VARIATIONAL BAYES SUB-GROUP ADAPTIVE SPARSE COMPONENT EXTRACTION FOR DIAGNOSTIC IMAGING SYSTEM

Bin Gao[1], Peng Lu[1], W.L. Woo[2], G.Y. Tian[1,2]

[1]School of Automation Engineering, University of Electronic Science and Technology of China, China.

[2]School of Electrical and Electronic Engineering, University of Newcastle, UK.

ABSTRACT

A novel unsupervised sparse component extraction algorithm for diagnosing micro defects in thermography imaging system is presented. The approach is optimized under Variational Bayesian framework, which is fully automated and does not require manual selection of the parameters in the solution. An internal sub sparse grouping mechanism and adaptive fine-tuning have been built into the proposed algorithm to control the sparsity. The proposed method is used to automatically detect the micro defects on metals. Other contending defect feature extraction and sparse pattern extraction methods are employed for comparison. The algorithm has been shown to improve the detection precision of both artificial and natural cracks.

Index Terms **—** Low-rank decomposition, variational Bayes, diagnostic imaging system, sparse decomposition.

1. INTRODUCTION

In recent years, the sparse decomposition has been widely used in applications such as image. The robust PCA (GSPCA) is proposed to separate the sparse patterns [1]. Peng *et al.* proposed sparse and low-rank decomposition for linearly correlated images [2]. Chen *et al.* proposed fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix [3]. The variational Bayesian (VB) sparse PCA and Markov chain Monte Carlo (MCMC) sparse PCA with specific prior are proposed for adaptive sparse decomposition [4-5]. In all cases, sparse treatment works well in limited application field where in such situation, the sparse decomposition will invariably suffer from either under- or over-sparseness which subsequently lead to ambiguity in extracting of target component. Thus, the above suggests that the present form of sparse control strategy is still technically lacking.

In this paper, a novel adaptive sub-group sparsity component decomposition method is proposed to extract anomalous patterns for micro-defects in the Eddy Current Pulsed Thermography (ECPT) system. Our proposed model allows: (i) Unlike the general model, the method imposes automatically sparseness control as well as rendering sub-group so that the decomposition can be iteratively optimized. This overcomes the problem of under- and over-sparse factorization. (ii) Both control parameter and decomposition is learned and adapted as part of the matrix factorization by using variational Bayes approach. The proposed method can significantly improve the detection precision of the defects, which has been demonstrated on the steel artificial and natural cracks.

The paper is organized as follows: Section 2 discusses the proposed methodology. Section 3 describes the experiment setup. The experimental results and discussion are presented in Sections 4. Finally, Section 5 describes the conclusions.

2. PROPOSED METHODOLOGY

2.1. Sparse Pattern Modeling and Extraction

The general model of factorization will invariably suffer from either under- or over-sparseness which subsequently lead to ambiguity in separating sparse patterns. In order to deal with the issue, the robust PCA can be replaced by the new sparse control model, which is expressed as [1]:

$$\mathbf{Y}' = \mathbf{L} + \lambda\mathbf{S} + \mathbf{N} \tag{1}$$

where λ is the parameter that controls the sparse level of $\mathbf{S}$. The algorithm uses an adaptive iteration algorithm to estimate the optimal $\mathbf{S}$ and λ.

$\mathbf{L}$ is the low-rank matrix, which is updated by using the factorization $\mathbf{L} = \mathbf{U}\mathbf{V}^{\mathbf{T}}$, where $\mathbf{U}$ is a $K \times r$ matrix, and $\mathbf{V}$ is a $N \times r$ matrix. $\mathbf{U}$ and $\mathbf{V}$ can be obtained by using the singular value decomposition.

