
Gao B, Lu P, Woo WL, Tian GY, Zhu Y, Johnston M. [Variational Bayesian Sub-group Adaptive Sparse Component Extraction for Diagnostic Imaging System](#). *IEEE Transactions on Industrial Electronics* (2018)

DOI link

<https://doi.org/10.1109/TIE.2018.2801809>

ePrints link

<http://eprint.ncl.ac.uk/246364>

Date deposited

27/02/2018

Copyright

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Variational Bayesian Sub-group Adaptive Sparse Component Extraction for Diagnostic Imaging System

Bin Gao, *Senior Member IEEE*, Peng Lu, W.L. Woo, *Senior Member, IEEE*, G.Y. Tian, *Senior Member, IEEE*, Yuyu Zhu, Martin Johnston

Abstract —A novel unsupervised sparse component extraction algorithm is proposed for detecting micro defects when employing a thermography imaging system. The proposed approach is developed using the Variational Bayesian framework. This enables a fully automated determination of the model parameters and bypasses the need for human intervention in manually selecting the appropriate image contrast frames. An internal sub-sparse grouping mechanism and adaptive fine-tuning strategy have been built to control the sparsity of the solution. The proposed algorithm is computationally affordable and yields a high accuracy objective performance. Experimental tests on both artificial and natural defects have been conducted to verify the efficacy of the proposed method.

Index Terms — Low-rank decomposition, Variational Bayesian (VB), diagnostic imaging system, sparse decomposition.

NOMENCLATURE

Notational symbols are listed below:

VB	Variational Bayesian
NDT	Non-Destructive Testing
ECPT	Eddy Current Pulsed Thermography
IT	Induction Thermography
EC	Eddy current
PCA	Principal Component Analysis
ICA	Independent Component Analysis
MCMC	Markov Chain Monte Carlo
IR	Infrared camera
EM	Electromagnetic
$\mathbf{Y}', \mathbf{X}', \mathbf{M}$	Input data, patterns, mixing parameter
$\mathbf{L}, \mathbf{S}, \mathbf{N}$	Low-rank matrix, sparse matrix, noise
q	Hyper-parameter
\mathbf{U}, \mathbf{V}	Factor matrixes
α, β	Precisions of the distribution
$p(\bullet)$	Conditional distribution
$p(\mathbf{Y}', \lambda, \mathbf{Z})$	Joint distribution
$q(\bullet)$	Posterior distribution
PPT	Pulsed Phase Thermography
TSR	Thermographic Signal Reconstruction
BRTF	Bayesian Robust Tensor Factorization

I. INTRODUCTION

Imaging diagnostic systems have been widely used in industry especially in the manufacturing sector. Xiao *et al.* [1] proposed a methodology of a hybrid x-ray imaging system and 2-D wavelet transform to automatically detect the internal defects of castings. An optical interference pattern sensing method [2] has been proposed for thin-film transistor defect inspection. A multilayer neural network based imaging inspection has been built for defect detection in for [3] solder joints. Picon *et al.* combined fuzzy spectral and spatial image

features for classification of non-ferrous materials [4]. Tsai *et al.* proposed an automated visual inspection scheme for multicrystalline solar wafers by using the mean-shift technique [5]. Juang *et al.* proposed a Takagi-Sugeno fuzzy real-time object detection system by using a support vector machine in principal component space [6]. Chen *et al.* proposed an effective traffic surveillance system for detecting and tracking moving vehicles [7]. Infrared thermography is a non-destructive and non-contact method [8] that has many advantages in defect detection such as rapid inspection, high spatial resolution, high sensitivity and high efficiency. Therefore, it has become an attractive technique for Non-Destructive Testing (NDT) applications. Eddy Current Pulsed Thermography (ECPT) is a multi-physics integrative NDT method that belongs to the Induction Thermography (IT) family [9]. ECPT is not limited to the sample surface but it can also reach a certain depth governed by the skin depth of eddy currents (ECs). Furthermore, it focuses the heat on the defect due to friction or eddy current distortion and subsequently increases the temperature contrast between the defective region and defect-free areas. Therefore, it has been used to detect defects in conductive material [10-12].

Various signal processing methods have been applied to detect defects in ECPT. However, most detection methods are limited to the manual selection of the proper contrast of the thermal images [11]. To enhance the signal-to-noise ratio (SNR), frequency analysis has been conducted by applying a Fourier Transform to find the flaw-contrast by using a phase map [9]. Pattern-based approach using Principal Component Analysis (PCA) and Independent Component Analysis (ICA) have also been used for thermography feature extraction [13]. However, all previous methods failed to build the connection between the physical mechanism and the intrinsic properties of the detection process. Although the results are acceptable to a certain extent, there is a crucial need to improve the overall detection precision in order to obtain a reliable diagnostic imaging system.

A sparsity model has recently been developed in various fields. Kim [14] proposed an efficient and scalable sparse coding algorithm based on a first-order smooth optimization. Luo *et al.* [15] used a sparse Bayesian extreme learning machine for the task of multi-classification. Gui *et al.* [16] provided a survey on various structure sparsity-inducing methods for feature selection. In addition, a sparse representation has been proposed to enhance image super-resolution [17]. Recently, sparsity has been exploited in unsupervised pattern recognition methods [18]. Robust PCA [19, 20] is proposed to separate sparse and low-rank decomposition for linearly correlated images. Shen *et al.* [21] proposed the sparse representation for intrinsic image decomposition. Jiang *et al.* [22] proposed sparse and dense

hybrid representation via dictionary decomposition for face recognition. Shekhar *et al.* [23] proposed joint sparse representation for robust multimodal biometrics recognition. Hu *et al.* [24] proposed multi-feature joint sparse representation for object tracking. Chen *et al.* [25] proposed fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. VB sparse PCA and Markov Chain Monte Carlo (MCMC) sparse PCA [26, 27] with specific priors have been proposed for adaptive decomposition. In a general setting, sparse decomposition will invariably suffer from either under- or over-sparseness, which subsequently lead to ambiguity in extracting the target components. This suggests that the present form of sparseness control strategy is still technically inadequate and therefore it is necessary to link the physical mechanism to the appropriate model generation.

In this paper, a novel adaptive sub-group sparsity component decomposition algorithm is proposed. The primary goal is to extract anomalous patterns for micro-defect detection in the ECPT imaging system. The contributions of the paper are as follows: (i) The proposed model allows automated sparseness control as well as sub-grouping of components so that the decomposition can be iteratively optimized. This overcomes the problem of under- and over-sparse factorization; (ii) In the proposed model, both control parameters and decomposition are simultaneously learned and adapted as part of the factorization by using a VB approach; (iii) The development of the mathematical link that bridges the physics mechanisms and the sparse decomposition for crack detection has been established. Conventional thermal-based defect feature extraction methods are used for comparison based on the event F-score. The proposed model has the potential to improve the defect detection precision significantly and this will be demonstrated on artificial and natural steel cracks.

The rest of the paper is organized as follows: Section II discusses the proposed methodology. Section III describes the experimental setup. The obtained results and discussion have been presented in Section IV. Finally, Section V concludes the paper.

II. PROPOSED METHODOLOGY

A. Inductive thermal conduction

An infrared camera (IR) system is used to record the time sequence images, which contain both spatial and transient information. This can be represented as a three dimensional array $\bar{Y}(N_x \times N_y \times N_d)$. The terms N_x and N_y denote the spatial dimension while N_d denotes the transient dimension. When an electromagnetic (EM) field is applied to the conductive material, Joule heating is induced. Therefore, the temperature of the material increases. Combining heat diffusion and Joule heating [8], the heat conduction equation of a specimen can be expressed as:

$$\frac{\partial T}{\partial t} = \frac{k_u}{\rho_1 C_p} \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) + \frac{1}{\rho_1 C_p} Q(x, y, z, t) \quad (1)$$

where $T = T(x, y, z, t)$ denotes the distribution of temperature, k_u denotes the thermal conductivity of the material (W/m K), ρ_1 denotes the density (kg/m³), C_p denotes the specific heat (J/kg K) and $Q(x, y, z, t)$ denotes the internal heat generation function per unit volume, which is the result of the eddy current excitation.

