Unsupervised Diagnostic and Monitoring of Defects Using Waveguide Imaging With Adaptive Sparse Representation

Bin Gao, *Senior Member, IEEE*, Wai Lok Woo, *Senior Member, IEEE*, Gui Yun Tian, *Senior Member, IEEE*, and Hong Zhang, *Student Member, IEEE*

***Abstract*—This paper proposes a new system for the unsupervised diagnostic and monitoring of defects in waveguide imaging. The proposed method is automatic and does not require manual selection of specific frequencies for defect diagnostics. The core of the method is a computational intelligent machine learning algorithm based on sparse non-negative matrix factorization. An internal functionality is built into the machine learning algorithm to adaptively learn and control the sparsity of the factorization, and to render better accuracy in detecting defects. This is achieved by using Bayesian statistics methodology. The proposed method is demonstrated on automatic detection of defect in metals. In addition, we show that the extraction of the spectrum signature corresponding to the defect is significantly more efficient with the proposed optimal sparsity, which subsequently led to better detection performance. Experimental tests and comparisons with other sparse factorization methods have been conducted to verify the efficacy of the proposed method.**

***Index Terms*—Computational intelligence, diagnosis and monitoring, instrumentation, machine learning, signal processing and analysis, waveguide imaging.**

I. Introduction

In RECENT years, many unsupervised machine learning algorithms have been developed for industrial diagnostic imaging system applications [1], [2]. This includes inspection of electronic chips in semiconductor production lines [3], use of unsupervised learning features and multilayer neural networks for defect detection on solder joints [4], defect inspection system of solar modules in electroluminescence (EL) images [5], and machine learning-based fuzzy spectral and spatial feature integration method for classification of nonferrous materials in hyperspectral data [6]. All of these methods recognize that machine learning and pattern-based diagnostic system is a wide group of analysis techniques used in science and industry to evaluate the properties of materials, components, or systems without causing damage [7], [8]. Common machine and pattern feature learning methods consist of principal component analysis (PCA) [9], independent component analysis (ICA) [10], and non-negative matrix factorization (NMF) [11], [12]. In comparison with PCA and ICA, NMF concentrates on the part-based decomposition and it is not necessary to have the constraints of orthogonality and independence. Thus, NMF attracts lots of research applications, such as feature extraction, pattern recognition, machine learning, object detection, and dimensionality reduction [13]–[20]. In this paper, we propose a new NMF-based method for solving the feature extraction problem. In a conventional NMF, given a data matrix $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_L] \in \Re_+^{K \times L}$ with $\mathbf{Y}_{k,l} > 0$, NMF factorizes non-negative input matrix into a product of two non-negative matrices

$$\mathbf{Y} \approx \mathbf{DH}. \tag{1}$$

In (1), $\mathbf{D} \in \Re_+^{K \times \ell}$ and $\mathbf{H} \in \Re_+^{\ell \times L}$ where K and L represent the total number of rows and columns in matrix $\mathbf{Y}$, respectively. It is expected that $\ell < L$, since $\mathbf{D}$ can be compressed and reduced to its integral components, such as $\mathbf{D}_{K \times \ell}$ is a representation as a set of dictionary vectors. $\mathbf{H}_{\ell \times L}$ is an active matrix that controls the amplitude of each dictionary vector at every time or space point. A common method using multiplicative update algorithm as well as families of parameterized cost functions to solve the NMF optimization problem has been introduced in [19] and [20]. The sparse NMF (SNMF) [21] has gained popularity since it enables the decomposition to be made more unique. Specifically, the term "sparseness" denotes a representational scheme where only a few units are effectively employed to represent data vectors. This implies most units taking values close to zero, while only few take significantly nonzero values. The different prior distribution has been incorporated to the update cost function [22]–[24],

Manuscript received November 07, 2014; revised April 05, 2015; accepted May 18, 2015. Date of publication October 26, 2015; date of current version February 02, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 51377015 and Grant 61401071; in part by NSAF under Grant U1430115; in part by FP7 Health Monitoring of Offshore Wind Farms (HERMOW IRSES) project 269202; in part by the China Postdoctoral Science Foundation under Grant 136413; in part by the Science and Technology Department of Sichuan Province, China, under Grant 2013HH0059; and in part by the Fundamental Research Funds for the Central Universities under Grant ZYGX2014J068. Paper no. TII-15-0262.

B. Gao is with the School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: bin_gao@uestc.edu.cn).

W. L. Woo is with the School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K. (e-mail: w.l.woo@ncl.ac.uk).

G. Y. Tian is with the School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China, and also with the School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K. (e-mail: g.y.tian@ncl.ac.uk).

H. Zhang is with the School of Electronic and Information Engineering, Fuqing Branch of Fujian Normal University, Fuzhou 350007, China (e-mail: zhhgw@hotmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2015.2492924

[35]. These include exponential density, inverse gamma density, and Gaussian distributions. The parameters and hyperparameters of these prior models are adapted by ether using the Markov chain Monte Carlo or maximum a posterior probability approaches. The benefit of this is that these approaches increase the accuracy of matrix factorization; however, the computational complexity increases significantly due to the adaptation of the parameters and its hyperparameters in every iteration.

Waveguide-based nondestructive diagnosis (ND) system for defect evaluation benefits both the science and practice of measuring the status of material properties without compromising its function. From an industrial point of view, the demands of ND system include low-cost, noncontact, automatic, accuracy detecting and imaging defects [25]. To achieve this, postsignal processing for defects analysis plays a critical role in waveguide imaging system. Many research studies have already been explored for spectral estimation and image reconstruction of defects by using such methods as Fourier-based, correlation-based, and super-resolution methods [26]. However, all these methods are limited by manual selection of the proper contrast components. Although the results are tolerable, they are generally not predictable and repeatable. This ambiguous case prevents the use of waveguide-based nondestructive testing and evaluation (NDT&E) in automated environments.

In this paper, an optimal sparse representation is proposed to extract the anomalous patterns in a waveguide imaging system. This method can automatically identify the defects in the spatial-frequency domain. The link between the physics model and signal processing has been developed with the aim of realizing an automated process of defect identification. The proposed algorithm will be derived from fundamental principles. Experimental tests on man-made metal defects have shown that our proposed method has resulted in superior detection performance.

This paper is organized as follows. In Section II, the background of waveguide imaging and how the NMF relates to the system is described. Section III presents the automatic sparse NMF model. Experimental results coupled with a series of performance comparison against conventional methods are presented in Section IV. Finally, Section V concludes this paper.

II. Background of Waveguide Imaging

Fine defects and fatigue defects detection on the surface of metal have gained an increasing interest in using microwave and millimeter-wave ND imaging systems since waves at micro or millimeter frequencies can penetrate the low-loss dielectric coating materials. In addition, these techniques with smaller wavelength can render high-spatial resolution images of the interior of various complexes, thick, and layered composite structures [27]–[29].

A. Principle of Waveguide Imaging

Open-ended rectangular waveguide probes are usually used at frequencies which allow only the dominant mode to propagate. The complex reflection coefficient at the aperture of the waveguide is expressed as [28]

$$\Gamma = |\Gamma|\, e^{j\phi} = \frac{1-\Upsilon}{1+\Upsilon} \tag{2}$$

where Γ is related to the terminating admittance of the waveguide Υ. The complex reflection coefficient for both phase and magnitude variations can be measured and calculated using a vector network analyzer. The higher the pronounced shift in the phase and magnitude of the reflection coefficient, the easier the defect can be detected. As this near-field open-ended waveguide for defect detection is based on surface current distribution, when there is a defect, this defect will cause perturbation of the induced surface current density on the metal plate. The dimensions information of defect is rendered by the presence of perturbation.

B. Frequency Spectrum of Defect and Nondefect Areas

Experiments have been conducted by moving the aperture of the open-ended waveguide over the metal surface and the probe is automatically controlled by stepping motor with an X–Y Scanner. In theory, when the defect is parallel to the waveguide (orthogonal to the electric field of the dominant TE_{10} mode), the frequency spectrum experiences a pronounced shift in location. When the defect is exposed to the aperture of the waveguide compared to when the defect is outside the aperture (nondefect area), this shift indicates changes in the reflection coefficient property of the metal surface perturbed by the defect. Thus, both the phase and amplitude of frequency spectrum at the defect area are different from that of the nondefect area. It was also observed that this shift is highly dependent on the relative location of the defect within the waveguide aperture.

There are two main challenges faced by traditional waveguide measurements. First, accurate detection and evaluation of defects is highly dependent on the types of waveguide system (operation frequency) and the interface due to the environment and the surface condition of material. Second, as the resolution of detection is determined by $\lambda_0/\sqrt{\varepsilon}$ where λ_0 is the wavelength in free space, and ε is the permittivity of material, the high-resolution results can be achieved by using high-frequency excitation at millimeter wavelength. However, the overall system becomes complicated, expensive, and may not be robust to noise or random perturbation. Thus, the demands of advanced signal processing algorithm become crucial when using low-frequency waveguide system.

In this case, when the aperture or part of aperture scans the defect edge, the reflected signal consists of both defective and nondefective information. The pattern to be extracted is the defective area when the aperture aligns at the start of the defect edge region extending to the aperture out of the defect region. Since NMF is an "addition model" which describes the positive contributions of each possible factor rather than a "net model" which consists of positive as well as negative contributions and the final result is a net value of these two contributions, the NMF is interpretable and directly models the physical phenomenon. These attributes are not shared by well-known models such as PCA and ICA which belong to the "net model"

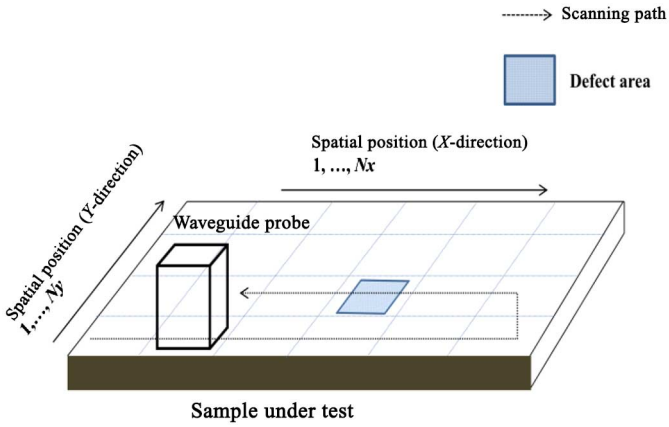


Fig. 1. Waveguide imaging C-scan.

categories. In addition, NMF is computationally more efficient than other sparse representations through the use of multiplicative update algorithm in which its theory shares similarity with the second-order optimization but with a carefully optimized step size leading to a multiplicative update parameter estimation rather than an additive update. This enables the algorithm to have fast convergence and yet computationally simple.