In this work, $\mathbf{S}$ is the sparse matrix, each entry of $\mathbf{S}$ can be assumed to obey the independent Gaussian distribution, that is:

$$p(\mathbf{S} \mid \lambda, \boldsymbol{\alpha}) = \prod_i \prod_j \mathcal{N}(s_{ij} \mid 0, (\lambda^q \alpha_{ij})^{-1}) \tag{2}$$

where q is the hyperparameter and can be heuristically set. The appropriate q can be determined by using the Monte-Carlo approach. Therefore, it does not appear in other distribution formula. α_{ij} can be assumed to obey the Jeffrey's priors and can be denoted as:

$$p(\alpha_{ij}) = (\alpha_{ij})^{-1} \quad \forall i, j \tag{3}$$

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

In (1), the conditional distribution for observation is expressed as:

$$p(\mathbf{Y}' \mid \mathbf{U}, \mathbf{V}, \mathbf{S}, \beta, \lambda) = \mathcal{N}(\mathbf{Y}' \mid \mathbf{U}\mathbf{V}^T + \lambda\mathbf{S}, \beta^{-1}\mathbf{E}_{KN}) \tag{4}$$

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

$$\begin{aligned} p(\mathbf{Y}', \mathbf{U}, \mathbf{V}, \mathbf{S}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \beta, \lambda) &= p(\mathbf{Y}' \mid \mathbf{U}, \mathbf{V}, \mathbf{S}, \beta, \lambda) \\ &\times p(\mathbf{U} \mid \boldsymbol{\gamma}) p(\mathbf{V} \mid \boldsymbol{\gamma}) p(\mathbf{S} \mid \lambda, \boldsymbol{\alpha}) p(\boldsymbol{\gamma}) p(\boldsymbol{\alpha}) p(\beta) \end{aligned} \tag{5}$$

2.2. Parameters Learning Strategy by VB Method

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

$$q(\mathbf{u}_{i\bullet}) = \mathcal{N}(\mathbf{u}_{i\bullet} \mid \langle \mathbf{u}_{i\bullet} \rangle, \boldsymbol{\Sigma}^{\mathbf{U}}) \tag{6}$$

The covariance and mean are expressed as follows:

$$\mathbf{\Sigma^U} = (\langle\beta\rangle\langle\mathbf{V^T V}\rangle + \mathbf{\Gamma})^{-1} \tag{7}$$

$$\langle\mathbf{u}_{i\bullet}\rangle^{\mathbf{T}} = \langle\beta\rangle\mathbf{\Sigma^U}\langle\mathbf{V}\rangle^{\mathbf{T}}(\mathbf{y}''_{i\bullet} - \lambda\mathbf{s}_{i\bullet})^{\mathbf{T}} \tag{8}$$

where $\mathbf{\Gamma}(\mathbf{\Gamma} = diag(\gamma_1,...,\gamma_r))$ 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:

$$q(\mathbf{v}_{j\bullet}) = \mathcal{N}(\mathbf{v}_{j\bullet} \mid \langle\mathbf{v}_{j\bullet}\rangle, \mathbf{\Sigma^V}) \tag{9}$$

The covariance and mean are denoted as:

$$\mathbf{\Sigma^V} = (\langle\beta\rangle\langle\mathbf{U^T U}\rangle + \mathbf{\Gamma})^{-1} \tag{10}$$

$$\langle\mathbf{v}_{j\bullet}\rangle^{\mathbf{T}} = \langle\beta\rangle\mathbf{\Sigma^V}\langle\mathbf{U}\rangle^{\mathbf{T}}(\mathbf{y}''_{\bullet j} - \lambda\mathbf{s}_{\bullet j}) \tag{11}$$

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

$$\mathbf{L} = \langle\mathbf{U}\rangle\langle\mathbf{V}\rangle^{\mathbf{T}} \tag{12}$$

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

$$\langle\gamma_j\rangle = \frac{2a_1 + K + N}{2b_1 + \langle\mathbf{u}_{\bullet j}^{\mathbf{T}}\mathbf{u}_{\bullet j}\rangle + \langle\mathbf{v}_{\bullet j}^{\mathbf{T}}\mathbf{v}_{\bullet j}\rangle} \tag{13}$$

