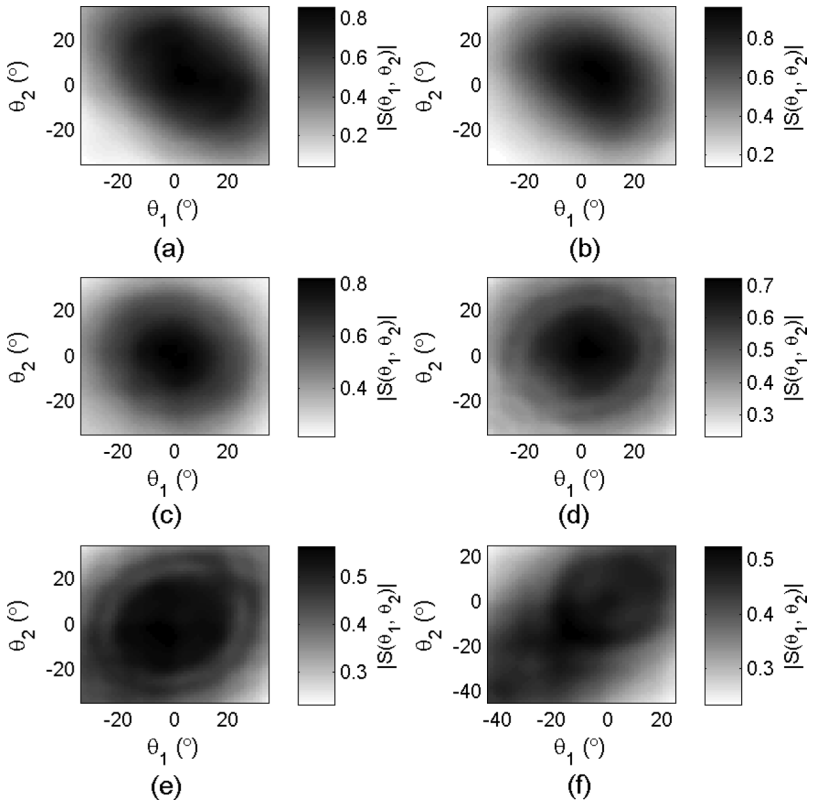

Fig. 10. The experimentally measured scattering matrices of defects 1 through 6 in the test specimen shown in Fig. 9. The amplitude of the scattering matrix of defect 6 (hole) is multiplied by a constant so that the measured mean pulse–echo amplitude becomes equal to the simulated value. The other scattering matrices are scaled by the same constant.

individual classifier could output undesirable classification results while correct results can be obtained by dynamic classifier selection.

Figs. 11(a)–11(f) show the estimated class probabilities of defects 1 through 6, respectively [calculated from (14) and (15)]. It can be seen that defects 2 through 6 are correctly classified, and there are second probable results found for defects 3, 5, and 6, which have lower probabilities compared with the correct results. The scattering matrices of slots are not modeled in the database, and defect 1 is classified as to have an aspect ratio between 0.1 and 0.2 (Class 2). This is an acceptable result because for the scattering matrix shown in Fig. 10(a), its nearest neighbor in the database is found to be the ellipse with the size 0.94λ and aspect ratio 0.2 (the corresponding SSIM index is 0.81). This result can be further verified through simulation, in which the scattering matrix of defect 1 (see Fig. 9 for the detailed shape) is calculated using the FELS model [24]. For the simulated scattering matrix of the slot, if θ_1 and θ_2 are limited to $[-35^\circ, 35^\circ]$ and the class definition shown in Table IV is used, the characterization result of the proposed approach would be 0.83λ (size) and 0.1 to 0.2 (aspect ratio). Alternatively, if θ_1 and θ_2 range from -90° to 90° and the class definition shown in Table I is adopted, we would obtain the result of 0.84λ (size) and 0.2 (aspect ratio). The sizing results as well as the optimal classifiers for each defect are given in Table V. The characterization results are accurate, and the largest error in the measured size is 0.14λ (for defect 2). It is interesting to note that if the globally best classifier (the SVM classifier with 4 PCs) is used to classify the defects, the classification results will be incorrect for defects 4, 5, and 6.