III. Proposed Methodology

A raster-like relative motion of waveguide ND imaging scan procedure is shown in Fig. 1.

The observation of Fig. 1 can be mathematically modeled as a three-dimensional representation (or a tensor) of the mixing spatial-frequency spectrum $\mathbf{Y}'$ which contains both nondefect and defect spatial-frequency spectrum sources. In $\mathbf{Y}'$, the frequencies are given by $f = 1, 2, \ldots, F$ where F represents the total frequency units. The tensor $\mathbf{Y}'$ can be converted into a two-dimensional matrix as $\overline{\mathrm{Y}} = [\mathrm{vec}(\mathbf{Y}'(1))\mathrm{vec}(\mathbf{Y}'(2)) \cdots \mathrm{vec}(\mathbf{Y}'(F))]^T$ where $\mathbf{Y}'(f)$ denotes the spatial-frequency spectrum matrix with dimensions $N_x \times N_y$ of the fth slice of $\mathbf{Y}'$. The mixing spatial-frequency spectrum observation at each frequency-spatial point becomes

$$\overline{\mathbf{Y}}(f,l) = \mathbf{X}^{\mathrm{defect}}(f,l) + \mathbf{X}^{\mathrm{nondefect}}(f,l) \tag{3}$$

Where $\overline{\mathbf{Y}}(f,l)$ is the mixed spatial-frequency spectrum component. The terms $\mathbf{X}^{\mathrm{defect}}(f,l)$ and $\mathrm{X}^{\mathrm{nondefect}}(\mathrm{f,l})$ denote the defect and nondefect spatial-frequency spectrum components, respectively. Here, the space slots are given by $l = 1, 2, \ldots, l_{\max}$ where $l_{\max} = N_x \times N_y$. Note that in (3), each component is a function of f and l. Given above, we define the spatial-frequency "power spectrum" as

$$\left|\overline{\mathrm{Y}}(f,l)\right|^2 \triangleq \overline{\mathrm{Y}}(f,l)\,\overline{\mathrm{Y}}^*(f,l) \tag{4a}$$

where "$*$" denotes complex conjugate. The test object for a given scanned location l is either defective or nondefective since these regions are mutually exclusive. Hence, when $\mathbf{X}^{\mathrm{defect}}(f,l)$ is active, $\mathrm{X}^{\mathrm{nondefect}}(\mathrm{f,l})$ becomes inactive, and vice versa. Additionally, both the phase and amplitude of frequency spectrum corresponding to a defect are different from that of the nondefect. Thus, $\mathbf{X}^{\mathrm{defect}}(f,l) \times \mathbf{X}^{\mathrm{nondefect}}(f,l) \cong 0$ for a given location l and frequency f. Therefore, (4a) reduces to

$$\begin{aligned}\left|\overline{\mathrm{Y}}(f,l)\right|^2 &= \left|\mathbf{X}^{\mathrm{defect}}(f,l)\right|^2 + \left|\mathbf{X}^{\mathrm{nondefect}}(f,l)\right|^2 \\ &\quad + 2\mathrm{Re}\left[\mathbf{X}^{\mathrm{defect}}(f,l) \times \mathbf{X}^{\mathrm{nondefect}}(f,l)\right] \\ &\cong \left|\mathbf{X}^{\mathrm{defect}}(f,l)\right|^2 + \left|\mathbf{X}^{\mathrm{nondefect}}(f,l)\right|^2 \end{aligned} \tag{4b}$$

where $\mathrm{Re}[\cdot]$ means extracting the real component. In matrix representation, (4b) is expressed as

$$\mathbf{Y} \cong \sum_{i=1}^{2} \mathbf{X}_i \tag{4c}$$

where $\mathbf{X}_1 = |\mathbf{X}^{\mathrm{defect}}(f,l)|^2$, $\mathbf{X}_2 = |\mathbf{X}^{\mathrm{nondefect}}(f,l)|^2$, and $\mathrm{Y} = |\overline{\mathrm{Y}}(\mathrm{f,l})|^2$(for all $f = 1, 2, \ldots, F, l = 1, 2, \ldots, l_{\max}$) are matrices (column and row vectors represent the frequency and spatial slots, respectively) denoting the spatial-frequency power spectrum representation of (3). Equation (4) is a generation equation since it describes how $\mathbf{Y}$ is generated as a mixing of $\mathbf{X}_i$. As there are differences in terms of the spatial-frequency power spectrum between defect and nondefect area, the NMF is an efficient tool for extracting these spatial-frequency characteristics. In this paper, we present an optimally sparse NMF where the sparsity parameter is learnt online from the data. The model is given by

$$\mathbf{Y} = \mathbf{DH} + \mathbf{V} = \sum_{d=1}^{d_{\max}} \mathbf{D}_d \mathbf{H}_d + \mathbf{V} \tag{5a}$$

where $\mathbf{Y} \in \Re_+^{F \times l_{\max}}$, $\mathbf{D} \in \Re_+^{F \times d_{\max}}$, $\mathbf{H} \in \Re_+^{d_{\max} \times l_{\max}}$, and $\mathbf{V} \in \Re_+^{F \times l_{\max}}$are noise terms and $\mathbf{H}$ is considered sparsely distributed by

$$p(\mathbf{H}|\lambda) = \prod_{d=1}^{d_{\max}} \prod_{l=1}^{l_{\max}} \lambda_{d,l} \exp(-\lambda_{d,l} \mathbf{H}_{d,l}). \tag{5b}$$

Comparing with a typical prior model, it is worth noting that in (5b), *each individual element* in $\mathbf{H}$ is constrained to an exponential distribution with independent decay sparsity parameter $\lambda_{d,l}$, which will be adaptively learnt to yield the optimal sparse solution. The matrix $\mathbf{D}_d$ is the dth column of $\mathbf{D}$, $\mathbf{H}_d$ is the dth row of $\mathbf{H}$, and $\mathbf{V}$ is assumed to be independently and identically distributed (i.i.d.) as Gaussian distribution with noise having variance σ^2. The terms $d_{\max}$ and $l_{\max}$ are the maximum number of columns in $\mathbf{D}$ and column length in $\mathbf{Y}$, respectively. This is in contrast with the typical SNMF where $\lambda_{d,l}$ is manually set to a fixed constant, i.e., $\lambda_{d,l} = \lambda$ for all d, l. This setting directly imposes uniform constant sparsity on all activation basis matrix $\mathbf{H}$ which enforces each element to be identical to a fixed distribution due to the constant sparsity parameter. This will lead to under- and oversparse decomposition during the optimization procedure. In Section III, we will present the details of the sparsity analysis for defect detection and evaluate its performance against with other existing methods. Given only the observed spatial-frequency spectrum $\mathbf{Y}$, the aim of the proposed algorithm is to determine the optimum parameters

$\mathbf{D}$, $\mathbf{H}$, and $\lambda_{d,l}$ such that the defect spatial-frequency power spectrum components can be estimated accurately. Note that the proposed algorithm does not require prior information or "supervisory knowledge" of the defect and nondefect sources from the system. Thus, the proposed system for the diagnosis and monitoring of defects is unsupervised and fully automated.

A. Development of Proposed L_1-Optimal Sparse NMF

To investigate spectral basis with sparse activation, we choose a prior distribution $p(\mathbf{D},\mathbf{H})$ over the factors $\{\mathbf{D},\mathbf{H}\}$. The posterior can be estimated by employing the Bayes' theorem as

$$p\left(\mathbf{D},\mathbf{H}\left|\mathbf{Y},\sigma^2,\lambda\right.\right)=\frac{p\left(\mathbf{Y}\left|\mathbf{D},\mathbf{H},\sigma^2\right.\right)p\left(\mathbf{D},\mathbf{H}\left|\lambda\right.\right)}{P\left(\mathbf{Y}\right)}. \quad (6)$$

Hence, the negative log likelihood serves as the cost function defined as

$$\begin{aligned}L&\propto\frac{1}{2\sigma^2}\left\|\mathbf{Y}-\sum_d\tilde{\mathbf{D}}_d\mathbf{H}_d\right\|_F^2+f\left(\mathbf{H}\right)\\&=\frac{1}{2\sigma^2}\left\|\mathbf{Y}-\sum_d\tilde{\mathbf{D}}_d\mathbf{H}_d\right\|_F^2+\sum_{d,l}\lambda_{d,l}\mathbf{H}_{d,l}. \quad (7)\end{aligned}$$

The sparsity term $f(\mathbf{H})$ represents the L_1-norm regularization and resolve the ambiguity by forcing all structure in $\mathbf{H}$ onto $\mathbf{D}$. Thus, the sparseness of the solution in (7) is highly dependent on the regularization sparse parameter $\lambda_{d,l}$.

1) Estimation of the Spectral and Activation Basis Matrix: With the normalization in $\mathbf{D}$, the spectral dictionary and activation basis can now be represented as $\tilde{\mathbf{Z}}=\sum_d\tilde{\mathbf{D}}_d\mathbf{H}_d$. By applying the multiplicative learning rules [12], [22], the updates become

$$\mathbf{H}\leftarrow\mathbf{H}\cdot\frac{\tilde{\mathbf{D}}^{\mathbf{T}}\mathbf{Y}}{\tilde{\mathbf{D}}^{\mathbf{T}}\tilde{\mathbf{Z}}+\lambda},\ \text{here } f(\mathbf{H})=\sum_{d,l}\lambda_{d,l}\mathbf{H}_{d,l} \quad (8)$$

$$\mathbf{D}\leftarrow\tilde{\mathbf{D}}\cdot\frac{\mathbf{Y}\mathbf{H}^{\mathbf{T}}+\tilde{\mathbf{D}}\ \text{diag}\left(\sum_\tau\mathbf{1}\left(\left(\tilde{\mathbf{Z}}\mathbf{H}^{\mathbf{T}}\right)\cdot\tilde{\mathbf{D}}\right)\right)}{\tilde{\mathbf{Z}}\mathbf{H}^{\mathbf{T}}+\tilde{\mathbf{D}}\ \text{diag}\left(\sum_\tau\mathbf{1}\left(\left(\mathbf{Y}\mathbf{H}^{\mathbf{T}}\right)\cdot\tilde{\mathbf{D}}\right)\right)}. \quad (9)$$

In (9), $\tilde{\mathbf{D}}\in\Re_+^{F\times d_{\max}}$, $\tilde{\mathbf{Z}}\in\Re_+^{F\times l_{\max}}$, $\mathbf{1}$ is the vector of $[1,\ldots,1]^{\mathbf{T}}$, "$\mathbf{T}$" denotes matrix transpose, "$\cdot$" is the element-wise product, and diag $(\cdot)$ represents a matrix with the argument on the diagonal.