The required expectations are given by

$$\langle\mathbf{u}_{\bullet j}^{\mathbf{T}}\mathbf{u}_{\bullet j}\rangle = \langle\mathbf{u}_{\bullet j}\rangle^{\mathbf{T}}\langle\mathbf{u}_{\bullet j}\rangle + K\left(\mathbf{\Sigma^U}\right)_{jj} \tag{14}$$

$$\langle\mathbf{v}_{\bullet j}^{\mathbf{T}}\mathbf{v}_{\bullet j}\rangle = \langle\mathbf{v}_{\bullet j}\rangle^{\mathbf{T}}\langle\mathbf{v}_{\bullet j}\rangle + N\left(\mathbf{\Sigma^V}\right)_{jj} \tag{15}$$

The posterior distribution of s_{ij} follows a Gaussian distribution and can be denoted as:

$$q(s_{ij}) = \mathcal{N}(s_{ij} \mid \langle s_{ij}\rangle, \Sigma_{ij}^{\mathbf{S}}) \tag{16}$$

The covariance and mean are denoted as follows:

$$\Sigma_{ij}^{\mathbf{S}} = \frac{1}{\lambda^2\langle\beta\rangle + \lambda^q\langle\alpha_{ij}\rangle} \tag{17}$$

$$\begin{gathered}\left(\Sigma_{ij}^{\mathbf{S}}\right)^{-1}\langle s_{ij}\rangle = \lambda\langle\beta\rangle\langle\left(y'_{ij} - l_{ij}\right)\rangle \\ \langle s_{ij}\rangle = \frac{\lambda\langle\beta\rangle}{\lambda^2\langle\beta\rangle + \lambda^q\langle\alpha_{ij}\rangle}\left(y'_{ij} - \langle\mathbf{u}_{i\bullet}\rangle\langle\mathbf{v}_{j\bullet}\rangle^{\mathbf{T}}\right)\end{gathered} \tag{18}$$

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

$$\langle\alpha_{ij}\rangle = \frac{1}{\lambda^q(\langle s_{ij}\rangle^2 + \Sigma_{ij}^{\mathbf{S}})} \tag{19}$$

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

$$\langle\beta\rangle = \frac{KN}{\langle\left\|\mathbf{Y}'' - \mathbf{UV^T} - \lambda\mathbf{S}\right\|_F^2\rangle} \tag{20}$$

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

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

Using Eqn. (1), the function of λ can be expressed as:

$$f(\lambda) = \sum_i\sum_j l_{ij} + \lambda\sum_i\sum_j s_{ij} + \sum_i\sum_j n_{ij} - \sum_i\sum_j y'_{ij} \tag{21}$$

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

$$\frac{df(\lambda)}{d\lambda} = \sum_i\sum_j s_{ij} \text{ and } \lambda^{n+1} = \lambda^n + \rho\sum_i\sum_j s_{ij} \tag{22}$$

where ρ is learning rate and n denotes iteration time.

There is a necessity to develop a stopping criteria for adaptive sparse control. In this work, the sub-grouping strategy is proposed to guarantee the stop criteria.

$\mathbf{S}$ is a sparse matrix where most entries are near to zero, where only a few of those take significance. Specifically, the *K*-means clustering algorithm is used to separate $\mathbf{s}_{\bullet 1}$ into two classes, and gets the center as well as the label of s_{i1}. $\mathbf{c} = [c_1, c_2]^{\mathbf{T}}$ denotes the clustering centroid locations, c_1 denotes the centroid location of the first class, and c_2 denotes the centroid location of the second class. The within-cluster sums of point-to-centroid distances $\mathbf{d} \in \mathbb{R}^{2\times 1}$ are computed by Euclidean distance. d_1 denotes within-first-cluster sums of point-to-centroid distances, and d_2 denotes within-second-cluster sums of point-to-centroid distances. d_j $j \in [1,2]$ and R are computed by using Eqn. (30). $R \le \delta$ (e.g. 10^{-6}), the iteration terminates, namely:

$$d_j = \left(\sum_{(i,1)\in\mathbf{C}_j}(s_{ij} - c_j)^2\right)^{\frac{1}{2}} \text{ and } R = \frac{d_1}{d_2} \tag{23}$$

In summary, the specific steps of the proposed method can be summarized in Table 1.

Table 1. Proposed Sparse Pattern Extraction

Input: $\mathbf{Y'}$ matrix representation of Z principal components of ECPT thermal video.

