

## Exploratory Factor Analysis of Infrared Thermographic Data for Defect Detection

by Kai-Lun Huang\*, Stefano Sfarra\*\*, Yuan Yao\*†

\* Department of Chemical Engineering, National Tsing Hua University, Hsinchu, 30013, Taiwan, ROC,

\*\* Department of Industrial and Information Engineering and Economics, University of L'Aquila, Piazzale E. Pontieri no. 1, Monteluco di Roio – L'Aquila (AQ), I-67100, Italy

† Corresponding author: yyao@mx.nthu.edu.tw

### Abstract

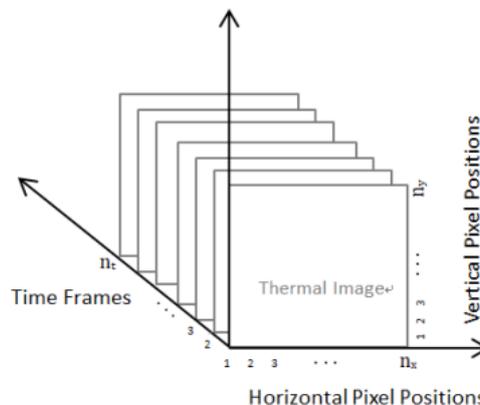
Active infrared thermography is an important non-destructive testing method used for revealing defect structures in materials. In many applications, thermographic data processing is necessary to extract defect features from a large number of thermal images. In this work, it is proposed to use exploratory factor analysis (EFA) for dimensionality reduction and feature extraction of thermographic data. By means of factor rotation, EFA minimizes the complexity of factor loadings and makes the results more interpretable. Hence, the defect information is highlighted while large signal-to-noise ratios are obtained. The feasibility of proposed method is illustrated with the experiment on a panel painting.

### 1. Introduction

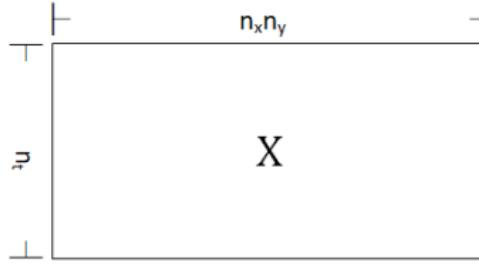
Active infrared thermography (AIRT) [1], which uses an external energy source to heat the investigated object and records the thermal images reflecting the material surface temperature by infrared camera, is widely adopted in non-destructive testing (NDT) for rapidly revealing defect structures. To highlight the defect features, thermographic data analysis methods are a necessity in many applications. In recent years, multivariate statistical methods, such as principal component analysis (PCA) [2], have been incorporated in this field [3] because of their abilities of dimensionality reduction and feature extraction. However, a main drawback of PCA is the lack of sparsity, which means that all elements in the loading vectors are typically non-zero. This characteristic limits the interpretability of the results and is not conducive to separating the defect information from the image backgrounds. Sparse PCA [4-6] can be adopted to overcome the limitation of PCA, while a remaining problem is how to specify the parameter controlling the sparsity. In this work, exploratory factor analysis (EFA) [7] is utilized as a better alternative of PCA for thermographic data analysis. By rotating the factor axes, each factor only focuses on a small number of highly correlated variables, resulting in more interpretable results. The remaining part of this paper is structured as follows. Section 2 describes the proposed methodology for thermographic data analysis. Then, the case study on a panel painting is presented in Section 3. Finally, Section 4 concludes the paper.

### 2. Exploratory factor analysis of thermographic data

Before applying EFA, it is necessary to transform the thermographic data structure from three-dimensional (3-D) to two-dimensional (2-D). As shown in **Fig. 1**, the thermal images collected in an AIRT experiment form a 3-D matrix whose size is  $n_t \times n_x \times n_y$ , where  $n_t$  is the number of frames and  $n_x \times n_y$  represents the total number of pixels contained in each image. After unfolding, the data can be reshaped to a 2-D matrix with a size of  $n_x n_y \times n_t$ . In the unfolded matrix shown in **Fig. 2**, there are totally  $n_x n_y$  variables each of which represents the temperature change at a specific pixel.