The proposed defect characterization approach requires a predetermined parameter K, which defines the size (number of the nearest neighbors) of the local region that is used to estimate the local classifier accuracy. If K is small, the characterization result is more likely to be affected by noise, especially for the measurements that lie near the class boundaries. On the other hand, if a large value of K is used, the characterization result could potentially be dominated by the test samples which are not similar (near) enough to the measurement. Defect characterization results shown in Section V and Fig. 11 were obtained by the default value of $K = 25$. For the experimental measurements, different values of K were also explored, and the results obtained when K is 50 are included in Table V. It is found that increasing the number of K affects the characterization results to a relatively small extent. The characterization results remain almost unchanged for defects 1 through 4. The sizing of defect 5 changes significantly (from underestimating to overestimating the defect size), but the error (0.09λ) is still small. The worst result is found for defect 6—the classification result is poor, which also causes the size of the defect to be measured inaccurately. For the limited angular coverage case where θ_1 and θ_2 are limited to $[-35^\circ, 35^\circ]$, the optimal

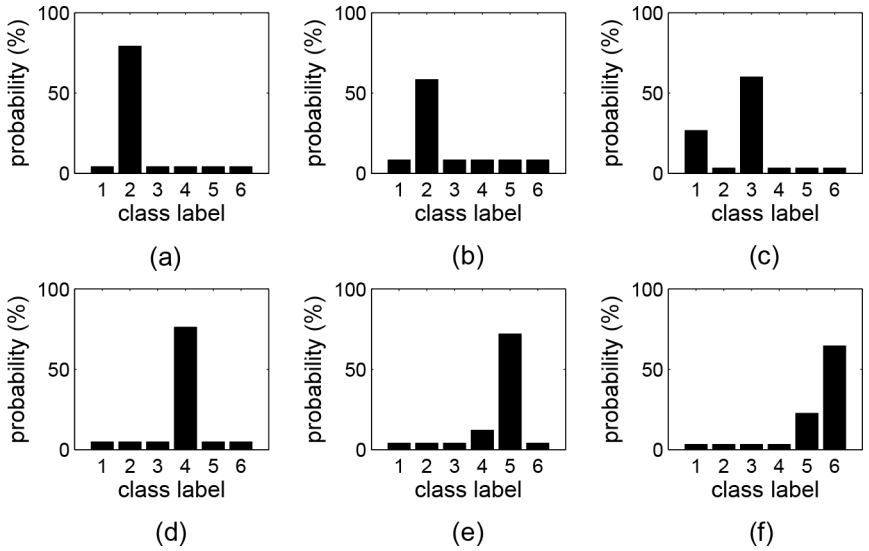

Fig. 11. Probability that experimentally measured defects belong to each of the classes (see Table IV for the class definition), (a)–(f) are the results of defects 1 through 6, respectively.

value of K was found to be 14 by leave-one-out cross-validation of the training set, and the corresponding WMR was 11.48%. When K ranged between 6 and 25, the WMR remained below 13%, which is still near the optimal value. The characterization results obtained when K was set to be the optimal value of 14 were almost identical to those when the default value of K is adopted, with the largest error in the measured size being 0.14λ (for defect 1).

VII. Conclusions

A defect characterization approach based on the scattering matrix is proposed in this paper. The studied problem is a pattern classification problem, where the defect classes are defined according to the aspect ratio of a defect. PCA is used to reduce the dimensionality of the data contained within the scattering matrices. For a given number of PCs, the best PC combination is sought through exhaustive searching, and the resulting PC subsets are used as input features of QDA and SVM classifiers. For a given measurement, the optimal classifier is dynamically selected based on the estimation of the local classifier accuracy, which is calculated from the local region (K nearest neighbors) of the measurement. In simulation, it is shown that 95.79% of the 4636 test samples are correctly classified (aspect ratio determined to within 0.1 relative to the size). Experimentally, the sizes and aspect ratios of six subwavelength defects (including one slot and five volumetric voids) are measured to within 0.14λ and 0.2, respectively.