B. Thermal patterns and sparse behaviors of defects

To interpret the thermal patterns, a penetrated slot specimen is used as a test sample to represent one type of fundamental defect. Specifically, when the eddy current encounters a discontinuity, e.g., a slot or notch, it is forced to divert, leading to areas of increased and decreased eddy current density. Therefore, in the heating phase, different areas will have different heat generation rates, which subsequently lead to spatial variation of temperature. Hot spots are observed around the tips of the slot while the cool areas are located at both sides of the slot. In the cooling phase, heat diffuses from a high temperature area to a low temperature area and reduces the contrast in the acquired image. Fig. 1(a) shows the numerical simulations for the fusion of eddy current and temperature distribution, which is performed by using the COMSOL multi-physical simulation software via the electro-thermal module. The heat transfer process and the magnetic field propagation were solved simultaneously by this module for an accurate description of the heating mechanism around a particular defect. Fig. 1(b) illustrates the simulation result for the slot after 200ms of heating. The flow of EC and thermal distribution for the defect are visualized by the streamline plot in Fig. 1(b). In the presence of a defect, the ECs will divert to complete their closed loop path, which leaves a unique EC distribution based on the defect geometry that can provide useful multi-physical information to identify a defect. The real test sample and its thermal spatial image after heating for 0.1s and four positions of thermal transient curves are shown in Fig. 1(c) as well.

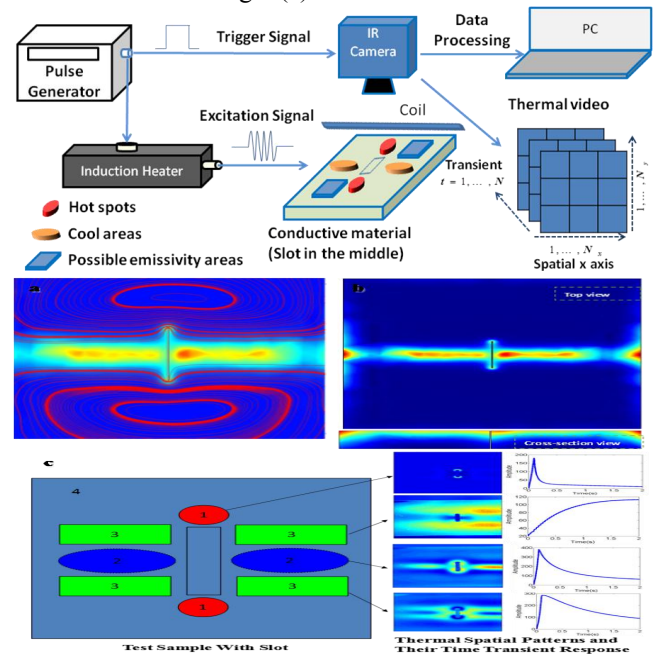


Fig. 1: Inductive Thermography and interpretation of thermal patterns illustration of Thermal-Electromagnetic patterns. a) Schematic of theoretical EC distribution and resultant heating for slot. b) Simulation of heat. c) Spatial-Transient pattern interpretation and separation results.

C. Sparse pattern modeling and adaptive extraction

1) Observation model

All of the above characteristic thermal-transient regions can be considered as “pattern regions” since each of them share similar transient responses. To avoid the influences of arbitrary selection of image frame from the transient thermal video, the extraction of abnormal patterns has become more crucial. Therefore, the task of the proposed method is to separate the thermography sequence into thermal pattern images as well as to automatically identify the specific pattern that relates to the defect. The mathematical model used in [28] can be described as:

$$\mathbf{Y}(t) \approx \sum_{i=1}^{N_s} m_i \mathbf{X}_i(t) \quad (2)$$

where $\mathbf{X}_i(t) \in \mathbb{R}^{N_x \times N_y}$ denotes the number of patterns. The term m_i denotes the mixing parameter that describes the contribution of the i^{th} thermal pattern (e.g. hot spots, cool area, etc) to the recorded thermography image, which is interpreted in Fig. 1. This work uses a fixed-length segment derived from a transient response and we have $\mathbf{Y}' = [\text{vec}(\mathbf{Y}(t)), \text{vec}(\mathbf{Y}(t+1)), \dots, \text{vec}(\mathbf{Y}(t+N_d-1))]$, where ‘vec’ is the vectorization operator. The constructed image sequences have a linear relationship with the different pattern regions and can be expressed as:

$$\mathbf{Y}' = \mathbf{X}'\mathbf{M} \quad (3)$$

where $\mathbf{Y}' \in \mathbb{R}^{K \times N_d}$, $K = N_x \times N_y$, $\mathbf{M} = [\mathbf{m}_1, \dots, \mathbf{m}_{N_s}]^T$, $\mathbf{M} \in \mathbb{R}^{N_s \times N_d}$ is the parameter matrix, \mathbf{m}_i is the i^{th} mixing parameter vector and $\mathbf{X}' \in \mathbb{R}^{K \times N_s}$ is denoted as $\mathbf{X}' = [\text{vec}(\mathbf{X}_1(t)), \text{vec}(\mathbf{X}_2(t)), \dots, \text{vec}(\mathbf{X}_{N_s}(t))]$.

2) Sparse pattern modeling and extraction

The model in (2) is general and does not emphasize or exploit the sparse pattern. However, the sparse pattern focuses on the significant regions (e.g. hot spots) that will benefit the quantitative detection of small defects (e.g. cracks). Thus, the sheer low-rank assumption for the data may not be fully suitable due to its over-simplistic structure. Therefore, (2) is modified and broken down into a linear combination of the sparse patterns (e.g. defect regions), the low-rank patterns and the noise. Therefore, the model of robust PCA [19] can be applied and expressed as:

$$\mathbf{Y}' = \underbrace{[\mathbf{X}'\mathbf{M}]_{i=1, \dots, N_s, i \neq j}}_{\mathbf{L}} + \underbrace{\mathbf{X}'_j \mathbf{M}_j}_{\mathbf{S}} + \mathbf{N} \quad (4)$$

where \mathbf{Y}' can be decomposed as the linear combination of three parts. The first part is the low-rank matrix \mathbf{L} (e.g. for cool area, emissivity area and background reflected patterns in Fig. 1), the second part is for the sparse patterns \mathbf{S} (e.g. defect

regions) that contain only few non-zero values and the final part is the noise \mathbf{N} , where $\mathbf{L} \in \mathbb{R}^{K \times N_d}$ is the low-rank matrix with rank $r \ll \min(K, N_d)$, $\mathbf{S} \in \mathbb{R}^{K \times N_d}$ is the sparse matrix and $\mathbf{N} \in \mathbb{R}^{K \times N_d}$ is the noise matrix. The proposed model focuses on the extraction of sparse patterns and preservation of the defect information. However, the general model of factorization will still invariably suffer from either under- or over-sparseness, which subsequently lead to ambiguity in separating defect sparse patterns from the background. In order to deal with the issue, a new sparse control model is introduced to replace (4) as:

$$\mathbf{Y}' = \mathbf{L} + \lambda \mathbf{S} + \mathbf{N} \quad (5)$$

where λ is the parameter that controls the sparseness of \mathbf{S} . The algorithm uses an adaptive iterative algorithm to estimate the optimal \mathbf{S} and λ . \mathbf{L} is the low-rank matrix, which is updated by using the factorization $\mathbf{L} = \mathbf{U}\mathbf{V}^T$, where \mathbf{U} is a $K \times r$ matrix and \mathbf{V} is a $N_d \times r$ matrix. Both \mathbf{U} and \mathbf{V} can be obtained by using the singular value decomposition. In this work, \mathbf{S} is the sparse matrix where each element can be assumed to obey the independent Gaussian distribution that is:

$$p(\mathbf{S} | \lambda, \boldsymbol{\alpha}) = \prod_i \prod_j \mathcal{N}(s_{ij} | 0, (\lambda^q \alpha_{ij})^{-1}) \quad (6)$$

where q is the hyper-parameter, which can be heuristically set or determined by using the Monte-Carlo approach. Therefore, it does not appear in other distribution formula. The term α_{ij} represents the precision, which can be assumed to obey the Jeffrey’s priors [26], namely,

$$p(\alpha_{ij}) = (\alpha_{ij})^{-1} \quad i = 1, \dots, K; j = 1, \dots, N_d \quad (7)$$

In reality, $\lambda^q \alpha_{ij}$ tends to become a large finite value while the corresponding s_{ij} will converge to zero.