2) Estimation of the Sparsity Parameter: Equation (7) can be rewritten as

$$F(\mathbf{H})=\frac{1}{2\sigma^2}\left\|\text{vec}\left(\mathbf{Y}\right)-\left(\mathbf{I}\otimes\tilde{\mathbf{D}}\right)\text{vec}\left(\mathbf{H}\right)\right\|_F^2+\underline{\lambda}^{\mathbf{T}}\text{vec}\left(\mathbf{H}\right) \quad (10)$$

where vec$(\cdot)$ is the column vectorization, "$\otimes$" denotes the Kronecker product, and $\mathbf{I}$ is the identity matrix. Defining the following terms:

$$\underline{\mathbf{y}}=\text{vec}(\mathbf{Y}),\ \overline{\mathbf{D}}=\left[\mathbf{I}\otimes\tilde{\mathbf{D}}\right],\ \underline{\boldsymbol{\lambda}}=\left[\text{vec}(\lambda)\right],\ \text{and}\ \underline{\boldsymbol{h}}=\left[\text{vec}(\mathbf{H})\right]. \quad (11)$$

$\boldsymbol{\lambda}$ is matrix form of $[\lambda_{d,l}]$with all elements $l=1,2,\ldots,l_{\max}$ and $d=1,2,\ldots,d_{\max}$. Thus, (10) can be rewritten in terms of $\underline{\boldsymbol{h}}$ as

$$F(\underline{\boldsymbol{h}})=\frac{1}{2\sigma^2}\left\|\underline{\mathbf{y}}-\overline{\mathbf{D}}\underline{\boldsymbol{h}}\right\|_F^2+\underline{\lambda}^{\mathbf{T}}\underline{\boldsymbol{h}} \quad (12)$$

where $\underline{\mathbf{y}}\in\Re_+^{S\times1}$, $S=F\times l_{\max}$, the $\underline{\boldsymbol{h}}$ and $\underline{\boldsymbol{\lambda}}$ are vectors of dimension $R\times1$ where $R=d_{\max}\times l_{\max}$ and $\overline{\mathbf{D}}\in\Re_+^{S\times R}$. To determine $\underline{\boldsymbol{\lambda}}$, the expectation-maximization (EM) algorithm [30] is used and $\underline{\boldsymbol{h}}$ is considered as the hidden variable. The log-likelihood function can be optimized with respect to $\underline{\boldsymbol{\lambda}}$ through the Jensen's inequality. For any distribution $Q(\underline{\boldsymbol{h}})$, the log-likelihood function satisfies the following equation:

$$\ln p(\underline{\mathbf{y}}|\underline{\boldsymbol{\lambda}},\overline{\mathbf{D}},\sigma^2)\geq\int Q\left(\underline{\boldsymbol{h}}\right)\ln\left(\frac{p\left(\underline{\mathbf{y}},\underline{\boldsymbol{h}}|\underline{\boldsymbol{\lambda}},\overline{\mathbf{D}},\sigma^2\right)}{Q\left(\underline{\boldsymbol{h}}\right)}\right)d\underline{\boldsymbol{h}}. \quad (13)$$

It is easily evident that the distribution that maximizes $\int Q(\underline{\boldsymbol{h}})\ln(\frac{p(\underline{\mathbf{y}},\underline{\boldsymbol{h}}|\underline{\boldsymbol{\lambda}},\overline{\mathbf{D}},\sigma^2)}{Q(\underline{\boldsymbol{h}})})d\underline{\boldsymbol{h}}$ is given by $Q\left(\underline{\boldsymbol{h}}\right)=p(\underline{\boldsymbol{h}}|\underline{\mathbf{y}},\underline{\boldsymbol{\lambda}},\overline{\mathbf{D}},\sigma^2)$, which is the posterior distribution of $\underline{\boldsymbol{h}}$. In this paper, the posterior distribution is represented as Gibbs distribution, namely

$$Q\left(\underline{\boldsymbol{h}}\right)=\frac{1}{Z_h}\exp\left[-F\left(\underline{\boldsymbol{h}}\right)\right],\quad\text{where } Z_h=\int\exp\left[-F\left(\underline{\boldsymbol{h}}\right)\right]d\underline{\boldsymbol{h}}. \quad (14)$$

The functional form of the Gibbs distribution in (14) is expressed in terms of $F\left(\underline{\boldsymbol{h}}\right)$ and it enables us to simplify the variational optimization of $\underline{\boldsymbol{\lambda}}$. The maximum likelihood estimation of $\underline{\boldsymbol{\lambda}}$ can be expressed as

$$\begin{aligned}\underline{\lambda}^{ML}&=\arg\max_{\underline{\lambda}}\ \ln p\left(\underline{\mathbf{y}}|\underline{\boldsymbol{\lambda}},\overline{\mathbf{D}},\sigma^2\right)\\&=\arg\max_{\underline{\lambda}}\int Q\left(\underline{\boldsymbol{h}}\right)\ln p\left(\underline{\boldsymbol{h}}|\underline{\boldsymbol{\lambda}}\right)d\underline{\boldsymbol{h}}. \quad (15)\end{aligned}$$

By the same token,

$$\begin{aligned}\sigma^2_{ML}&=\arg\max_{\sigma^2}\int Q\left(\underline{\boldsymbol{h}}\right)\left(\ln p\left(\underline{\mathbf{y}}|\underline{\boldsymbol{h}},\sigma^2,\overline{\mathbf{D}}\right)+\ln p\left(\underline{\boldsymbol{h}}|\underline{\boldsymbol{\lambda}}\right)\right)d\underline{\boldsymbol{h}}\\&=\arg\max_{\sigma^2}\int Q\left(\underline{\boldsymbol{h}}\right)\ln p\left(\underline{\mathbf{y}}|\underline{\boldsymbol{h}},\sigma^2,\overline{\mathbf{D}}\right)d\underline{\boldsymbol{h}}. \quad (16)\end{aligned}$$

From (5b), each element in $\mathbf{H}$ is constrained to be exponentially distributed with independent decay parameter. This gives $p\left(\underline{\boldsymbol{h}}|\underline{\boldsymbol{\lambda}}\right)=\prod_p\lambda_p\exp\left(-\lambda_p h_p\right)$ and the update of $\underline{\boldsymbol{\lambda}}$ is expressed as

$$\lambda_p=\frac{1}{\int h_pQ\left(\underline{\boldsymbol{h}}\right)d\underline{\boldsymbol{h}}}\quad\text{for } p=1,2,\ldots,R \quad (17)$$

where λ_p is the pth element of $\underline{\boldsymbol{\lambda}}$. Similarly, the update for σ^2_{ML} is given by

$$\begin{aligned}\sigma^2_{ML}&=\arg\max_{\sigma^2}\int Q\left(\underline{\boldsymbol{h}}\right)\left(-\frac{N_0}{2}\ln\left(2\pi\sigma^2\right)\right.\\&\qquad\left.-\frac{1}{2\sigma^2}\left\|\underline{\mathbf{y}}-\overline{\mathbf{D}}\underline{\boldsymbol{h}}\right\|^2\right)d\underline{\boldsymbol{h}}\\&=\frac{1}{N_0}\int Q\left(\underline{\boldsymbol{h}}\right)\left(\left\|\underline{\mathbf{y}}-\overline{\mathbf{D}}\underline{\boldsymbol{h}}\right\|^2\right)d\underline{\boldsymbol{h}}. \quad (18)\end{aligned}$$

The integral in (17) and (18) is difficult to solve analytically and as such, we seek an approximation to $Q(\underline{\boldsymbol{h}})$. The solution $\underline{\boldsymbol{h}}$ naturally partitions its elements into distinct subsets $\underline{\boldsymbol{h}}_P$ and $\underline{\boldsymbol{h}}_M$ where components $\forall p \in P$ such that $h_p = 0$, and components $\forall m \in M$ such that $h_m > 0$. Given the behavior, $F(\underline{\boldsymbol{h}})$ can be approximated as $F(\underline{\boldsymbol{h}}) \approx F(\underline{\boldsymbol{h}}_M) + F(\underline{\boldsymbol{h}}_P)$. Thus, $Q(\underline{\boldsymbol{h}})$ can be factorized as

$$\begin{aligned} Q(\underline{\boldsymbol{h}}) &= \frac{1}{Z_h} \exp\left[-F(\underline{\boldsymbol{h}})\right] \\ &\approx Q_P(\underline{\boldsymbol{h}}_P) Q_M(\underline{\boldsymbol{h}}_M). \end{aligned} \tag{19}$$

Since $\underline{\boldsymbol{h}}_P = \underline{\mathbf{0}}$ is on the boundary of the distribution, it is natural to invoke the Taylor series expansion about the MAP estimate $\underline{\boldsymbol{h}}^{\text{MAP}}$

$$\begin{aligned} Q_P(\underline{\boldsymbol{h}}_P \geq 0) &\propto \exp\left\{ -\left[\left(\frac{\partial F}{\partial \underline{\boldsymbol{h}}} \right)\Big|_{\underline{\mathbf{h}}^{\text{MAP}}} \right]_P^{\mathbf{T}} \underline{\boldsymbol{h}}_P - \frac{1}{2} \underline{\boldsymbol{h}}_P^{\mathbf{T}} \overline{\boldsymbol{\Lambda}}_P \underline{\boldsymbol{h}}_P \right\} \\ &= \exp\left[-\left(\overline{\boldsymbol{\Lambda}} \underline{h}^{\text{MAP}} - \frac{1}{\sigma^2} \overline{\mathbf{D}}^{\mathbf{T}} \underline{\mathbf{y}} + \underline{\boldsymbol{\lambda}} \right)_P^{\mathbf{T}} \underline{\boldsymbol{h}}_P - \frac{1}{2} \underline{\boldsymbol{h}}_P^{\mathbf{T}} \overline{\boldsymbol{\Lambda}}_P \underline{\boldsymbol{h}}_P \right] \end{aligned} \tag{20}$$

where $\overline{\boldsymbol{\Lambda}}_P = \frac{1}{\sigma^2} \overline{\mathbf{D}}_P^{\mathbf{T}} \overline{\mathbf{D}}_P$, $\overline{\boldsymbol{\Lambda}}_P \in \Re_+^{R \times R}$, and $\overline{\boldsymbol{\Lambda}} = \frac{1}{\sigma^2} \overline{\mathbf{D}}^{\mathbf{T}} \overline{\mathbf{D}}$, $\overline{\boldsymbol{\Lambda}} \in \Re_+^{R \times R}$. The variational approximation to $Q_P(\underline{\boldsymbol{h}}_P)$ can be achieved by using the exponential distribution

$$\hat{Q}_P(\underline{\boldsymbol{h}}_P \geq 0) = \prod_{p \in P} \frac{1}{u_p} \exp(-h_p / u_p). \tag{21}$$