The proposed defect characterization approach is suitable for real-time computation, because the most time-consuming learning part (the preparation of the individual classifiers using the training samples, i.e., Step 1 in Section IV-E) can be conducted off-line. In addition, because each of the individual classifiers is input to the dynamic classifier selection system in parallel, the approach can be easily extended to include different classification as well as feature extraction algorithms. Porosity/inclusions are not modeled in the database and attempts to characterize such defects can be rejected because the SSIM index of such scattering matrices and their nearest neighbors in the database is below the threshold (0.8). All of the defects considered in this paper are assumed to have a 2-D geometry, but a more comprehensive 3-D analysis is desirable when the defect geometry also shows variation in the third dimension. The aim of the future research is to extend the current database to include even more defect types, and to develop techniques that could characterize defects whose orientation angles are unknown.

TABLE V. Defect Characterization Results (Size a and Aspect Ratio r) and Optimal Classifiers of the Experimental Measurements.

Defect	Characterization result ($K = 25$, default)	Optimal classifier ($K = 25$, default)	Characterization result ($K = 50$)	Optimal classifier ($K = 50$)
1	$a = 0.93\lambda$, $r = 0.1$–0.2	SVM (PCs: 1–4)	$a = 0.94\lambda$, $r = 0.1$–0.2	SVM (PCs: 1–4)
2	$a = 0.94\lambda$, $r = 0.1$–0.2	SVM (PCs: 1–4)	$a = 0.94\lambda$, $r = 0.1$–0.2	SVM (PCs: 1–3)
3	$a = 0.81\lambda$, $r = 0.3$–0.4	QDA (PCs: 2–3)	$a = 0.81\lambda$, $r = 0.3$–0.4	SVM (PCs: 1–4)
4	$a = 0.73\lambda$, $r = 0.5$–0.6	QDA (PCs: 1–3)	$a = 0.71\lambda$, $r = 0.5$–0.6	QDA (PCs: 1– 3)
5	$a = 0.76\lambda$, $r = 0.7$–0.8	QDA (PCs: 1–3)	$a = 0.89\lambda$, $r = 0.7$–0.8	QDA (PCs: 1–3)
6	$a = 0.92\lambda$, $r = 0.9$–1.0	SVM (PCs: 1–2)	$a = 1.01\lambda$, $r = 0.7$–0.8	QDA (PCs: 1–4)

References

[1] J. D. Achenbach, "Quantitative nondestructive evaluation," *Int. J. Solids Struct.*, vol. 37, no. 1–2, pp. 13–27, 2000.

[2] W. D. Callister, *Materials Science and Engineering, an Introduction*. New York, NY, USA: Wiley, 2006.

[3] Y. Humeida, P. D. Wilcox, M. D. Todd, and B. W. Drinkwater, "A probabilistic approach for the optimisation of ultrasonic array inspection techniques," *NDT Int.*, vol. 68, pp. 43–52, Dec. 2014.

[4] B. W. Drinkwater and P. D. Wilcox, "Ultrasonic arrays for non-destructive evaluation: A review," *NDT Int.*, vol. 39, no. 7, pp. 525–541, 2006.

[5] C. Holmes, B. W. Drinkwater, and P. D. Wilcox, "Post-processing of the full matrix of ultrasonic transmit-receive array data for non-destructive evaluation," *NDT Int.*, vol. 38, no. 8, pp. 701–711, 2005.

[6] N. Pörtzgen, D. Gisolf, and G. Blacquiere, "Inverse wave field extrapolation: A different NDI approach to imaging defects," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 54, no. 1, pp. 118–127, 2007.

[7] J. Zhang, B. W. Drinkwater, and P. D. Wilcox, "The use of ultrasonic arrays to characterise crack-like defects," *J. Nondestruct. Eval.*, vol. 29, no. 4, pp. 222–232, 2010.

[8] L. Bai, A. Velichko, and B. W. Drinkwater, "Ultrasonic characterization of crack-like defects using scattering matrix similarity metrics," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 62, no. 3, pp. 545–559, 2015.