If the precision connects with only one parameter such as $p(\mathbf{S} | \boldsymbol{\alpha}) = \prod_i \prod_j \mathcal{N}(s_{ij} | 0, \alpha_{ij}^{-1})$, then the sparsity of s_{ij} may not be optimally controlled by the simple adjustment of α_{ij}^{-1} . Therefore, this work augments the sparseness control parameter λ^q for α_{ij} to enhance its ability to render better sparsity control. In Fig. 2, the variation of the sparse distribution of s_{ij} actually depends on all parameters of λ , α_{ij} and q . Fig. 2(a) illustrates the case of a Gaussian distribution $p(s_{ij} | \lambda, \alpha_{ij})$ that connects λ , $\alpha_{ij} = 1$ and $q = 1$. When λ increases, the distribution of s_{ij} tends to be more compact. When $\lambda < 1$, the distribution spreads flatter. Fig. 2(b) illustrates the case of $p(s_{ij} | \lambda^2, \alpha_{ij})$ that connects λ , $\alpha_{ij} = 1$ and $q = 2$. In Comparison, the distribution of s_{ij} is even more super-Gaussian. Therefore, this enables flexibility and adaptability of the data under different situations. Hence, the sparsity degree of \mathbf{S} is determined by λ , hyper-parameter q and $\boldsymbol{\alpha}$ to satisfy the different situations.

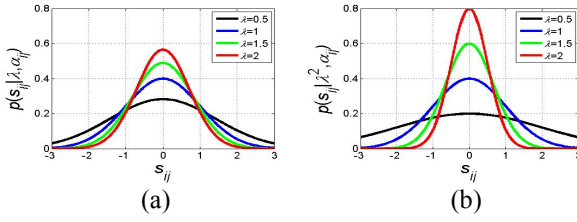


Fig. 2: The Gaussian distribution $p(s_{ij} | \lambda^q, \alpha_{ij})$ of (a) λ with $\lambda=0.5, 1, 1.5, 2$. (b) λ^2 with $\lambda=0.5, 1, 1.5, 2$.

In (5), the conditional distribution is expressed as:

$$p(\mathbf{Y}' | \mathbf{U}, \mathbf{V}, \mathbf{S}, \beta, \lambda) = \mathcal{N}(\mathbf{Y}' | \mathbf{U}\mathbf{V}^T + \lambda\mathbf{S}, \beta^{-1}\mathbf{E}_{KN_d}) \quad (8)$$

where β is the precision of Gaussian distribution and follows the Jeffrey's priors [26], $p(\beta) = \beta^{-1}$. Therefore, the joint distribution is expressed as:

$$p(\mathbf{Y}', \mathbf{U}, \mathbf{V}, \mathbf{S}, \gamma, \alpha, \beta, \lambda) = p(\mathbf{Y}' | \mathbf{U}, \mathbf{V}, \mathbf{S}, \beta, \lambda) \times p(\mathbf{U} | \gamma) p(\mathbf{V} | \gamma) p(\mathbf{S} | \lambda, \alpha) p(\gamma) p(\alpha) p(\beta) \quad (9)$$

In this paper, the mean field VB inference [26] is adopted. The posterior distribution is evaluated by minimizing the Kullback-Leibler divergence in an alternating mechanism for each latent variable. This model uses a vector \mathbf{Z} to denote all latent variables, such as $\mathbf{Z} = (\mathbf{U}, \mathbf{V}, \mathbf{S}, \gamma, \alpha, \beta)$. The approximation of posterior distribution $q(Z_k) (k=1, \dots, 6)$ at each latent variable $Z_k \in \mathbf{Z}$ is expressed as:

$$\ln q(Z_k) = \langle \ln p(\mathbf{Y}', \lambda, \mathbf{Z}) \rangle_{\mathbf{Z} \setminus Z_k} + \text{const} \quad (10)$$

where $\langle \bullet \rangle$ denotes the expectation, $\mathbf{Z} \setminus Z_k$ denotes the set \mathbf{Z} with Z_k removed. The term *const* denotes a constant. $p(\mathbf{Y}', \lambda, \mathbf{Z})$ is the joint distribution of \mathbf{Y}' , λ and \mathbf{Z} . The posterior of each latent variable is independent.

i). Estimation of factors \mathbf{U} and \mathbf{V}

The posterior distribution of the i^{th} row of \mathbf{U} , which can be expressed as $\mathbf{u}_{i\bullet}$, obeys the multivariate Gaussian distribution and it can be denoted as:

$$q(\mathbf{u}_{i\bullet}) = \mathcal{N}(\mathbf{u}_{i\bullet} | \langle \mathbf{u}_{i\bullet} \rangle, \Sigma^U) \quad (11)$$

The covariance and mean are expressed as follows:

$$\Sigma^U = (\langle \beta \rangle \langle \mathbf{V}^T \mathbf{V} \rangle + \Gamma)^{-1} \quad (12)$$

$$\langle \mathbf{u}_{i\bullet} \rangle^T = \langle \beta \rangle \Sigma^U \langle \mathbf{V} \rangle^T (\mathbf{y}'_{i\bullet} - \lambda \mathbf{s}_{i\bullet})^T \quad (13)$$

where $\Gamma (\Gamma = \text{diag}(\gamma_1, \dots, \gamma_r))$ is a diagonal matrix. Similarly, the posterior distribution of the j^{th} row of \mathbf{V} is expressed as $\mathbf{v}_{j\bullet}$ and obeys the multivariate Gaussian distribution:

$$q(\mathbf{v}_{j\bullet}) = \mathcal{N}(\mathbf{v}_{j\bullet} | \langle \mathbf{v}_{j\bullet} \rangle, \Sigma^V) \quad (14)$$

The covariance and mean are denoted as:

$$\Sigma^V = (\langle \beta \rangle \langle \mathbf{U}^T \mathbf{U} \rangle + \Gamma)^{-1} \quad (15)$$

$$\langle \mathbf{v}_{j\bullet} \rangle^T = \langle \beta \rangle \Sigma^V \langle \mathbf{U} \rangle^T (\mathbf{y}'_{j\bullet} - \lambda \mathbf{s}_{j\bullet}) \quad (16)$$

where $\langle \mathbf{U}^T \mathbf{U} \rangle$ and $\langle \mathbf{V}^T \mathbf{V} \rangle$ can be computed by combining the mean, the correlation coefficient and the covariance. \mathbf{L} can be computed as follows:

$$\mathbf{L} = \langle \mathbf{U} \rangle \langle \mathbf{V} \rangle^T \quad (17)$$

ii). Estimation of γ

The posterior distribution of γ_j is a Gamma distribution and the mean estimation is expressed as:

$$\langle \gamma_j \rangle = \frac{2a_1 + K + N_d}{2b_1 + \langle \mathbf{u}_{j\bullet}^T \mathbf{u}_{j\bullet} \rangle + \langle \mathbf{v}_{j\bullet}^T \mathbf{v}_{j\bullet} \rangle} \quad (18)$$