The parameters $\underline{\mathbf{u}} = \{u_p\}$ for $\forall p \in P$ are obtained by minimizing the Kullback–Leibler divergence [31] between Q_P and $\hat{Q}_P$

$$\begin{aligned} \underline{\mathbf{u}} &= \arg\min_{\underline{\mathbf{u}}} \int \hat{Q}_P(\underline{\boldsymbol{h}}_P) \ln \frac{\hat{Q}_P(\underline{\boldsymbol{h}}_P)}{Q_P(\underline{\boldsymbol{h}}_P)} d\underline{\boldsymbol{h}}_P \\ &= \arg\min_{\underline{\mathbf{u}}} \int \hat{Q}_P(\underline{\boldsymbol{h}}_P) \left[\ln \hat{Q}_P(\underline{\boldsymbol{h}}_P) - \ln Q_P(\underline{\boldsymbol{h}}_P) \right] d\underline{\boldsymbol{h}}_P \end{aligned} \tag{22}$$

which gives

$$\min_{u_p} \hat{\underline{\mathbf{b}}}_P^{\mathbf{T}} \underline{\mathbf{u}} + \frac{1}{2} \underline{\mathbf{u}}^{\mathbf{T}} \hat{\boldsymbol{\Lambda}} \underline{\mathbf{u}} - \sum_{p \in P} \ln u_p \tag{23}$$

where $\hat{\underline{\mathbf{b}}}_P = (\overline{\boldsymbol{\Lambda}} \underline{\boldsymbol{h}}^{\text{MAP}} - \frac{1}{\sigma^2} \overline{\mathbf{D}}^{\mathbf{T}} \underline{\mathbf{y}} + \underline{\boldsymbol{\lambda}})_P$ and $\hat{\boldsymbol{\Lambda}} = \overline{\boldsymbol{\Lambda}}_P + \text{diag}\left(\overline{\boldsymbol{\Lambda}}_P\right)$. Solving (23) for u_p leads to the following update [32], [36]:

$$u_p \leftarrow u_p \frac{-\hat{b}_p + \sqrt{\hat{b}_p^2 + 4 \frac{(\hat{\boldsymbol{\Lambda}} \underline{\mathbf{u}})_p}{\tilde{u}_p}}}{2\left(\hat{\boldsymbol{\Lambda}} \underline{\mathbf{u}}\right)_p} \tag{24}$$

where $\underline{\mathbf{u}} \in \Re_+^{R \times 1}$. As for components $\underline{\boldsymbol{h}}_M$, as the non-negative constraints are not active, $Q_M(\underline{\boldsymbol{h}}_M)$ can be approximated as unconstrained Gaussian with mean $\underline{\boldsymbol{h}}_M^{MAP}$. By using the factorized approximation $Q(\underline{\boldsymbol{h}}) = \hat{Q}_P(\underline{\boldsymbol{h}}_P) Q_M(\underline{\boldsymbol{h}}_M)$, this gives

$$\lambda_p = \begin{cases} \frac{1}{h_p^{\text{MAP}}}, & \text{if } p \in M \\ \frac{1}{u_p}, & \text{if } p \in P \end{cases} \tag{25}$$

TABLE I
PROPOSED L_1 -OPTIMAL SPARSE NMF

1. Initialize $\mathbf{D}$ and $\mathbf{H}$ with nonnegative random values.
2. Normalize $\tilde{\mathbf{D}}_d = \mathbf{D}_d / \|\mathbf{D}_d\|$ and Compute $\tilde{\mathbf{Z}} = \sum_d \tilde{\mathbf{D}}_d \mathbf{H}_d$ and update u_p using (24).
3. Assign $\boldsymbol{\lambda}_g = \begin{cases} \frac{1}{\mathbf{H}_g} & \text{if } g \in M \\ \frac{1}{\mathbf{U}_g} & \text{if } g \in P \end{cases}$ where $\mathbf{H}_g \in \Re_+^{d_{\max} \times l_{\max}}$ and $\mathbf{U}_g \in \Re_+^{d_{\max} \times l_{\max}}$ are the matrix representations of u_p and "$\dot{-}$" is element divide.
4. Compute $\sigma^2 = \frac{1}{N_0}\left[\left(\underline{\mathbf{y}} - \overline{\mathbf{D}} \underline{\hat{\mathbf{h}}} \right)^{\mathbf{T}} \left(\underline{\mathbf{y}} - \overline{\mathbf{D}} \underline{\hat{\mathbf{h}}} \right) + \text{Tr}\left(\overline{\mathbf{D}}^{\mathbf{T}} \overline{\mathbf{D}} \mathbf{C} \right) \right]$ where $\mathbf{C} \in \Re_+^{R \times R}$ is computed by using (27)
5. Update $\mathbf{H} \leftarrow \mathbf{H} \bullet \frac{\tilde{\mathbf{D}}^{\mathbf{T}} \mathbf{Y}}{\tilde{\mathbf{D}}^{\mathbf{T}} \tilde{\mathbf{Z}} + \boldsymbol{\lambda}_g}$ $\boldsymbol{\lambda}_g \in \Re_+^{d_{\max} \times l_{\max}}$
6. Compute $\tilde{\mathbf{Z}} = \sum_d \tilde{\mathbf{D}}_d \mathbf{H}_d$ using the updated $\mathbf{H}$.
7. Update $\mathbf{D} \leftarrow \tilde{\mathbf{D}} \bullet \frac{\mathbf{Y}\mathbf{H}^{\mathbf{T}} + \tilde{\mathbf{D}} \, diag\left(\sum_\tau \mathbf{1}\left(\left(\tilde{\mathbf{Z}} \mathbf{H}^{\mathbf{T}} \right) \bullet \tilde{\mathbf{D}} \right) \right)}{\tilde{\mathbf{Z}}\mathbf{H}^{\mathbf{T}} + \tilde{\mathbf{D}} \, diag\left(\sum_\tau \mathbf{1}\left(\left(\mathbf{Y}\mathbf{H}^{\mathbf{T}} \right) \bullet \tilde{\mathbf{D}} \right) \right)}$
8. Repeat steps 2 to 7 until convergence to a desired threshold is reached.

for $p = 1, 2, \ldots, R$ and h_p^{MAP} refers to the pth element of sparse code $\underline{\boldsymbol{h}}_P$ computed from (8) and its covariance $\mathbf{C}$ is given by

$$C_{pm} = \begin{cases} \left(\overline{\boldsymbol{\Lambda}}_P^{-1} \right)_{pm}, & \text{if } p, m \in M \\ u_p^2 \delta_{pm}, & \text{otherwise.} \end{cases} \tag{26}$$

The update rule for σ^2 can be estimated as

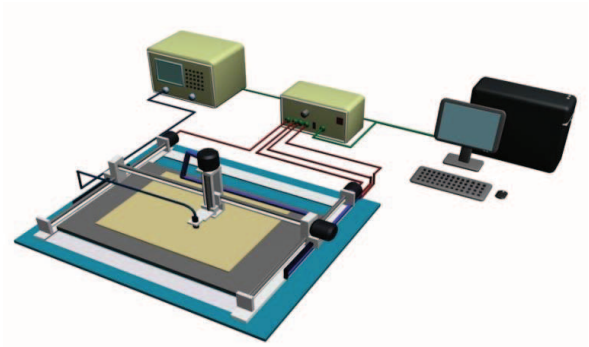
$$\sigma^2 = \frac{1}{N_0} \left[\left(\underline{\mathbf{y}} - \overline{\mathbf{D}} \underline{\widehat{h}} \right)^{\mathbf{T}} \left(\underline{\mathbf{y}} - \overline{\mathbf{D}} \underline{\widehat{h}} \right) + \text{Tr}\left(\overline{\mathbf{D}}^{\mathbf{T}} \overline{\mathbf{D}} \mathbf{C} \right) \right] \tag{27}$$

where $\widehat{h}_p = \begin{cases} h_p^{\text{MAP}}, & \text{if } p \in M \\ u_p, & \text{if } p \in P \end{cases}$. In order to test the efficacy of our proposed method, we evaluate and compare the proposed method with other existing sparse NMF methods. The specific steps of the proposed method have been summarized in Table I.

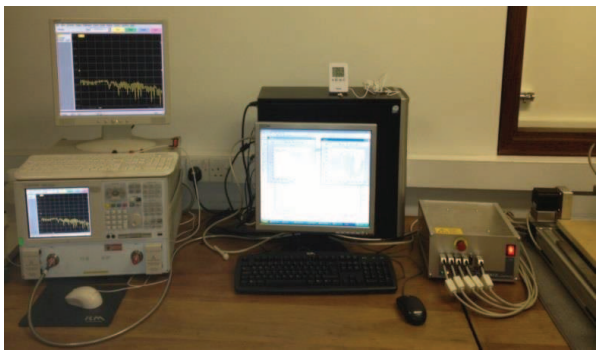
IV. MEASUREMENT SETUP

A. Experimental Platform and Sample Preparation

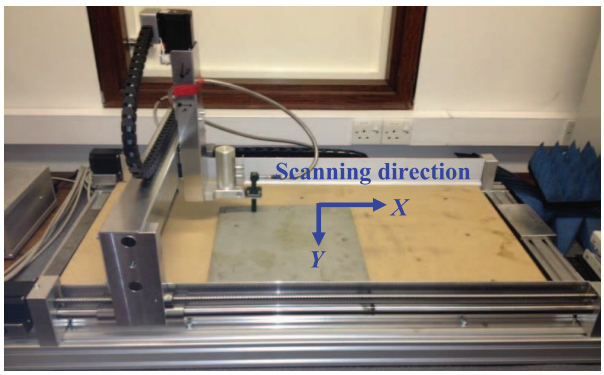
The experimental setup is shown in Fig. 2. An X-band (frequency range 8.2–12.4 GHz) open-ended rectangular waveguide is moved using an X–Y scanner. The type of probe is WR-90 waveguide with the aperture dimensions of 22.86 mm × 10.16 mm ($a \times b$) and the flange dimensions are 42.2 mm × 42.2 mm. A vector network analyzer (Agilent PNA E8363B) is used to provide signal source and obtain the frequency spectrum of the reflected signal. A computer is used to control and acquire the measurement data from the vector network analyzer through IEEE-488 GPIB (general purpose interface bus). In addition, the X–Y scanner is also automatically controlled by PC.