[9] K. M. M. Tant, A. J. Mulholland, and A. Gachagan, "A model-based approach to crack sizing with ultrasonic arrays," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 62, no. 5, pp. 915–926, 2015.

[10] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed., New York, NY, USA: Springer-Verlag, 2009.

[11] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 3rd ed., Orlando, FL, USA: Academic Press, 2006.

[12] M. Cacciola, S. Calcagno, F. C. Morabito, and M. Versaci, "Computational intelligence aspects for defect classification in aeronautic composites by using ultrasonic pulses," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 55, no. 4, pp. 870–878, 2008.

[13] Y. Chen, H. W. Ma, and G. M. Zhang, "A support vector machine approach for classification of welding defects from ultrasonic signals," *Nondestruct. Test. Eval.*, vol. 29, no. 3, pp. 243–254, 2014.

[14] P. A. Dickstein, S. Girshovich, Y. Sternberg, A. N. Sinclair, and H. Leibovitch, "Ultrasonic feature-based classification of the interfacial condition in composite adhesive joints," *Res. Nondestruct. Eval.*, vol. 2, no. 4, pp. 207–224, 1990.

[15] F. W. Margrave, K. Rigas, D. A. Bradley, and P. Barrowcliffe, "The use of neural networks in ultrasonic flaw detection," *Measurement*, vol. 25, no. 2, pp. 143–154, 1999.

[16] A. Masnata and M. Sunseri, "Neural network classification of flaws detected by ultrasonic means," *NDT Int.*, vol. 29, no. 2, pp. 87–93, 1996.

[17] P. Rizzo, I. Bartoli, A. Marzani, and F. L. di Scalea, "Defect classification in pipes by neural networks using multiple guided ultrasonic wave features extracted after wavelet processing," *ASME J. Press. Vessel Technol.*, vol. 127, no. 3, pp. 294–303, 2005.

[18] S. Sambath, P. Nagaraj, and N. Selvakumar, "Automatic defect classification in ultrasonic NDT using artificial intelligence," *J. Nondestruct. Eval.*, vol. 30, no. 1, pp. 20–28, 2011.

[19] J. Saniie, E. Oruklu, and S. Yoon, "System-on-chip design for ultrasonic target detection using split-spectrum processing and neural networks," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 59, no. 7, pp. 1354–1368, 2012.

[20] K. Worden, C. R. Farrar, G. Manson, and G. Park, "The fundamental axioms of structural health monitoring," *Proc. R. Soc. A*, vol. 463, no. 2082, pp. 1639–1664, 2007.

[21] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed., New York, NY, USA: Springer-Verlag, 2002.

[22] E. Glushkov, N. Glushkova, A. Ekhlakov, and E. Shapar, "An analytically based computer model for surface measurements in ultrasonic crack detection," *Wave Motion*, vol. 43, no. 6, pp. 458–473, 2006.

[23] R. J. Brind, J. D. Achenbach, and J. E. Gubernatis, "High-frequency scattering of elastic waves from cylindrical cavities," *Wave Motion*, vol. 6, no. 1, pp. 41–60, 1984.

[24] A. Velichko and P. D. Wilcox, "A generalized approach for efficient finite element modelling of elastodynamic scattering in two and three dimensions," *J. Acoust. Soc. Am.*, vol. 128, no. 3, pp. 1004–1014, 2010.

[25] E. B. Flynn, M. D. Todd, P. D. Wilcox, B. W. Drinkwater, and A. J. Croxford, "Maximum-likelihood estimation of damage location in guided-wave structural health monitoring," *Proc. R. Soc. A*, vol. 471, no. 2183, pp. 2575–2596, 2011.

[26] R. E. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton, NJ, USA: Princeton Univ. Press, 1961, p. 94.

[27] E. J. Cross, G. Manson, K. Worden, and S. G. Pierce, "Features for damage detection with insensitivity to environmental and operational variations," *Proc. R. Soc. A*, vol. 468, no. 2148, pp. 4098–4122, 2012.