The required expectations are given by

$$\langle \mathbf{u}_{j\bullet}^T \mathbf{u}_{j\bullet} \rangle = \langle \mathbf{u}_{j\bullet} \rangle^T \langle \mathbf{u}_{j\bullet} \rangle + K (\Sigma^U)_{jj} \quad (19)$$

$$\langle \mathbf{v}_{j\bullet}^T \mathbf{v}_{j\bullet} \rangle = \langle \mathbf{v}_{j\bullet} \rangle^T \langle \mathbf{v}_{j\bullet} \rangle + N_d (\Sigma^V)_{jj} \quad (20)$$

iii). Estimation of \mathbf{S}

The posterior distribution of \mathbf{S} can be decomposed into each coefficient s_{ij} . The derivation can be expressed as:

$$\ln q(\mathbf{S}) = \langle \ln p(\mathbf{Y}', \lambda, \mathbf{Z}) \rangle_{\mathbf{Z} \setminus \mathbf{S}} + \text{const}$$

$$q(s_{ij}) \propto \exp\left(-\frac{1}{2}((\lambda^2 \langle \beta \rangle + \lambda^q \langle \alpha_{ij} \rangle) s_{ij}^2 - 2\lambda \langle \beta \rangle (y'_{ij} - l_{ij}) s_{ij})\right) \quad (21)$$

In (21), s_{ij} obeys a Gaussian distribution and can be denoted as:

$$q(s_{ij}) = \mathcal{N}(s_{ij} | \langle s_{ij} \rangle, \Sigma^S_{ij}) \quad (22)$$

If a scalar value a obeys a Gaussian distribution whose mean is u and variance is σ^2 , then it can be expressed as:

$$p(a) \propto \exp\left(-\frac{1}{2\sigma^2}(a - u)^2\right) \quad (23)$$

$$p(a) \propto \exp\left(-\frac{1}{2\sigma^2}(a^2 - 2au)\right)$$

Combining (22) and (23), the covariance and mean are denoted as follows:

$$\Sigma^S_{ij} = \frac{1}{\lambda^2 \langle \beta \rangle + \lambda^q \langle \alpha_{ij} \rangle} \quad (24)$$

$$(\Sigma^S_{ij})^{-1} \langle s_{ij} \rangle = \lambda \langle \beta \rangle (\langle y'_{ij} - l_{ij} \rangle) \quad (25)$$

$$\langle s_{ij} \rangle = \frac{\lambda \langle \beta \rangle}{\lambda^2 \langle \beta \rangle + \lambda^q \langle \alpha_{ij} \rangle} (\langle y'_{ij} - \langle \mathbf{u}_{i\bullet} \rangle \langle \mathbf{v}_{j\bullet} \rangle^T)$$

iv). Estimation of α and noise precision β

The posterior probability distribution of α_{ij} obeys a Gamma distribution and the mean of α_{ij} is expressed as:

$$\langle \alpha_{ij} \rangle = \frac{1}{\lambda^q (\langle s_{ij} \rangle^2 + \Sigma^S_{ij})} \quad (26)$$

The posterior probability distribution of β obeys a Gamma distribution and the mean of β is expressed as:

$$\langle \beta \rangle = \frac{KN_d}{\langle \|\mathbf{Y}' - \mathbf{U}\mathbf{V}^T - \lambda\mathbf{S}\|_F^2 \rangle} \quad (27)$$

where $tr(\bullet)$ denotes the trace operator and

$$\begin{aligned} \left\| \mathbf{Y}' - \mathbf{U}\mathbf{V}^T - \lambda \mathbf{S} \right\|_F^2 &= \left\| \mathbf{Y}' - \langle \mathbf{U} \rangle \langle \mathbf{V} \rangle^T - \lambda \langle \mathbf{S} \rangle \right\|_F^2 + tr \left(N_d \langle \mathbf{U} \rangle^T \langle \mathbf{U} \rangle \mathbf{\Sigma}^V \right) \\ &+ tr \left(K \langle \mathbf{V} \rangle^T \langle \mathbf{V} \rangle \mathbf{\Sigma}^U \right) + tr \left(K N_d \mathbf{\Sigma}^U \mathbf{\Sigma}^V \right) + \lambda^2 \sum_{i=1}^K \sum_{j=1}^{N_d} \mathbf{\Sigma}_{ij}^S \end{aligned}$$

vi). Estimation of λ

Using (5), the function of λ can be expressed as:

$$f(\lambda) = \sum_i \sum_j l_{ij} + \lambda \sum_i \sum_j s_{ij} + \sum_i \sum_j n_{ij} - \sum_i \sum_j y'_{ij} \quad (28)$$

By adopting the gradient descent method, the update for λ assumes the form

$$\frac{df(\lambda)}{d\lambda} = \sum_i \sum_j s_{ij} \text{ and } \lambda^{n+1} = \lambda^n + \rho_2 \sum_i \sum_j s_{ij} \quad (29)$$

where ρ_2 is the learning rate and n denotes iteration time.

vi). Stopping criterion

There is a necessity to develop a stopping criterion for the adaptive sparsity control step. In this work, the sub-grouping strategy is proposed to guarantee that the algorithm terminates effectively. As \mathbf{S} is a sparse matrix where only a few of the elements take significant values, we propose to use the K -means clustering algorithm to separate the first column ($\mathbf{s}_{\cdot 1}$) of \mathbf{S} into two classes (i.e. C_j , $j=1,2$) and the members are labelled as $s_{i1} \in C_j$. Let us denote $\mathbf{c} = [c_1, c_2]^T$ as the clustering centroid locations where c_1 and c_2 denote the centroid location of the first class and second class, respectively. In addition, the within-cluster sums of point-to-centroid distances $\mathbf{d} \in \mathbb{R}^{2 \times 1}$ can be computed by Euclidean distance, where d_1 and d_2 denote the within-first-cluster sums of point-to-centroid distances and within-second-cluster sums of point-to-centroid distances, respectively. The iteration terminates when the ratio $R = d_1/d_2$ is less than a small threshold i.e. $R \leq \delta$ (e.g. 10^{-6}).

The calculations for d_1 , d_2 and R are given by

$$d_j = \left(\sum_{i, s_{i1} \in C_j} (s_{i1} - c_j)^2 \right)^{\frac{1}{2}} \text{ and } R = \frac{d_1}{d_2} \quad \begin{cases} i=1, \dots, K \\ j=1, 2 \end{cases} \quad (30)$$

The specific steps of the proposed method can be summarized in Table I.

Table I: Proposed sparse pattern extraction

Input: A matrix $\mathbf{Y}' \in \mathbb{R}^{K \times N_d}$, which denotes N_d principal components of ECPT thermal video.

Procedure:

Initialization: $\mathbf{U}, \mathbf{V}, H, \lambda, \delta, \rho_2, q, R$

while $R \leq \delta \parallel R \geq \delta^{-1}$

for $h=1:H$

Update the posterior \mathbf{u}_{\cdot} by (13)

Update the posterior \mathbf{v}_{\cdot} by (16)

Update the posterior \mathbf{L} by (17)

Update the posterior γ_j by (18)

Update the posterior s_{ij} by (25)

Update the posterior α_{ij} by (26)

Update the posterior β by (27)

end

Update λ by (29)

Update d_1, d_2 and R by (30)

Note: MATLAB[®] demo code of the proposed method can be found in <http://faculty.uestc.edu.cn/gaobin/en/lwgc/153408/list/index.htm>

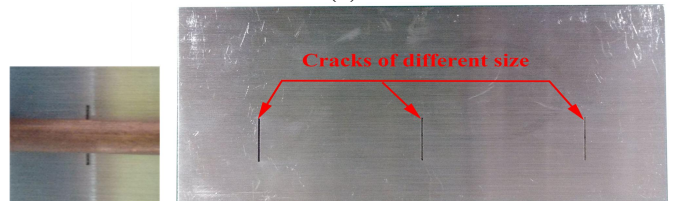
III. EXPERIMENT SETUP

A. Sample preparation and experiments setup

The experimental setup is shown in Fig.3. An Easyheat 224 from Cheltenham Induction Heating is used for coil excitation. The Easyheat has a maximum excitation power of 2.4 kW, a maximum current of 400 A and an excitation frequency range of 150-400 kHz (380 Arms and 256 kHz are used in this study). The IR camera, FLIR A655sc is equipped with an uncooled, maintenance free, Vanadium Oxide (VoX) microbolometer detector that produces thermal images of 640 x 480 Pixels. These pixels generate crisp and clear detailed images that are easy to interpret with high accuracy. Two kinds of samples are prepared: 1) Three stainless steel samples (120mm×60mm×5mm) with different size of cracks have been prepared (i.e. one sample is shown in Fig. 3(c)). In the experiment, the coil is placed in the middle of the crack, which can be seen in Fig. 3(b). In this study, the frame rate of 100 Hz is chosen and 200 millisecond videos are recorded in the experiments. 2) Thermal natural fatigue cracks in a steel blade are provided by Alstom for validation. In the blade, flaws are produced in-situ with controlled thermal fatigue loading. In this study, one natural crack, 150BBB1353, is used for testing. The crack location is marked with red circles in Fig. 3(e). A Helmholtz coil is selected for inspection. In the study, the setting $q=2$ is used. The event based F-score is used for evaluating the detection performance of the different algorithms [29]. The definition of the F-score can be visualized in Appendix A.