(a)



(b)

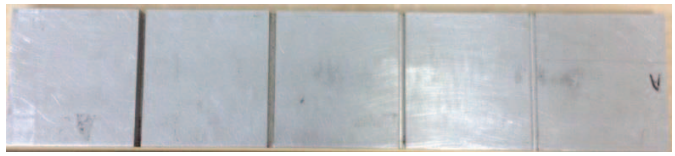


(c)

Fig. 2. (a) Waveguide system. (b) Controller and signal processing platform. (c) Scanning platform.

During the measurement, the linear sweep is applied over the X-band frequency range (frequency resolution is 0.02 GHz with 201 linear sweep points). The reflected frequency spectrum is obtained using linear sweep frequency technology in the vector network analyzer (i.e., measuring the reflection coefficient for each operation frequency over whole sweep frequency range). An aluminum sample with different defects as shown in Figs. 3 and 4 is used for testing. Fig. 4 shows the experimental architecture of waveguide detection system based on waveguide probe (a is broad dimension of the open-ended rectangular waveguide aperture, b is narrow dimension of the open-ended rectangular waveguide aperture, W is the defect width, L is the defect length, δ is location of the defect within waveguide aperture), and all specific experiment setting and sample description is summarized in Table II.

The resolution of the waveguide system can be found from $\rho = \frac{L\lambda}{D_{x,y}}$ where $D_{x,y}$ denotes the length of the aperture in the



(a)

(b)

Fig. 3. Sample under test. (a) Plane view. (b) Side view.

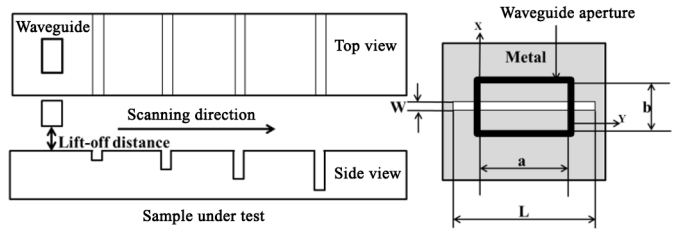


Fig. 4. Schematic of the aluminum sample and probe scanning direction.

TABLE II
TESTING SETUP PARAMETERS FOR MICROWAVE IMAGING

Parameter	Condition
Scanning length for line scan (Lx)	260 mm
Scanning step for line scan (Δx)	1 mm
Defect length $L \times$ Width W	50 mm × 4 mm
Defect depths D	2, 4, 6, 8 mm
Lift-off	1.5 mm

corresponding direction, λ is the wavelength ($= c/f$), and L is the lift-off. With 12.4 GHz and 1.5 mm lift-off, 1.58 mm (the minimum defect should bigger than $\rho/2$) is the minimum defect size, which can be detected with this waveguide in X-direction ($Dx = 10.16$ mm).

B. Probability of Defect Detection

To determine the probability of detection (POD) of the proposed method, the following criteria have been defined.

1) The sample contains a defect and the diagnostic method indicates a defect present [true positive (TP)].
2) The sample contains a defect and the diagnostic method does not indicate a defect present [false negative (FN)].

Since we are interested with the detection of the presence of defect, the POD can be defined as POD $\cong$ TP/(TP + FN). This will be used to validate the proposed method. In order to obtain ground true results, the test sample underwent human annotation with three separate persons per sample to control for any interannotator reliability issues. The basis of annotation is as follows. When the scanning direction is normal to the defect lips, the probe observes a non-normal reflection coefficient for a distance $b + W$ (where b is the narrow dimension of the open-ended rectangular waveguide aperture, W is the

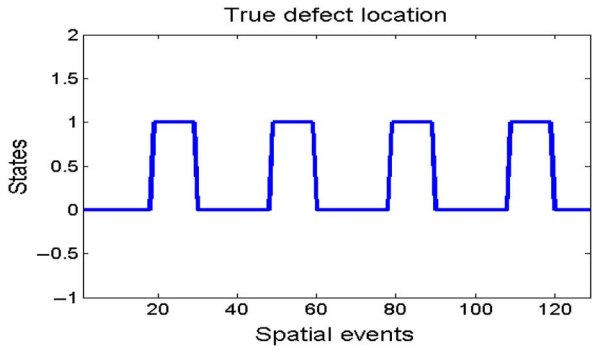


Fig. 5. Ground truth of defect location and size.

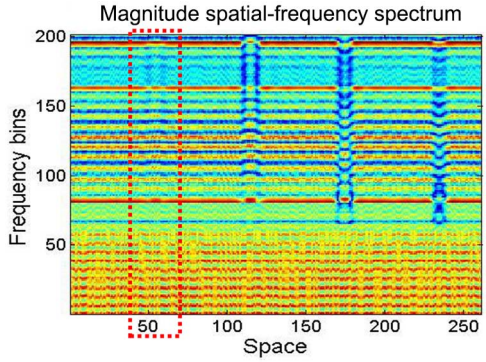


Fig. 6. Mixing spatial-frequency power spectrum of four defects in aluminum plate with different depth (left to right: 2, 4, 6, and 8 mm).

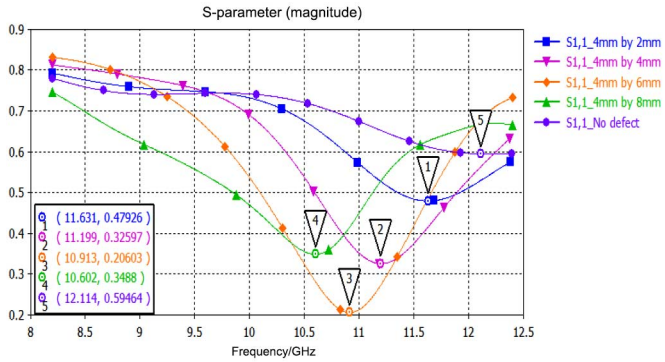


Fig. 7. Simulation for aluminum sample with four different depth cracks and the magnitude results of reflected coefficients and the selected frequency point for defects under different depth situations.

defect width) such that the received signal is not constant. During measurement, the scanning direction of waveguide is along the defect width, and the obtained width of the defect is approximately $b + W = 10.16\,\text{mm} + 4\,\text{mm} = 14.16\,\text{mm}$. This can manually locate and determine the size of the defect from the raw data as shown in Fig. 6.

In order to obtain a robust detection result, event-based technique [34] is used as postprocessing. We consider the signal with window length of four samples as one event and choose the maximum value among these samples which termed. The step size is set to two with 50% overlap between adjacent frames. Specifically, for manually event-based annotation as shown in Fig. 5, the defect at that position is active when the state equals "1." The defect at that position is nonactive when the state equals "0." Once the event-based signal has been obtained, the maximum value of the first 15 samples of data is calculated (known as background nondefect signal), which is used as the threshold to determine whether the further samples belong to a defect or not. This selection is based on Monte Carlo approach where the process is repeated over many realizations within the range between the first 2 and 20 samples and the selection of the first 15 samples to obtain the optimal POD results. The defect activation point is considered when the signal peak value is 40% larger than the variance of nondefect signal values. This selection is also based on Monte Carlo approach within the range between 10% and 70%. Through experiment, we set 40% as the threshold as this gives the best detection results. As for the sparse factorization, the regions between the highest two peaks are considered as the defect activation points.

V. Results and Analysis

A. Aluminum Sample Under Test

Fig. 6 shows the spatial-frequency power spectrum that contains both defect and nondefect areas. The challenging task for waveguide imaging system is that it requires the processing algorithm to not only precisely locate the defect position but also accurately measure the size and depth of the defects. From Fig. 6, however, it is difficult to detect the shallow defect (as marked by the red dotted box) as well as measuring the width and depth of defects. This indicates that the reflection spectrum is unable to provide specific measurements as to: 1) the differences of the frequency spectrum characteristic between nondefect and defect areas; and 2) how to precisely locate and estimate the width of the defect areas (especially for low depth defects). In order to validate the proposed method, we use CST Waveguide Studio 2012 software to simulate the attenuation of the reflection coefficient when the signal meets different types of defects. In addition, we use the standard method [31] to manually select one frequency (whose magnitude attenuation decreases the lowest which marked by triangular box) for comparison. Fig. 7 shows the simulation result of the magnitude of reflected coefficients for both nondefect and defect situations. Fig. 8 shows the experimental detection results by selecting frequency spectrum according the results for different depth of defect.

As shown in Fig. 8, if only a specific frequency is selected for defect detection, it lacks the ability to mine the whole band information and leads to poorer detection results (as none of the selected frequency spectrum can display the detection of all defects). This is clearly shown in the plot where each frequency fails to detect one or two depth defects as marked with the red boxes. By applying the proposed algorithm, it is now possible to resolve these issues.

Fig. 9 shows the factorization results using the previous study of Smooth Itakura-Saito NMF method [33]. The method factorizes the observation matrix into a product of the basis matrix and activation matrix. This is shown in Fig. 8 which shows the factorization results based on only one basis vector for the spatial-frequency power spectrum of the defect area and another one basis vector for the nondefect area (i.e., $d_{\max} = 2$).

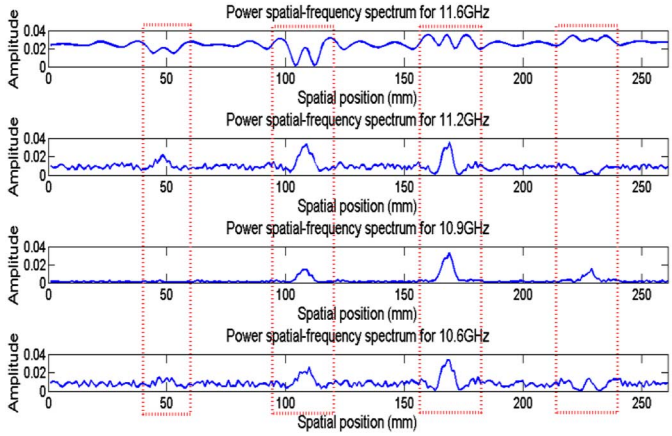


Fig. 8. Experimental detection results by selecting frequency spectrum according the simulation results for different depths of defect.