Output: thermal low-rank pattern $\mathbf{UV^T}$, sparse pattern $\mathbf{S}$, optimal parameter λ

Procedure:

Initialize: $\mathbf{U} = \mathbf{A}\mathbf{\Sigma}^{\frac{1}{2}}, \mathbf{V^T} = \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{D^T}, M, \lambda_0, \delta, \rho, q,$ ratio

while $R \le \delta \;||\; R \ge \delta^{-1}$

$\lambda^{n+1} = \lambda^n + \rho\sum_i\sum_j s_{ij}$

for $k=1: M$

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

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

$\langle\gamma_j\rangle = \frac{2a + K + N}{2b + \langle\mathbf{u}_{\bullet j}\rangle^{\mathbf{T}}\langle\mathbf{u}_{\bullet j}\rangle + K\left(\mathbf{\Sigma^U}\right)_{jj} + \langle\mathbf{v}_{\bullet j}\rangle^{\mathbf{T}}\langle\mathbf{v}_{\bullet j}\rangle + N\left(\mathbf{\Sigma^V}\right)_{jj}}$

$\langle\alpha_{ij}\rangle = \frac{1}{\lambda^q(\langle s_{ij}\rangle^2 + \Sigma_{ij}^{\mathbf{S}})}$; $\langle\beta\rangle = \frac{KN}{\langle\left\|\mathbf{Y'} - \mathbf{UV^T} - \lambda\mathbf{S}\right\|_F^2\rangle}$

end

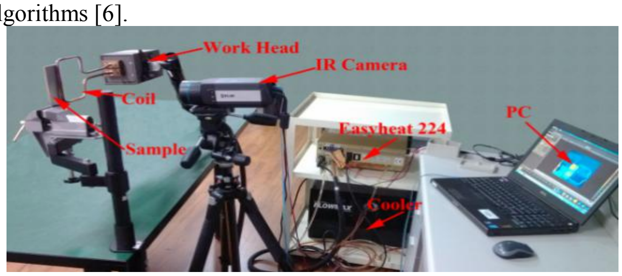
Compute d_1, d_2 in (23) by using the K-means algorithm.

$R = d_1/d_2$

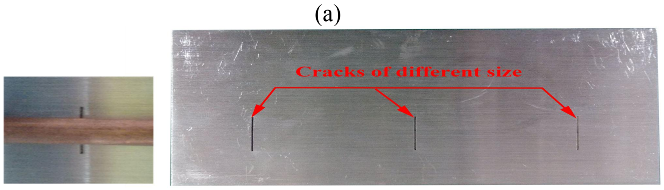
Note: MATLAB© demo code of the proposed method can be found in http://faculty.uestc.edu.cn/gaobin/en/lwcg/153408/list/index.htm

3. EXPERIMENT SETUP

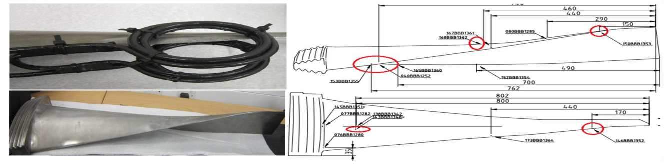
The experimental setup is shown in Fig.1. An Easyheat 224 from Cheltenham Induction Heating is used for coil excitation. The Easyheat has a maximum excitation power of 2.4 kW, a maximum current of 400 Arms and an excitation frequency range of 150-400 kHz (380 Arms and 256 kHz are used in this study). 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. The FLIR A655sc will make temperature differences as small as 50 mK clearly visible. Two kinds of samples are prepared: 1) three stainless steel samples (120mm×60mm×5mm) with three different size of cracks in each have been prepared (i.e. one sample is shown in Fig. 1c). In the experiment, coil is placed in the middle of the crack which can be seen in Fig.3b. 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 steel blade is 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. 1e. A Helmholtz coil is selected for inspection. In study, the setting q=2 . The event based F-score is used for evaluating the detection performance of the different algorithms [6].