(a)



(b)

(c)

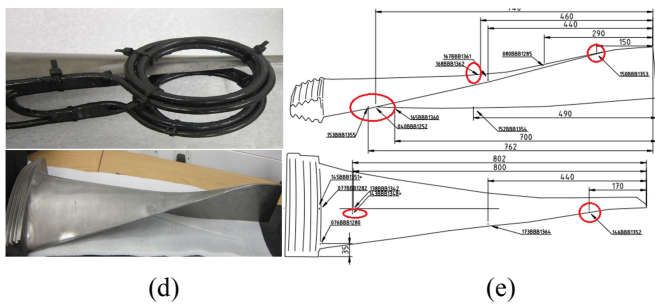


Fig. 3: (a) Inductive thermography system. (b) The coil is placed in the middle of the crack. (c) Steel test sample with artificial cracks. (d) Steel blade with thermal fatigue natural cracks. (e) Natural cracks location map

IV. RESULTS AND DISCUSSION

A. Comparison of common adopted thermal feature extraction methods

General thermal based defect feature extraction methods are used for comparison purposes. These include the method of manual selection of the original thermal image for defect detection, ICA [13], Pulsed Phase Thermography (PPT) [30], Thermographic Signal Reconstruction (TSR) [31] and PCA [32]. Fig. 4 shows one example of an artificial crack with a 2.8mm depth to show the comparison results. It can be seen that most extraction methods highlight the singular pattern

around the tip portion of the slot. However, the proposed method enables a heightened state of emphasis with higher resolution in the zone around the tip portion.

Fig. 4(a) refers to the result obtained by manual selection of the original thermal image frame. The second and the fifth red circles mark the defect regions. It is also noted that the results of the aforementioned adopted thermal feature extraction methods are heavily influenced by the background and noise. Specifically, TSR (Fig. 4(d)) and ICA (Fig. 4(b)) preserve both defect and unwanted coil edge information, where the results are influenced by the background. PPT (Fig. 4(c)) performs even worse than the manual selection of original thermal image and it fails to remove the influence of the heating from the coil edge. PCA (Fig. 4(e)) only extracts the highest temperature information around the crack tips. However, the extracted pattern loses part of the defect information. Fig. 4(f) presents the result of the proposed method. The contrast between defect and non-defect patterns is clearly visible and the proposed method has retained superior performance over the other methods. To validate the proposed system, a natural fatigue crack (a 1 mm length crack) in the steel blade is used for testing. With prior knowledge of other NDT techniques, the hot spot of crack is located by using human judgment and can be visually identified in Fig. 5(a). However, it is extremely difficult to detect the crack due to the complex geometric shape and the small size of the defect. Fig. 5 shows the comparison study of natural crack detection.

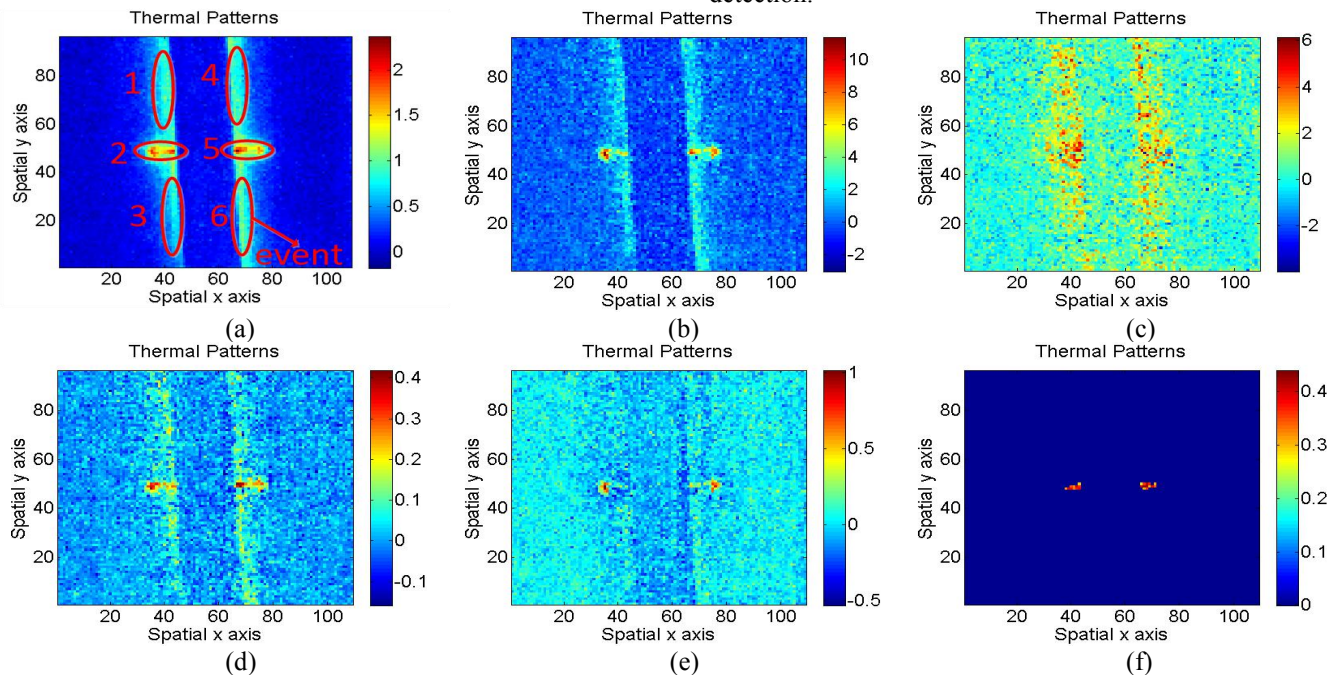


Fig. 4: Artificial crack with 2.8mm depth thermal patterns of (a) Original thermal image, (b) ICA, (c) PPT, (d) TSR, (e) PCA, (f) Proposed method

Fig. 5(a) is the human selection of the original thermal image. In comparison, it can be clearly seen that the selection method of the thermal image, ICA, TSR and PCA methods fail to determine the correct spatial pattern of defect. From Fig. 5 (a) to Fig. 5 (e), the panels show a considerable level of mixing ambiguities that have not been accurately resolved. The PPT works acceptably with defect location. However, the

extracted singular pattern has an issue of pattern dispersion (in reality, the 1mm crack only contains a few pixels) and it has not fully reduced the background interface. Moreover, it requires human selection with a specific frequency frame for visualization. On the other hand, the proposed method has not only successfully extracted the defect spatial pattern with high accuracy but also almost completely suppressed the

background interference. In order to quantitatively evaluate the results, the event based F-score is computed. The different regions of the extracted thermal patterns are considered as different events and the purpose is to objectively evaluate the detectability as well as the ability of noise suppression for each method. The examples of events organization are shown in Fig. 4(a), which is the standard template of events

arrangement. Fig. 5(a) is the standard template of events arrangement for the steel blade with thermal fatigue natural crack. The events in region 2 and 5 in Fig. 4(a) represent the defect events, whereas the events in region 1, 3, 4 and 6 denote the interference. In Fig. 5(a), the event in region 5 is the defect event and other circled regions are interference.