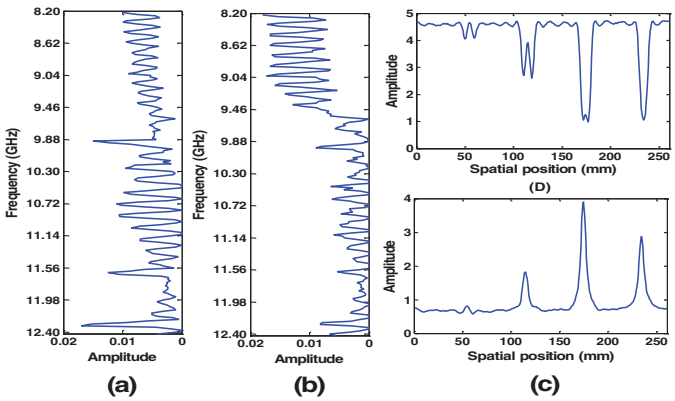


Fig. 9. Results of an aluminum sample with four defects at 1.5 mm lift-off. (a) Spectral basis of the nondefect area. (b) Spectral basis of the defect area. (c) Activation basis of the nondefect area. (d) Activation basis of the defect area.

The factorized basis matrix is used to characterize the spectral basis between the defect and nondefect parts of the spectrum. Fig. 9(a) and (b) shows the factorized spectral basis corresponding to the nondefect and defect spectrum, respectively. Fig. 9(c) and (d) is the factorized activation matrices which can be used further to estimate the defect location, width, and inference of depth information. In particular, Fig. 9(c) shows that when the activation peaks reduce in magnitude, this indicates that the nondefect spectral basis at that spatial position (this position actually refers to defect position) has a lesser degree of being active. Conversely, the activation peaks increase in magnitude at the same position and this is revealed in Fig. 9(d). It is observed that the smooth IS NMF has successfully estimated both the defect location and the defect width from the power spatial-frequency spectrum. In addition, the estimated activation basis has indicated the trend of increasing defects depth, according to attenuation except the last one, which should be with the deepest depth. For the case of 2-mm up to 6-mm defects, the trend of depth information can be predicted. However, for 8-mm depth defect, the peak is shorter than 6-mm depth. The reason the 8-mm depth defect attenuates less than the 6-mm depth defect can be attributed to the 8-mm depth being too deep and is beyond the X-band waveguide system range (8.20–12.40 GHz). Therefore, the waveguide signal is unable to accurately predict the 8-mm defect depth as the attenuation of the reflected signal decreases less than that of the 6-mm depth defect. This can be also validated in the simulation results as shown in Fig. 7. An additional issue of this method is computational complexity, which is mainly attributed to the IS divergence and the extra smooth factor calculation. The least-square cost function is significantly more computationally efficient than the IS divergence. Hence, this is advantageous from the practical point of view. The following section will illustrate and compare the proposed method with other factorization methods.

B. Learnt Sparsity Versus Manually Fixed Sparsity

In this implementation, we conducted several experiments to compare the performance of the proposed method with smooth IS NMF and least-square-based NMF or SNMF under different sparsity regularization. To investigate the impact of sparsity regularization on defect detection for least-square NMF, five cases are conducted.

- Case 1: Nonsparseness (NMF) [12], $\lambda_{d,l} = \lambda = 0$ for all d, l.
- Case 2: Uniform constant sparsity (SNMF) [21] with improper sparseness setting, e.g., $\lambda_{d,l} = \lambda = 0.01$ for all d, l.
- Case 3: Uniform constant sparsity (SNMF) [21] with proper sparseness setting (the uniform regularization is chosen as $c = 0, 0.5, \ldots, 10$ for all sparsity parameters, i.e., $\lambda_{d,l} = \lambda = c$. The best result is retained), $\lambda_{d,l} = \lambda = 7.5$ for all d, l.
- Case 4: Automatic relevance determination ARD-NMF [15]
- Case 5: Proposed learnt L_1-sparsity.

1) Estimated Activation Basis: Since the dominant mode of open-ended waveguide propagating in the z-direction is incident upon the waveguide aperture, defect detection is possible by measuring the total electric field at the location in the interrogating waveguide probe as the probe scans the metal surface. When the defect is absent within the aperture of the probe, the reflected signal remains at a constant level, relating to a short-circuited waveguide. When the probe encounters the opening of a defect, the reflection coefficient is changed, indicating the existence of a crack within the probe aperture. This feature can be used to determine the defect width and depth.

Figs. 10–14 show the matrix factorization results in terms of the estimated activation basis $\mathbf{H}^{\text{defect}}$ which corresponds to the detect location and the prediction of the depth of defects for cases 1–5, respectively. The red dashed lines are the corresponding measurements of the true defect location and width. Figs. 10 and 11 show the cases of "nonsparse" and "improper-sparse" factorization, respectively, which clearly depict the spreading of the estimated activation basis and therefore fail to detect the location and depth of the defects. Figs. 11 and 12 show the cases of uninformed sparse factorization which requires manual setting of λ through conducting many trial-and-error runs. The obtained results, however, are not guaranteed optimal as ambiguities (several spikes as marked in

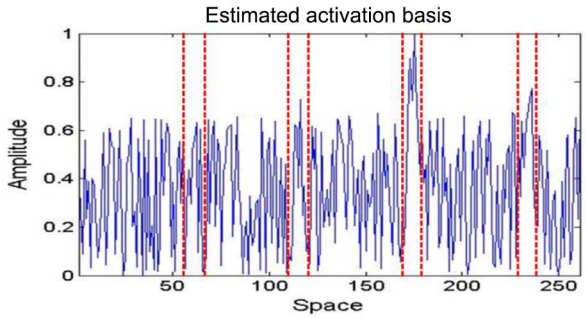


Fig. 10. Estimated $\mathbf{H}^{\text{defect}}$ for case 1.

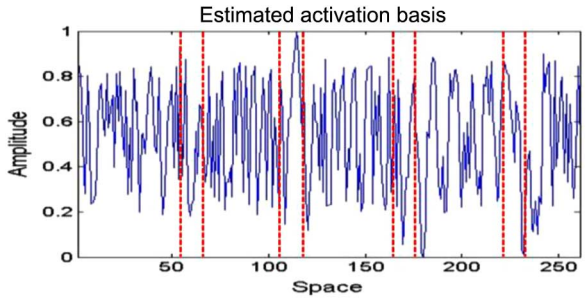


Fig. 11. Estimated $\mathbf{H}^{\text{defect}}$ for case 2.

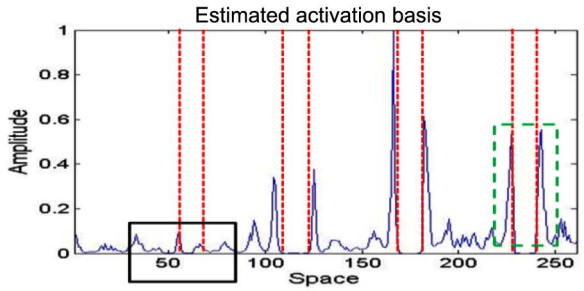


Fig. 12. Estimated $\mathbf{H}^{\text{defect}}$ for case 3.

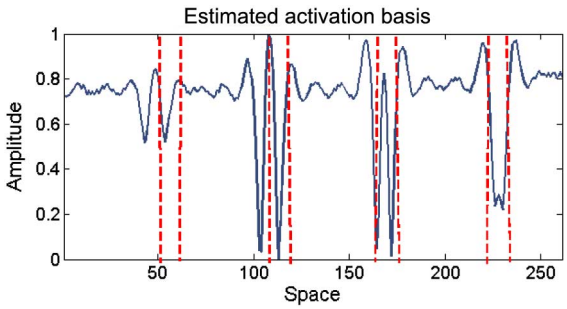


Fig. 13. Estimated $\mathbf{H}^{\text{defect}}$ for case 4.

the black box) still present in detecting the first defect. Also, it fails to predict the depth of the last defect (as marked in green box) which should be the largest spike. There is a significant positive correlation in the reflected amplitude of the spectrum. The reflected scattering signal is more pronounced with deeper defect. Figs. 13 and 14 show the respective "adaptive-sparse" factorization based on the ARD-NMF and the proposed algorithm. It is clearly seen that both methods have successfully detected the defects location. However, the ARD-NMF works

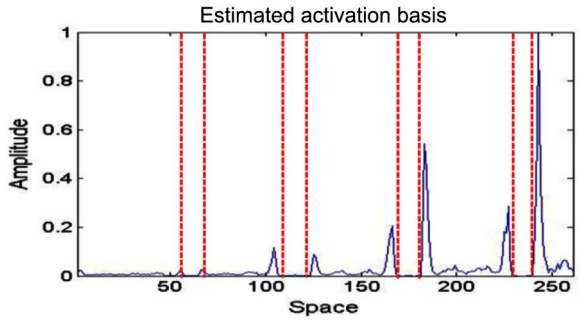


Fig. 14. Estimated $\mathbf{H}^{\text{defect}}$ for case 5.

TABLE III
PROBABILITY OF DETECTION

Methods	Depth of defect 2 mm	4 mm	6 mm	8 mm
NMF estimates (%)	0	0	24.3	0
sm IS NMF (%)	54.3	63.6	85.6	85.6
SNMF case 2 (%)	0	24.3	0	0
SNMF case 3 (%)	22.3	68.6	85.6	85.6
ARD-NMF (%)	54.3	68.6	85.6	85.6
Proposed method (%)	68.4	68.6	85.6	85.6

TABLE IV
DETECTION OF THE LOCATION OF THE DEFECTS

Methods	Depth of defect 2 mm	4 mm	6 mm	8 mm
NMF estimates	Fail	Fail	Fail	Fail
sm IS NMF	Detected	Detected	Detected	Detected
SNMF case 2	Fail	Fail	Fail	Fail
SNMF case 3	Fail	Detected	Detected	Detected
ARD-NMF	Detected	Detected	Detected	Detected
Proposed method	Detected	Detected	Detected	Detected

less than satisfactorily since the trend of peaks fails to predict the depth of the defects. This is further evidenced in Table IV. For the proposed method, both defect location and the trend of the depth have been successfully estimated. The peaks indicate "significant activation" at the boundary of defects, and the "zero activation" inside the four spatial intervals corresponds to the main lobe of the defect signal which is now represented in $\mathbf{V}$ in (5a) as the background of the spatial-frequency power spectrum. These defects can be detected when the probe scans the defect along its width and the waveguide produces strong inflections in the signal at the two locations of the defect entry and exit. This brings the benefit of easy assessment of the width of the defect. Similarly, when the probe scans the defect along its length, the value of length can be readily determined by measuring the region where the detected signal is changing. The detection performance between the true location and depth based on different cases are summarized in Tables III–V.