(a)



(b) (c)



(d) (e)

Fig. 1: (a) Inductive thermography system. (b) The coil which is placed in the middle of the crack. (c) Steel test sample with artificial cracks. (d) Steel blade with thermal fatigue natural crack. (e) Natural crack location map

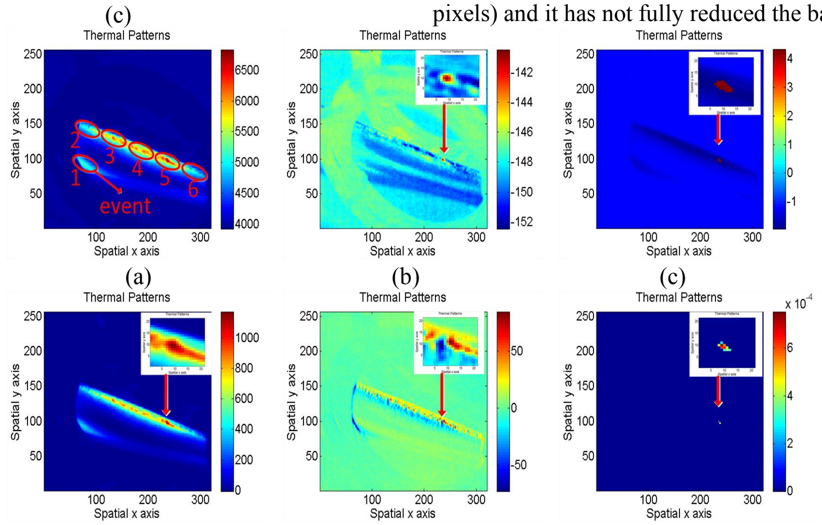
4. RESULTS AND DISCUSSION

4.1. Comparison of Common Adopted Thermal Feature Extraction Methods

General thermal based defect feature extraction methods are employed for comparison. These include manually selection of original thermal image for defect detection, Independent Component Analysis (ICA), Pulsed Phase Thermography (PPT), Thermographic Signal Reconstruction (TSR), and Principle Component Analysis (PCA) [7-10].

The contrast between defect and non-defect patterns is clearly visible, the proposed method has retained superior performance than other methods. To verify the proposed system, thermal fatigue nature crack (a 1 mm length crack) in steel blade is used for testing. With prior knowledge of other NDT technique, the hot spot of crack is located by human judgment and can be visually identified in Fig. 2(a). However, it is extremely difficult for human detection due to the complex geometric shape and crack are significantly small. These challenges also indicate that the target information in thermal images suffer significantly from background and noise. Fig. 2 shows the comparison study of natural crack detection.

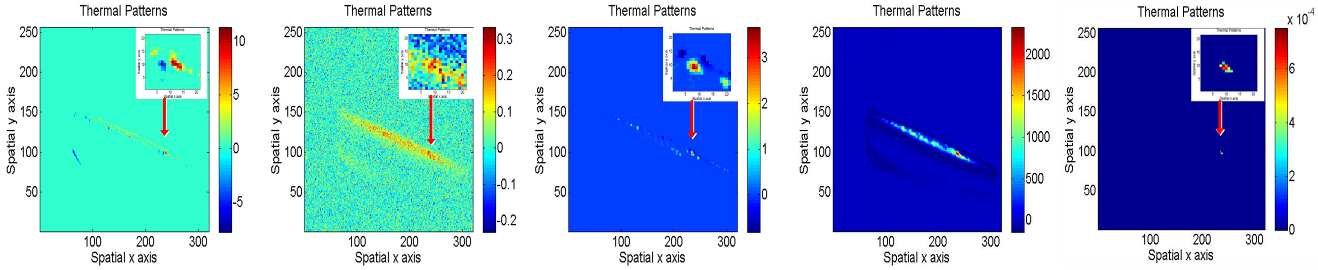
Fig. 2(a) is the human selection of original thermal image. In comparison, it is clearly seen that the selection method of thermal image, ICA, TSR and PCA methods fail to determine the correct spatial pattern of defect. From Fig. 2 (a) to Fig. 2 (e) panels, they show a considerable level of mixing ambiguities which have not been accurately resolved. The PPT works acceptable with defect location. However, the extracted singular pattern has issue of pattern dispersion (In reality, the 1mm crack only contains few pixels) and it has not fully reduced the background interface.