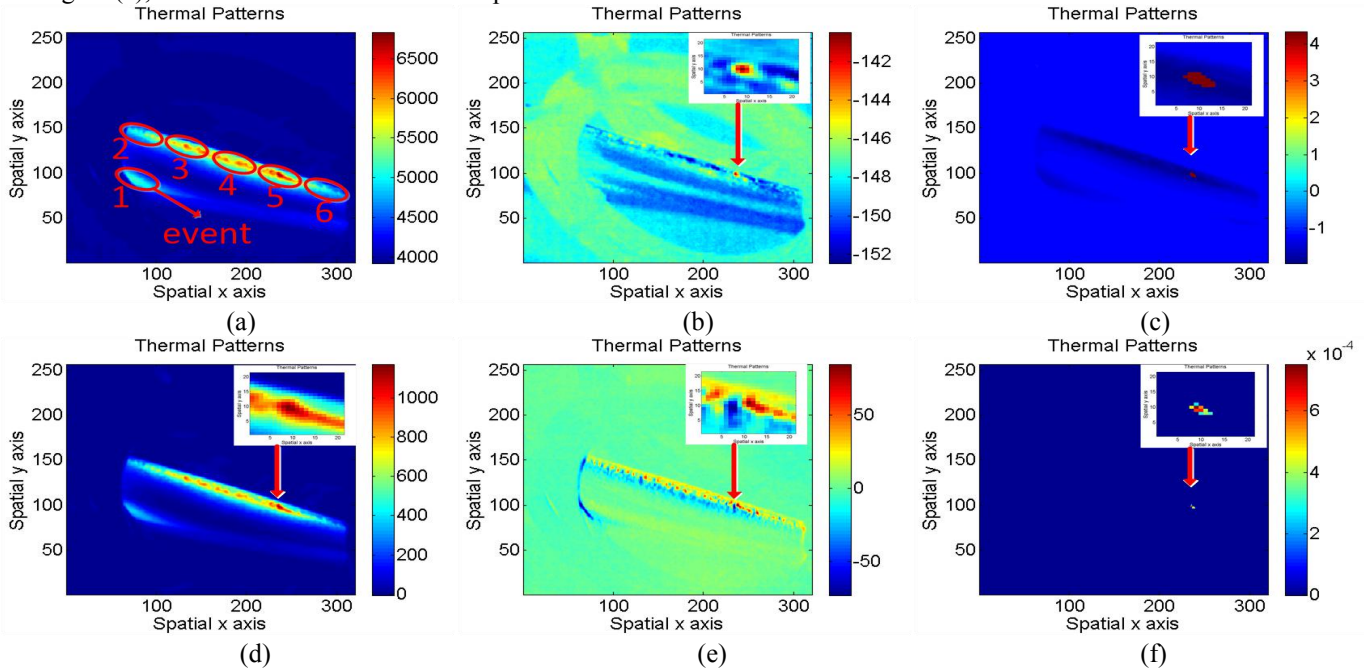


Fig. 5: Natural crack thermal patterns of (a) Original thermal image, (b) ICA, (c) PPT, (d) TSR, (e) PCA, (f) Proposed method

Table II: Performance comparison of F-score

	Natural crack	Artificial cracks(different depths)		
		13.mm	2.8mm	3.5mm
ICA	0.67	0.50	0.50	0.50
PPT	1.00	0.00	0.29	0.00
TSR	0.29	0.29	0.50	0.00
PCA	0.33	0.00	0.50	0.80
Proposed	1.00	1.00	1.00	1.00

The F-score is summarized in Table II. All event selections are based on human annotation, which are termed as ground truth.

The F-score has been calculated for detecting artificial defects with different depths and natural crack, respectively. The results of TSR, PCA and PPT give the worst performance since the F-score falls below 50% on average. The ICA gives mediocre performance with an average F-score of around 50%. The proposed method has significantly improved on the F-score rate for all artificial defects. In addition, the average improvement is more than 60% compared with other methods.

B. Comparison of different sparse decomposition methods

The previous sections have shown that the sparse pattern extraction plays an important role in quantitatively analyzing the cracks. This section compares the proposed method with other well-known sparse pattern extraction algorithms for defect detection. These algorithms are the greedy sparse PCA [25], VB sparse PCA [26], MCMC sparse PCA [27] and BRTF [33]. The results are compared in terms of accuracy with the same specimen. In our proposed method, the sparse

pattern extraction is applied by updating the sparse control parameters that gives superior results. Fig. 6 and Fig. 7 show the extraction results.

Fig. 6 indicates that without sparse control, the edge of the crack tips is difficult to be detected and has a worse effect on the quantitative sizing of the defect. This is attributed to the greedy sparse PCA, where the decomposition has the issue of over-sparsity such that all information has been removed from the extracted pattern. The MCMC sparse PCA emphasizes not only hot spot regions but also background and noise. The edge along the hot spots has been incorrectly detected. In terms of validation, the obtained results indicate that the greedy sparse PCA, MCMC sparse PCA, VB sparse PCA and BRTF methods lead to poor accuracy and the results are highly influenced by the background information. The F-score has been summarized in Table III. The results for greedy sparse PCA, VB sparse PCA, BRTF and MCMC sparse PCA give inferior performance since the F-score falls below 50% on average. In contrast, the proposed method has significantly improved on the F-score for both artificial defects and the natural crack. The average improvement is more than 60% better compared with the other methods.

In summary, the automatic sparsity control is necessary for the attainment of optimal sparse pattern extraction. The uniform constant sparsity control raises a consequential issue, since it is not possible to determine a priori, which decomposition should be assigned the degree of sparseness [34, 35]. This poses a difficult problem for sparse decomposition methods that require manual setting of the sparsity parameters.

Although the update parameters have the advantages of incorrect selection of prior distribution for the model bypassing human intervention, it introduces the drawbacks of parameters.

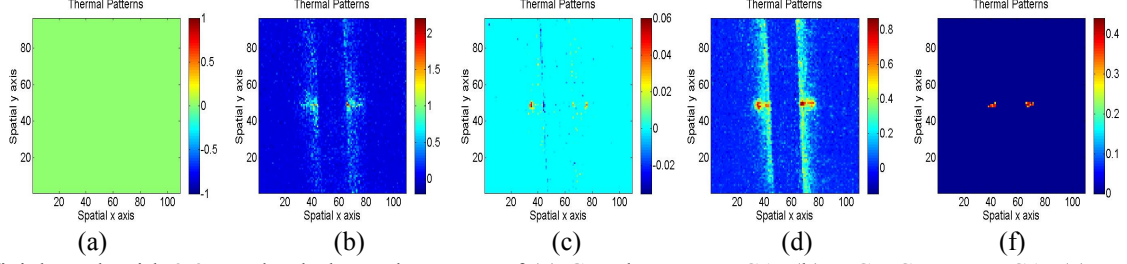


Fig. 6 Artificial crack with 2.8mm depth thermal patterns of (a) Greedy sparse PCA, (b) MCMC sparse PCA, (c) VB sparse PCA, (d) BRTF, (e) Proposed method, (f) Proposed method.

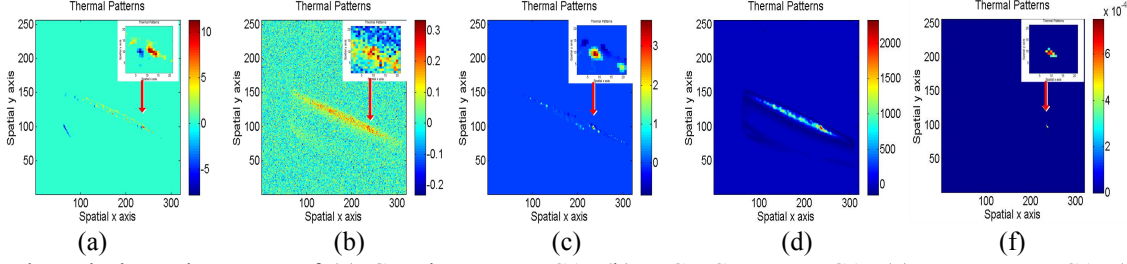


Fig. 7 Natural crack thermal patterns of (a) Greedy sparse PCA, (b) MCMC sparse PCA, (c) VB sparse PCA, (d) BRTF, (e) Proposed method, (f) Proposed method.