The POD has been calculated for defects at different depths, respectively, as shown in Table III. The results for NMF and SNMF of case 2 give the worst performance since POD falls below 50%. A higher performance is attained by the smooth IS

TABLE V
PREDICTION OF THE DEPTH OF THE DEFECTS

Methods	Depth of defect 2 mm	4 mm	6 mm	8 mm
NMF estimates	Fail	Fail	Fail	Fail
sm IS NMF	Detected	Detected	Detected	Fail
SNMF case 2	Fail	Fail	Fail	Fail
SNMF case 3	Fail	Detected	Detected	Fail
ARD-NMF	Detected	Detected	Fail	Fail
Proposed method	Detected	Detected	Detected	Detected

NMF with an average POD around 70%. The SNMF of case 3 gives mediocre performance with an average POD around 60%. Both ARD-NMF and the proposed method have significantly improved the POD rate for different defect depths, especially for a defect which has a depth of 2 mm. In addition, the average improvement is more than 60% compared with the lowest performance.

This is the results as shown in Fig. 8 which uses the non-sparse NMF that exactly capture the feature of the main lobe (the aperture directly reflecting the signal from the defect areas). When the defect is near to the waveguide aperture, the aperture causes both the electric (E) and magnetic (H) fields to bend around the defect. The motivation of this study is to show that besides using the main lobe of the reflected signal, the sidelobe signal (part of aperture that corresponds to the defect edge) also provides information which can be used to detect as well as predicting the depth of the deep defects if the degree of sparseness is correctly imposed on the activation basis.

2) Tuning Behavior of Learnt Sparsity Parameter: In this section, the tuning behavior of the learnt sparsity parameters by using the proposed method will be demonstrated. Several sparsity parameters have been selected to illustrate its adaptive behavior. During the iteration, all sparsity parameters are initialized as $\lambda_{d,l} = 1$ for all d, l and are subsequently adapted. After 100 iterations, the above sparsity parameters are significantly different for each sparsity parameter, e.g., $\lambda_{1,1} = 96$, $\lambda_{1,5} = 120$, $\lambda_{1,10} = 99$, and $\lambda_{1,15} = 118$ even though they started at the identical initial condition. This shows that each activation basis has its own sparsity. In addition, it is worth noting that in the case of defect detection, all location and trend of depth of defects have been successfully predicted when $\lambda_{d,l}$ is learnt adaptively. This represents a large improvement over the case of uniform constant sparsity (where no defects are detected). On the other hand, when sparsity is not imposed onto the activation basis, the defect detection fails as well. This extends through the manual selection of the sparse parameter where the first defect location and the depth of last defect fail to be predicted. Thus, the results indicate that the performance of defect detection has been undermined when the uniform constant sparsity scheme is used. On the contrary, improved performance can be obtained by enabling the sparsity parameters to be individually adapted for each element activation basis. We have plotted the histogram of the all adaptive sparsity parameters after 100 iterations in Fig. 15. The plot shows that the histogram is not a unimodal distribution.

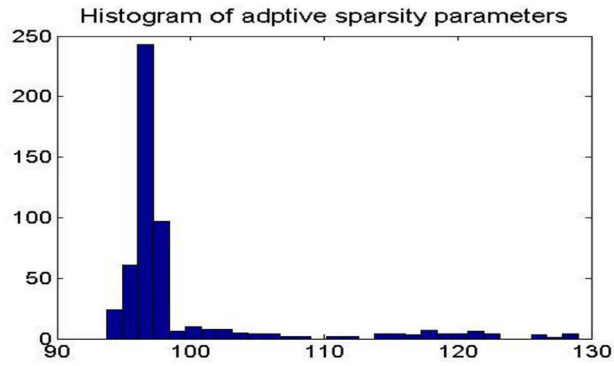


Fig. 15. Histogram of the learnt sparsity parameter.

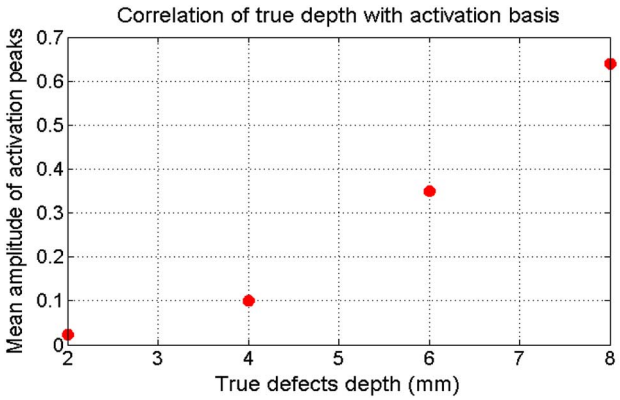


Fig. 16. Correlation between true defect depths with mean value of activation peaks.

We have used the Gaussian mixture model (GMM) to learn the distribution of this histogram and the result produces three Gaussian distributions with mean 98, 103, and 120. The global mean of the GMM tends to 98.

Based on the above analysis, it is clear that the various sparseness values are necessary for the attainment of optimally sparse factorization. However, the uniform constant sparsity matrix factorization raises a consequential issue since it is not possible to determine *a priori* which of the particular activation basis should be assigned the degree of sparseness. This poses a difficult problem in conventional SNMF which requires manual setting of the sparsity parameters. This therefore calls for the need to impose adaptive sparseness on each element of the activation basis so that they may be individually and adaptively optimized. In addition, we have taken the mean value of activation peaks to measure the correlation between the true depths of each defect with the estimated activation basis.

In Fig. 16, we have also calculated the correlation p-value where $p = 0.026 < 0.05$ (if the p-value is less than 0.05, then the two parameters are significantly correlated). This indicates that the estimated activation basis using the proposed learnt sparse NMF can be used to predict the defects depth in waveguide imaging system.

3) Comparison of Convergence Study and Computational Time: The convergence study and computational time of each method have been compared. These

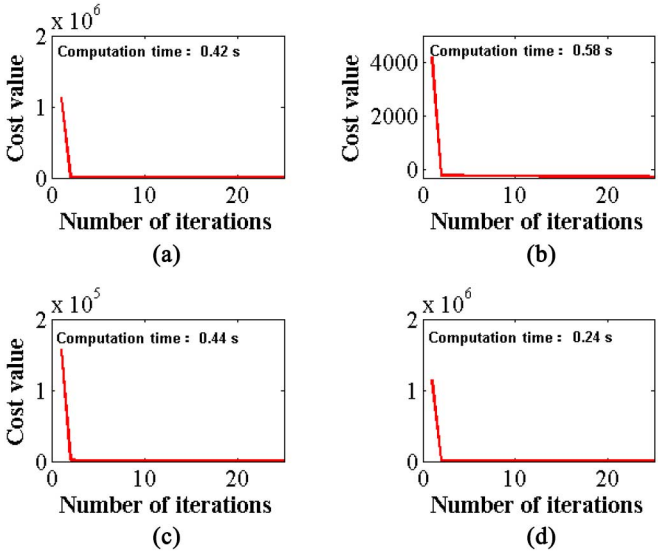


Fig. 17. Convergence trajectory and computation time of the different methods. (a) Proposed method. (b) ARD-NMF. (c) Smooth NMF. (d) General NMF/SNMF.

include the proposed method, NMF-ARD, smooth NMF, and general NMF/SNMF (cases 1–3) and are shown in Fig. 17.

Fig. 17 shows the convergence trajectory of each algorithm as well as the computation time for 25 iterations. The threshold for determine convergence is when the rate of change of the cost value is less than 10^{-3}. In terms of computation time, the general NMF is the fastest since it does not require any computation of hyperparameters. For the rest, hyperparameters learning is required and yet the proposed method renders high computation speed, while the ARD-NMF is the slowest.

VI. Conclusion

This paper presents a machine learning algorithm based on structured NMF for automated diagnosis and monitoring of defects. The proposed sparse representation is adaptively learnt from the underlying data statistics without using prior "supervisory knowledge" of the defect and nondefect spectra. The regularization term is learnt online using the Bayesian approach to yield the desired L_1-optimal sparse decomposition, thus enabling the activation basis to be estimated more effectively in the spatial-frequency power spectrum domain. This has been verified concretely based on our real test results. In addition, the proposed method has yielded significant improvements in defect detection when compared with conventional matrix factorization methods.

Appendix

A. Formulation of the Update Rule for u_p

The optimization of (23) can be accomplished be expanding (23) as follows:

$$G\left(\underline{\mathbf{u}},\tilde{\underline{\mathbf{u}}}\right) = \hat{\underline{\mathbf{b}}}_P^{\mathbf{T}}\underline{\mathbf{u}} + \frac{1}{2}\sum_{p\in P}\frac{\left(\hat{\boldsymbol{\Lambda}}\tilde{\underline{\mathbf{u}}}\right)_p}{\tilde{u}_p}u_p^2 - \sum_{p\in P}\ln u_p \qquad \text{(A1)}$$

Taking the derivative of $G\left(\underline{\mathbf{u}},\tilde{\underline{\mathbf{u}}}\right)$ in (A1) with respect to $\underline{\mathbf{u}}$ and setting it to be zero, we have

$$\frac{\left(\hat{\boldsymbol{\Lambda}}\tilde{\underline{\mathbf{u}}}\right)_p}{\tilde{u}_p}u_p + \hat{b}_p - \frac{1}{u_p} = 0. \qquad \text{(A2)}$$

The above equation is equivalent to the following quadratic equations:

$$\frac{\left(\hat{\boldsymbol{\Lambda}}\tilde{\underline{\mathbf{u}}}\right)_p}{\tilde{u}_p}u_p^2 + \hat{b}_p u_p - 1 = 0. \qquad \text{(A3)}$$

References

[1] H. Gao, C. Ding, C. Song, and J. Mei, "Automated inspection of E-shaped magnetic core elements using K-tSL-center clustering and active shape models," *IEEE Trans. Ind. Informat.*, vol. 9, no. 3, pp. 1782–1789, Aug. 2013.

[2] W. C. Li and D. M. Tsai, "Defect inspection in low-contrast LCD images using Hough transform-based nonstationary line detection," *IEEE Trans. Ind. Informat.*, vol. 7, no. 1, pp. 136–147, Feb. 2011.

[3] X. L. Bai, Y. M. Fang, W. S. Lin, L. P. Wang, and B. F. Ju, "Saliency-based defect detection in industrial images by using phase spectrum," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2135–2145, Nov. 2014.