[28] Y. Wang, Y. Sun, P. Lv, and H. Wang, "Detection of line weld defects based on multiple thresholds and support vector machine," *NDT Int.*, vol. 41, no. 4, pp. 517–524, 2008.

[29] K. Worden and A. J. Lane, "Damage identification using support vector machines," *Smart Mater. Struct.*, vol. 10, no. 3, pp. 540–547, 2001.

[30] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM T. Intel. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

[31] S. Romdhani, P. Torr, B. Schölkopf, and A. Blake, "Efficient face detection by a cascaded support-vector machine expansion," *Proc. R. Soc. A*, vol. 460, no. 2051, pp. 3283–3297, 2004.

[32] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*. Cambridge, UK: Cambridge Univ. Press, 2000, ch. 3.

[33] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, 2014.

[34] X. H. Chen, Y. Yamaguchi, and J. Chen, "Weighted misclassification rate: A new measure of classification error designed for landscape pattern index," *Remote Sens. Lett.*, vol. 3, no. 1, pp. 57–65, 2012.

[35] K. Woods, W. P. Kegelmeyer Jr., and K. Bowyer, "Combination of multiple classifiers using local accuracy estimates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 405–410, 1997.

[36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[37] C. Manning, P. R. Raghavan, and H. Schitze, *Introduction to Information Retrieval*. Cambridge, UK: Cambridge Univ. Press, 2008, ch, 13.

[38] J. Zhang, B. W. Drinkwater, and P. D. Wilcox, "Longitudinal wave scattering from rough crack-like defects," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 58, no. 10, pp. 2171–2180, 2011.

[39] F. Shi, W. Choi, M. J. S. Lowe, E. A. Skelton, and R. V. Craster, "The validity of Kirchhoff theory for scattering of elastic waves from rough surfaces," *Proc. R. Soc. A*, vol. 471, no. 2178, art. no. 20140977, 2015.

[40] A. Velichko and P. D. Wilcox, "Reversible back-propagation imaging algorithm for postprocessing of ultrasonic array data," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 56, no. 11, pp. 2492–2503, 2009.

Long Bai was born in Yanji, P.R. China, in 1987. He received the B.Sc. degree in electronic and information science and technology and the M.Eng. degree in dynamical systems and control from Peking University, P.R. China, in 2010 and 2013, respectively. Since 2013, he has been a Ph.D. student in the Ultrasonics and NDT group at the University of Bristol, and is currently working on defect characterization using ultrasonic arrays for NDE applications.

Alexander Velichko was born in Krasnodar, Russia, in 1975. He received the M.Sc. degree in applied mathematics from the Kuban State University, Krasnodar, Russia, in 1998, and the Ph.D. degree from the Rostov State University, Rostov-on-Don, Russia, in 2002. His doctoral research was on investigation of wave fields caused by internal vibration sources in layered elastic medium.

From 2005 to 2012, he was employed as a research associate in the Ultrasonics and Non-Destructive Testing Research Group at the Univer-

sity of Bristol. In 2012, he was appointed as a lecturer in the Department of Mechanical Engineering at the University of Bristol, England. His current research interests include mathematical modeling of propagation and scattering of elastic waves, ultrasonic imaging using arrays, and guided waves and signal processing.

Bruce W. Drinkwater was born in Hexham, England, in 1970. He received B.Eng. and Ph.D. degrees in mechanical engineering from Imperial College, London, England, in 1991 and 1995 respectively. His Ph.D. thesis was on the subject of solid coupled ultrasonic devices for nondestructive evaluation.

Since 1996, he has worked as an academic in the Mechanical Engineering Department at the University of Bristol, England. He was promoted to Professor in 2007. He has published more than 100 journal articles covering ultrasonics, physical acoustics and NDE. He is currently working on 1) acoustic radiation force devices for human-computer interaction and ultrasonic assembly in biology and materials engineering; and 2) ultrasonic array imaging and defect characterization for NDE.

Between 2000 and 2005, he was an EPSRC Advanced Research Fellow and in 2010, he received the Roy Sharpe Prize for his significant contribution to NDE research.