Table III: The F-score by different sparse methods

	Natural crack	Artificial cracks(different depths)		
		13.mm	2.8mm	3.5mm
GSPCA	0.33	0.00	0.00	0.00
MCMCSPCA	0.29	0.50	0.50	0.50
VBSPCA	0.33	0.29	0.50	0.50
BRTF	0.33	0.50	0.50	0.50
Proposed	1.00	1.00	1.00	1.00

C. Computational complexity of the proposed method

The computational complexity of each iteration is $O(N_d r^2 + r^3)$ for \mathbf{U} , $O(Kr^2 + r^3)$ for \mathbf{V} , $O(N_d Kr)$ for \mathbf{S} , $O((N_d + K) \times r)$ for γ , $O(N_d K)$ for α , $O((N_d + K)r^2 + r^3 + N_d Kr)$ for β and $O(N_d K)$ for λ . The total computational complexity of each iteration is $O(r^3 + (N_d + K)r^2)$. Therefore, the total computational complexity of the proposed method is $O(nHr^3 + (N_d + K)Hnr^2)$. Table IV shows the running times (on a 3.3GHz Core(TM) i5-4590 Intel CPU) and the unit is seconds.

Table IV: The running times by different sparse methods

	Natural crack	Artificial cracks(different depths)		
		13.mm	2.8mm	3.5mm
ICA	0.8	0.5	0.6	0.4
PPT	127.3	8.6	8.2	8.3
TSR	66.3	8.4	7.9	8.1
PCA	0.6	0.3	0.3	0.3
GSPCA	0.8	0.2	0.1	0.1
MCMCSPCA	2078.0	261.6	218.0	210.2
VBSPCA	12.0	1.2	1.2	1.2
BRTF	25.3	2.2	3.0	3.1
Proposed	8935.5	22.5	23.7	21.3

V. CONCLUSIONS

In this paper, a Variational Bayesian sub-group adaptive sparse component extraction algorithm has been proposed for diagnostic thermal NDT&E imaging for the first time. The physical interpretation of thermal patterns as well as the sparse decomposition has been established. The proposed sparse pattern extraction method allows abnormal patterns to be extracted automatically for flaw contrast enhancement. The proposed method is able to reduce interference from the background. In order to validate the algorithm, specimens with natural crack as well as artificial cracks of different depths provided by our industrial partners have been used. In this work, the F-score has been used to objectively evaluate the performance of the different methods. Compared with the other methods, the proposed method has significantly improved the accuracy of the defect detection by 60% in terms of the F-score.

Future optimization of the proposed method will focus on samples with natural cracks and complex surface conditions, e.g. roughness and emissivity variation. Image capturing and diagnostic of complex defect detection, e.g. subsurface defects in metallic material, will also be investigated.

VI. ACKNOWLEDGEMENT

The work was supported by National Natural Science Foundation of China (No. 61401071, No. 61527803), Supported by NSAF (Grant No. U1430115) and EPSRC IAA Phase 2 funded project: "3D super-fast and portable eddy current pulsed thermography for railway inspection" (EP/K503885/1)

REFERENCES

- [1] X. Li, S. K. Tso, X. P. Guan, Q. Huang, "Improving Automatic Detection of Defects in Castings by Applying Wavelet Technique", *IEEE Transactions on Industrial Electronics*, vol. 53, no.6, pp. 1927 -

- 1934, 2006.
- [2] T. Y. Li, J. Z. Tsai, R. S. Chang, L. W. Ho, C. F. Yang, "Pretest Gap Mura on TFT LCDs Using the Optical Interference Pattern Sensing Method and Neural Network Classification", *IEEE Transactions on Industrial Electronics*, vol. 60, no.9, pp. 3976 - 3982, 2013.
- [3] G. Acciani, G. Brunetti, and G. Fornarelli, "Application of neural networks in optical inspection and classification of solder joints in surface mount technology," *IEEE Transactions on Industrial Informatics*, vol. 2, no. 3, pp. 200-209, 2006.
- [4] 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 Transactions on Industrial Informatics*, vol. 5, no. 4, pp. 483-494, 2009.
- [5] D. M. Tsai, and J. Y. Luo, "Mean Shift-Based Defect Detection in Multicrystalline Solar Wafer Surfaces," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 1, pp. 125-135, 2011.
- [6] Chia-Feng Juang ; Guo-Cyuan Chen, "A TS Fuzzy System Learned Through a Support Vector Machine in Principal Component Space for Real-Time Object Detection", *IEEE Transactions on Industrial Electronics*, vol. 59, no.8, pp. 3309 - 3320, 2011.
- [7] Y. Chen; B. Wu; H. Huang; C. Fan, "A Real-Time Vision System for Nighttime Vehicle Detection and Traffic Surveillance", *IEEE Transactions on Industrial Electronics*, vol. 58, no.5, pp. 2030 - 2044, 2010.
- [8] B. Gao, Y. He, W. L. Woo, G. Y. Tian, J. Liu, and Y. Hu, "Multidimensional Tensor-Based Inductive Thermography With Multiple Physical Fields for Offshore Wind Turbine Gear Inspection," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 10, pp. 6305-6315, 2016.
- [9] B. Oswald-Tranta and G. Wally, "Thermo-inductive surface crack detection in metallic materials," in Proc. 9th Eur. Conf. NDT, Berlin, Germany, 2006, paper We.3.8.3.
- [10] G. Zenzinger, J. Bamberg, W. Satzger, and V. Carl, "Thermographic crack detection by eddy current excitation," *Nondestructive Testing & Evaluation*, vol. 22, no. 2-3, pp. 101-111, 2007.
- [11] B. Gao, W. L. Woo, Y. He, and G. Y. Tian, "Unsupervised Sparse Pattern Diagnostic of Defects With Inductive Thermography Imaging System," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 371-383, 2016.
- [12] W. Li, Y. Ye, K. Zhang, Z. Feng, "A Thickness Measurement System for Metal Films Based on Eddy-Current Method With Phase Detection", *IEEE Transactions on Industrial Electronics*, vol. 64, no.5, pp. 3940 - 3949, 2017.
- [13] B. Gao, W. L. Lok, and G. Y. Tian, "Electromagnetic Thermography Nondestructive Evaluation: Physics-based Modeling and Pattern Mining," *Scientific Reports*, vol. 6, pp. 25480, 2016.
- [14] M. Kim, "Efficient Kernel Sparse Coding Via First-Order Smooth Optimization," *IEEE Transactions on Neural Networks & Learning Systems*, vol. 25, no. 8, pp. 1447-1459, 2014.
- [15] J. Luo, C. M. Vong, and P. K. Wong, "Sparse Bayesian Extreme Learning Machine for Multi-classification," *IEEE Transactions on Neural Networks & Learning Systems*, vol. 25, no. 4, pp. 836-843, 2014.
- [16] J. Gui, Z. Sun, S. Ji, D. Tao, and T. Tan, "Feature Selection Based on Structured Sparsity: A Comprehensive Study," *IEEE Transactions on Neural Networks & Learning Systems*, vol. PP, no. 99, pp. 1-18, 2016.
- [17] D. Qiu, and Y. Liu, "Improved image super-resolution via sparse representation," *Video Engineering*, 2016.
- [18] J. Bobin, J. L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 16, no. 11, pp. 2662-2674, 2007.
- [19] E. J. Cand, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the Acm*, vol. 58, no. 3, pp. 1-37, 2009.
- [20] Peng, Yigang, et al. "RASL: Robust Alignment by Sparse and Low-Rank Decomposition for Linearly Correlated Images." *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 34, no. 11, pp. 2233-46, 2012.
- [21] Shen, Li, C. Yeo, and B. S. Hua. "Intrinsic Image Decomposition Using a Sparse Representation of Reflectance." *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 35, no. 12, pp. 2904-2915, 2013.
- [22] X. Jiang, and J. Lai, "Sparse and Dense Hybrid Representation via Dictionary Decomposition for Face Recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 37, no. 5, pp. 1067-79, 2015.
- [23] S. Shekhar, V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Joint Sparse Representation for Robust Multimodal Biometrics Recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 36, no. 1, pp. 113-26, 2014.
- [24] W. Hu, W. Li, X. Zhang, and S. Maybank, "Single and Multiple Object Tracking Using a Multi-Feature Joint Sparse Representation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 37, no. 4, pp. 816-33, 2015.
- [25] M. Chen, A. Ganesh, Z. Lin, Y. Ma, J. Wright, and L. Wu, "Fast Convex Optimization Algorithms for Exact Recovery of a Corrupted Low-Rank Matrix," *Journal of the Marine Biological Association of the UK*, vol. 56, no. 3, pp. 707-722, 2009.
- [26] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos, "Sparse Bayesian Methods for Low-Rank Matrix Estimation," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 3964-3977, 2012.
- [27] X. Ding, L. He, and L. Carin, "Bayesian Robust Principal Component Analysis," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 20, no. 12, pp. 3419-30, 2011.
- [28] G.J. Jang and T.W. Lee, "A maximum likelihood approach to single channel source separation," *Journal of Machine Learning Research*, vol. 4, pp. 1365-1392, 2003.
- [29] X. Li, B. Gao, W. L. Woo, G. Y. Tian, L. Gu, and X. Qiu, "Quantitative Surface Crack Evaluation based on Eddy Current Pulsed Thermography," *IEEE Sensors Journal*, vol. PP, no. 99, pp. 1-1, 2017.
- [30] C. Ibarra-Castaneda, and X. P. V. Maldague, "Interactive Methodology for Optimized Defect Characterization by Quantitative Pulsed Phase Thermography," *Research in Nondestructive Evaluation*, vol. 16, no. 4, pp. 175-193, 2005.
- [31] R. E. Martin, A. L. Gyekenyesi, and S. M. Shepard, "Interpreting the results of pulsed thermography data," *Materials Evaluation*, vol. 61, no. 5, pp. 611-616, 2003.
- [32] S. Marinetti, E. Grinzato, P. G. Bison, E. Bozzi, M. Chimenti, G. Pieri, and O. Salvetti, "Statistical analysis of IR thermographic sequences by PCA," *Infrared Physics & Technology*, vol. 46, no. 1, pp. 85-91, 2004.
- [33] Q. Zhao, G. Zhou, L. Zhang, A. Cichocki, and S. I. Amari, "Bayesian Robust Tensor Factorization for Incomplete Multiway Data," *IEEE Transactions on Neural Networks & Learning Systems*, vol. 27, no. 4, pp. 736-748, 2017.
- [34] Wei He and Shuzhi Sam Ge, "Cooperative Control of a Nonuniform Gantry Crane with Constrained Tension", *Automatica*, vol. 66, no. 4, pp. 146-154, 2016.
- [35] Wei He, Yuhao Chen and Zhao Yin, "Adaptive Neural Network Control of an Uncertain Robot with Full-State Constraints", *IEEE Transactions on Cybernetics*, vol. 46, no. 3, pp. 620-629, 2016.