[4] G. Acciani, G. Brunetti, and G. Fornarelli, "Application of neural networks in optical inspection and classification of solder joints in surface mount technology," *IEEE Trans. Ind. Informat.*, vol. 2, no. 3, pp. 200–209, Aug. 2006.

[5] D.-M. Tsai, I.-Y. Chiang, and Y.-H. Tsai, "A shift-tolerant dissimilarity measure for surface defect detection," *IEEE Trans. Ind. Informat.*, vol. 8, no. 1, pp. 128–137, Feb. 2012.

[6] A. Picon, O. Ghita, P. F. Whelan, and P. M. Iriondo, "Fuzzy spectral and spatial feature integration for classification of nonferrous materials in hyperspectral data," *IEEE Trans. Ind. Informat.*, vol. 5, no. 4, pp. 483–494, Nov. 2009.

[7] Y. Si, J. Mei, and H. Gao, "Novel approaches to improve robustness, accuracy and rapidity of iris recognition systems," *IEEE Trans. Ind. Informat.*, vol. 8, no. 1, pp. 110–117, Feb. 2012.

[8] D. M. Tsai and J. Y. Luo, "Mean shift-based defect detection in multicrystalline solar wafer surfaces," *IEEE Trans. Ind. Informat.*, vol. 7, no. 1, pp. 125–135, Feb. 2011.

[9] B. Gao, L. Bai, G. Y. Tian, W. L. Woo, and Y. Cheng, "Automatic defect identification of Eddy current pulsed thermography using single channel blind source separation," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 913–922, Apr. 2014.

[10] G. J. Brown and M. Cooke, "Computational auditory scene analysis," *Comput. Speech Lang.*, vol. 8, no. 4, pp. 297–336, 1994.

[11] X. Luo, M. Zhou, Y. Xia, and Q. Zhu, "An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1273–1284, May 2014.

[12] D. Lee and H. Seung, "Learning the parts of objects by nonnegative matrix factorisation," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[13] P. Parathai, W. L. Woo, and S. S. Dlay, "Single-channel blind separation using L_1-sparse complex nonnegative matrix factorization for acoustic signals," *J. Acoust. Soc. Amer.*, vol. 137, p. EL124, 2015.

[14] B. Gao, A. Yin, G. Tian, and W. L. Woo, "Thermography spatial-transient-stage mathematical tensor construction and material property variation track," *Int. J. Therm. Sci.*, vol. 85, pp. 112–122, 2014.

[15] V. Y. F. Tan and C. Févotte, "Automatic relevance determination in nonnegative matrix factorization with the beta-divergence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1592–1605, Jul. 2013.

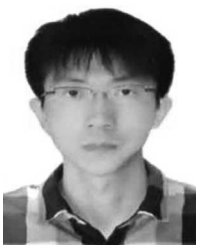
[16] D. Guillamet and J. Vitrià, "Introducing a weighted nonnegative matrix factorization for image classification," *Pattern Recognit. Lett.*, vol. 24, pp. 2447–2454, 2014.

[17] F. J. Theis and G. A. García, "On the use of sparse signal decomposition in the analysis of multi-channel surface electromyograms," *Signal Process.*, vol. 86, no. 3, pp. 603–623, Mar. 2006.

[18] O. Okun and H. Priisalu, "Unsupervised data reduction," *Signal Process.*, vol. 87, no. 9, pp. 2260–2267, Sep. 2007.

[19] R. Kompass, "A generalized divergence measure for nonnegative matrix factorization," *Neural Comput.*, vol. 19, no. 3, pp. 780–791, 2007.

[20] A. Cichocki, R. Zdunek, and S. I. Amari, "Csiszár's divergences for non-negative matrix factorization: Family of new algorithms," in *Proc. Int. Conf. Ind. Compon. Anal. Blind Signal Sep. (ICABSS'06)*, Charleston, SC, USA, Mar. 2006, vol. 3889, pp. 32–39.
[21] B. Gao, W. L. Woo, and S. S. Dlay, "Variational regularized two-dimensional nonnegative matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 5, pp. 703–716, May 2012.
[22] B. Gao, W. L. Woo, and S. S. Dlay, "Sparsity nonnegative matrix factorization for single channel source separation," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 989–1001, Sep. 2011.
[23] A. T. Cemgil, "Bayesian inference for nonnegative matrix factorisation models," *Comput. Intell. Neurosci.*, 2009, 17 pp., doi: 10.1155/2009/785152.
[24] B. Gao, W. L. Woo, and S. S. Dlay, "Unsupervised single channel separation of non-stationary signals using Gammatone filterbank and Itakura-Saito nonnegative matrix two-dimensional factorizations," *IEEE Trans. Circuits Syst. I Reg. Pap.*, vol. 60, no. 3, pp. 662–675, Mar. 2013.
[25] B. Gao, L. Bai, W. L. Woo, and G. Tian, "Thermography pattern analysis and separation," *Appl. Phys. Lett.*, vol. 104, 5pp., 2014, doi: 10.1063/1.4884644.
[26] G. C. Giakos, L. Fraiwan, N. Patnekar, S. Sumrain, G. B. Mertzios, and S. Periyathamby, "A sensitive optical polarimetric imaging technique for surface defects detection of aircraft turbine engines," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 1, pp. 216–222, Feb. 2014.
[27] S. Barbarossa, L. Marsili, and G. Mungari, "SAR super-resolution imaging by signal subspace projection techniques," *AEU-Archiv fur Elektronik und Ubertragungstechnik*, vol. 50, no. 2, pp. 133–138, 1996.
[28] S. Kharkovsky and R. Zoughi, "Microwave and millimeter wave nondestructive testing and evaluation—Overview and recent advances," *IEEE Instrum. Meas. Mag.*, vol. 10, no. 2, pp. 26–38, Apr. 2007.
[29] C. Yeh and R. Zoughi, "Microwave detection of finite surface cracks in metals using rectangular waveguides," *Res. Nondestruct. Eval.*, vol. 6, no. 1, pp. 35–55, 1994.
[30] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.
[31] S. Kullback, "Letter to the editor: The Kullback–Leibler distance," *Amer. Statist.*, vol. 41, no. 4, pp. 340–341, 1987.
[32] C. Wei, W. L. Woo, and S. S. Dlay, "Nonlinear underdetermined blind signal separation using Bayesian neural network approach," *Digit. Signal Process.*, vol. 17, no. 1, pp. 50–68, 2007.
[33] B. Gao, H. Zhang, W. L. Woo, G. Y. Tian, L. Bai, and A. Yin, "Smooth nonnegative matrix factorization for defect detection using microwave non-destructive testing and evaluation," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, 923–934, Apr. 2014.
[34] J. Ortiz Laguna, A. Olaya, and D. Borrajo, "A dynamic sliding window approach for activity recognition," in *User Modeling, Adaption and Personalization, Lecture Notes in Computer Science*, J. Konstan, R. Conejo, J. Marzo, and N. Oliver, Eds. New York, NY, USA: Springer, 2011, pp. 219–230.
[35] B. Gao, W. L. Woo, and B. W.-K. Ling, "Machine learning source separation using maximum a posteriori nonnegative matrix factorization," *IEEE Trans. Cybern.*, vol. 44, no. 7, pp. 1169–1179, Jul. 2014.
[36] B. Gao, W. L. Woo, and S. S. Dlay, "Single channel source separation using EMD-subband variable regularized sparse features," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, no. 4, pp. 961–976, May 2011.



Bin Gao (M'12–SM'14) received the B.S. degree in communications and signal processing from Southwest Jiao Tong University, Chengdu, China, in 2005, and the M.Sc. degree (with distinction) in communications and signal processing and the Ph.D. degree in research of blind source separation and statistic signal processing from Newcastle University, Newcastle upon Tyne, U.K., in 2007 and 2011, respectively.

From 2011 to 2013, he worked as a Research Associate with Newcastle University. Currently, he is an Associate Professor with the School of Automation Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China. He is also a very active Reviewer for many international journals and long standing conferences. He has coordinated several research projects for the National Natural Science Foundation of China. His research interests include sensor signal processing, machine learning, social signal processing, and nondestructive testing and evaluation.

Wai Lok Woo (M'09–SM'11) received the B.Eng. degree (first class Hons.) in electrical and electronics engineering and the Ph.D. degree in blind source separation and statistic signal processing from Newcastle University, Newcastle upon Tyne, U.K., in 1993 and 1998, respectively.

He is currently a Senior Lecturer and Director of Operations with the School of Electrical and Electronic Engineering, Newcastle University. He has authored over 250 papers on these topics on various journals and international conference proceedings. His research interest include mathematical theory and algorithms for nonlinear signal and image processing, machine learning for signal processing, data mining, blind source separation, multidimensional signal processing, and high performance computational intelligence.

Dr. Woo is a Member of the Institution Engineering Technology. Currently, he is an Associate Editor of several international journals and has served as Lead-Editor of journal special issues. He was the recipient of the IEE Prize and the British Scholarship to continue his research work.

Gui Yun Tian (M'01–SM'03) received the B.Sc. degree in metrology and instrumentation, and the M.Sc. degree in precision engineering from the University of Sichuan, Chengdu, China, in 1985 and 1988, respectively, and the Ph.D. degree in nondestructive testing and evaluation and comupter vision from the University of Derby, Derby, U.K., in 1998.

From 2000 to 2006, he was a Lecturer, Senior Lecturer, Reader, Professor, and Head of the Group of Systems Engineering, respectively, with the University of Huddersfield, Huddersfield, U.K. Since 2007, he has been with Newcastle University, Newcastle upon Tyne, U.K., where he has been a Chair Professor in Sensor Technologies. Currently, He is also with the School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu. He has coordinated several research projects from the Engineering and Physical Sciences Research Council, the Royal Academy of Engineering, and FP7. In addition, he also has good collaboration with leading industrial companies such as Airbus, Rolls Royce, BP, nPower, and TWI among others.

Hong Zhang (S'11) received the Bachelor's degree electrical engineering and automation from a joint program between Nanjing Normal University, Nanjing, China, and Northumbria University, Newcastle upon Tyne, U.K., in 2009, and the Master's and Ph.D. degrees in electrical and electronic engineering from Newcastle University, Newcastle upon Tyne, in 2010 and 2014, respectively.

He is currently a Lecturer with the School of Electronic and Information Engineering, Fuqing Branch of Fujian Normal University, Fuzhou, China. He is also a Reviewer for many international journals. His research interests include electromagnetic nondestructive testing, RFID, microwave, and communication science and technology.