(a) (b) (c)

(d) (e) (f)

Fig. 2: Natural crack thermal patterns of (a) Original thermal image, (b) ICA, (c) PPT, (d) TSR, (e) PCA, (f) Proposed method



(a) (b) (c) (d) (f)

Fig. 3: Natural crack thermal patterns of (a) Greedy sparse PCA, (b) MCMC sparse PCA, (c) VB sparse PCA, (d) BRTF, (f) the proposed method

Moreover, it requires human selection with specific frequency frame for visualizing. On the other hand, the proposed methods has successfully not only extracted defect spatial pattern with high accuracy but also completely suppress the background interface. In order to quantitative evaluate the results, the event based F-score is computed. Fig. 2(a) is the standard template of events arrangement for steel blade with thermal fatigue natural crack. In Fig. 2(a), the event of 5 is the defect event, and others are interference. The F-score of the natural crack and the artificial crack is summarized in Table 2. All events selection are based on human annotation which are termed as ground truth.

Table 2: 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 has been calculated for detecting artificial defects with different depth and natural cracks, respectively. The results for TSR, PCA and PPT give the worst performance since F-score falls below 50% in average. The ICA gives mediocre performance with an average F-score around 50%. The proposed method have significantly improved the F-score rate for all artificial defects. In addition, the average improvement is more than 60% compared with other methods.

4.2. Comparison of Different Sparse Decomposition Methods

Previous sections prove that the sparse patterns extraction takes important role in quantitatively analyzing the cracks. This section compares the proposed method with other well-known sparse pattern extraction algorithms on defect detection. They are the Greedy Sparse PCA, Variational Bayesian (VB) sparse PCA, MCMC sparse PCA and BRTF [11]. The results are compared in term of accuracy with the same specimen. In our proposed method, the sparse patterns extraction is applied by updating the sparse control parameters that gives superior results. Fig. 3 show the extraction results.

In terms of validation, the obtained result indicated that the greedy sparse PCA, MCMC sparse PCA, VB sparse PCA, and BRTF methods lead to poor accuracy and the result is highly influenced by the background information. The F-score has been summarized in Table 3. The results for greedy sparse PCA, VB sparse PCA, BRTF and MCMC sparse PCA give worse performance since the F-score falls below 50% in average. By contrast, the proposed method has significantly improved the F-score for both artificial defects and natural crack where the average improvement is more than 60% better compared with the other methods.

Table 3: 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

In summary, the automatic sparseness control is necessary for the attainment of optimal sparse pattern exaction. The uniform constant sparsity control raises a consequential issue, since it is not possible to determine a priori which the decomposition should be assigned the degree of sparseness. This poses a difficult problem in conventional methods which requires manual setting of the sparsity parameters such as greedy sparse PCA. For MCMC sparse PCA and Variational Bayesian sparse PCA, although the update parameters have advantages to bypass human intervention whereas it brings the drawbacks to the incorrect selection of prior distribution for the model parameters.

5. CONCLUSIONS

In this paper, variational Bayes sub-group adaptive sparse component extraction algorithm has been proposed for thermal NDT&E. The physics interpretation of thermal patterns as well as the sparse decomposition has been conducted. The proposed sparse pattern extraction method allows abnormal pattern to be extracted automatically for flaw contrast enhancement. The proposed method has been able to reduce interference from background information. In order to validate the algorithm, the natural crack and the artificial crack with the different depth have been used. In this work, F-score has been used to objective measurement the performance of the different methods. Compared with the other methods, the proposed method has significantly improved the accuracy of the defect detection by approximately 60%.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] B. Gao, W. W. Lok, and G. Y. Tian, "Electromagnetic Thermography Nondestructive Evaluation: Physics-based Modeling and Pattern Mining," *Scientific Reports*, vol. 6, pp. 25480, 2016.

[8] C. Ibarra-Castanedo, 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.

[9] 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.

[10] 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.

[11] 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.