Bin Gao (M' 12–SM' 14) received his B.Sc. degree in communications and signal processing from Southwest Jiao Tong University (2001-2005), China, MSc degree in communications and signal processing with Distinction and PhD degree from Newcastle University, UK (2006-2011). He worked as a Research Associate (2011-2013) with the same university on wearable acoustic sensor technology. Currently, he is a Professor with the School of Automation Engineering, University of Electronic Science and Technology of China (UESTC), China. His research interests include sensor signal processing, machine learning, social signal processing, nondestructive testing and evaluation where he actively publishes in these areas. He has published over 60 papers on these topics on various journals and international conference proceedings. He has coordinated several research projects from National Natural Science Foundation of China



Peng Lu received the B.Sc. degree from the School of Automation, Hangzhou Dianzi University (2011-2015), Hangzhou, China. He is currently pursuing the M.Sc. degree in Research Center of Non-destructive Evaluation and Structural Health Monitoring at the University of Electronic Science and Technology of China, Chengdu, China. His research interests include

sensor signal processing, image processing, tensor factorization and machine learning.



Wai Lok Woo was born in Malaysia. He received the BEng degree (1st Class Hons.) in Electrical and Electronics Engineering and the PhD degree from the Newcastle University, UK. He was awarded the IEE Prize and the British Scholarship to continue his research work. He is currently a Senior Lecturer and Director of Operations with the School of Electrical and Electronic Engineering. His major research is in the mathematical theory

and algorithms for nonlinear signal and image processing. This includes areas of machine learning for signal processing, blind source separation, multidimensional signal processing, signal/image deconvolution and restoration. He has an extensive portfolio of relevant research supported by a variety of funding agencies. He has published over 250 papers on these topics on various journals and international conference proceedings. Currently, he is Associate Editor of several international journals and has served as lead-editor of journals' special issues.



Gui Yun Tian (M'01-SM'03) received the B.Sc. degree in metrology and instrumentation and M.Sc. degree in precision engineering from the University of Sichuan, Chengdu, China, in 1985 and 1988, respectively, and the Ph.D. degree 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, U.K. Since 2007, he has been based at Newcastle University, Newcastle upon Tyne, U.K., where he has been Chair Professor in Sensor Technologies. Currently, He is also an adjunct professor with School of Automation Engineering, University of Electronic Science and Technology of China. He has coordinated several research projects from the Engineering and Physical Sciences Research Council (EPSRC), Royal Academy of Engineering and FP7, on top of this he also has good collaboration with leading industrial companies such as Airbus, Rolls Royce, BP, nPower, Networkrail and TWI.



Yuyu Zhu received the B.S. degree in automation from the Southwest University of Science and Technology, China, in 2002, and the M.Sc. degree in control theory and control engineering with the Southwest University of Science and Technology, in 2009. He is currently pursuing the Ph.D. degree with the University of Electronic Science and Technology of China. He is an Associate

Professor with the School on Monitoring and Control Technology and Power Electronics Technology. His research interests include nondestructive testing and evaluation, power electronics technology.

He has developed several equipment and coordinated several research projects.



Martin Johnston received his BSc (Hons) degree in Physics with Electronics from Birmingham University, UK, in 1999, his MSc degree in Electronic Engineering from Staffordshire University, UK, in 2001 and his PhD degree in 2006 from Newcastle University, UK. From 2006 - 2014 he worked as a Research Associate at the School of Engineering in Newcastle University

and he is now employed as a Lecturer. His research interests include the design of advanced error-correcting schemes and low-complexity decoding algorithms, vehicle-to-vehicle communications and physical-layer security.

Appendix

A The definition of *F-score*

The F-score is used as a measure of test accuracy, namely

$$F = \frac{(\zeta^2 + 1) \times \text{Recall} \times \text{Precision}}{\zeta^2 \times (\text{Recall} + \text{Precision})} \quad (31)$$

where ζ is a default value that applies relative weights to the Precision and Recall parameters, which is greater than one when Precision has a larger weight relative to Recall, otherwise it is less than one when Recall has higher weight. In this paper, ζ equal to one means Recall is as important as Precision.

The term Recall is defined as:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (32)$$

Similarly, Precision is expressed as:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (33)$$

where false positive (FP) refers to no defect existing but one is identified; True positive (TP) refers to a defect existing and is detected; True negative (TN) refers to no defect existing and none is detected and False negative (FN) refers to a defect existing but is not detected. Both FP and TN outcomes give the total opportunities for false alarms (effective number of no defects). On the other hand, TP and FN outcomes give the total opportunities for positive calls (effective number of defects).