**Fig. 1** : 3-D thermographic data structure



**Fig. 2:** Unfolded thermographic data matrix

This matrix is processed by EFA to describe the variability among the variables in terms of a relatively small number of latent factors. The mathematical expression is as follows.

$$\mathbf{x}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda} \mathbf{f}_i + \boldsymbol{\varepsilon}_i \quad (1)$$

where  $\mathbf{x}_i \in R^{n_x n_y \times 1}$  is the  $i$ -th column in the unfolded thermographic data matrix,  $\boldsymbol{\mu} \in R^{n_x n_y \times 1}$  is the mean vector,  $\mathbf{f}_i \in R^{m \times 1}$  is the corresponding factor vector whose length  $m$  is usually smaller than that of  $\mathbf{x}_i$ ,  $\boldsymbol{\Lambda} \in R^{n_x n_y \times m}$  is the factor loading matrix, and  $\boldsymbol{\varepsilon}_i \in R^{n_x n_y \times 1}$  is the error term. It is clear that EFA models observed variables as linear combinations of several factors, plus error terms. There are two common approaches, i.e. maximum likelihood and principal axis factoring, that can be adopted to fit an EFA model [8]. For more details about the algorithms, please refer to the cited reference.

Different from the conventional PCA, EFA often uses factor rotation [9] to enhance the interpretability of the loadings. Usually, a rotation method aims to minimize the number of variables that load on a factor. In doing this, the loadings become sparser and the model interpretation is improved. The factor rotation methods can be generally divided into two groups, namely oblique rotation and orthogonal rotation. An important difference between these two groups is that the former create factors correlated with each other while the latter leads to orthogonal factors. Herein, orthogonal rotation is adopted to better separate the information from different sources. Specifically, varimax rotation [10], which searches for a rotation of the original factors  $\boldsymbol{\Lambda} \mathbf{R}$  such that the sum of the variances of the squared loadings on each factor is maximized, is used. Mathematically, the orthonormal rotation matrix  $\mathbf{R}$  can be calculated by solving the following optimization problem.

$$\operatorname{argmax}_{\mathbf{R}} \left( \sum_{j=1}^m \sum_{i=1}^p (\boldsymbol{\Lambda} \mathbf{R})_{ij}^4 - \frac{1}{p} \sum_{j=1}^m \left( \sum_{i=1}^p (\boldsymbol{\Lambda} \mathbf{R})_{ij}^2 \right)^2 \right) \quad (2)$$

where  $m$  is the number of factors,  $p$  is the number of variables, and  $(\boldsymbol{\Lambda} \mathbf{R})_{ij}$  denotes the element in the  $i$ -th row and  $j$ -th column in matrix  $\boldsymbol{\Lambda} \mathbf{R}$ . In the applications to thermographic data analysis,  $p = n_x n_y$ .

After varimax rotation, each rotated factor tends to have high loadings for a smaller number of variables and low loadings for the rest, leading to nearly sparse properties. Therefore, the loading structure is simplified and much easier to interpret. In thermographic data analysis, this means that each factor only highlights the information of a small number of pixels, which is beneficial to separating the defects from the backgrounds. By reshaping each column of the rotated loading matrix  $\boldsymbol{\Lambda} \mathbf{R}$  to an image sized  $n_x \times n_y$ , the shapes and locations of the defects can be visualized.

### 3. Case study

The sample to investigate in the case study is a panel painting named “Madonna” and sized  $15 \times 21 \times 2 \text{ cm}^3$ , where the panel is made of poplar wood. By inserting four Mylar sheets at different depths in the panel, the fabrication defects were generated as shown in **Fig. 3**. In the AIRT experiment, the energy source consisted of two 250W lamps (Siccatherm E27, OSRAM). The entire experiment lasted for 270 s, while the first 90 s belonged to the heating stage and the remaining time period was the cooling stage. The sampling interval was 1 s. The distance between each lamp and the painting was 48 cm, while the two lamps were distanced by 50 cm. An infrared camera (ThermaCAM S65HS, FLIR) was used to record the thermal images during both the heating and cooling stages, whose resolution is  $320 \times 240$  pixels. **Fig. 4** shows some of the captured thermal images. It is difficult to identify the defects because of the emissivity variations contained in these images. As observed in **Fig. 4**, the colors constitute by different pigments have different chemical compositions and hence different thermal conductivities, resulting in the outline of the portrait in each thermal image.

The entire thermographic dataset can be viewed as a 3-D matrix whose size is  $270 \times 240 \times 320$ . Before conducting further analysis, this matrix was unfolded to a 2-D matrix with size of  $270 \times 76800$ . In the unfolded matrix, each row corresponds to a thermal image and each column represents the temporal changes of a pixel value.

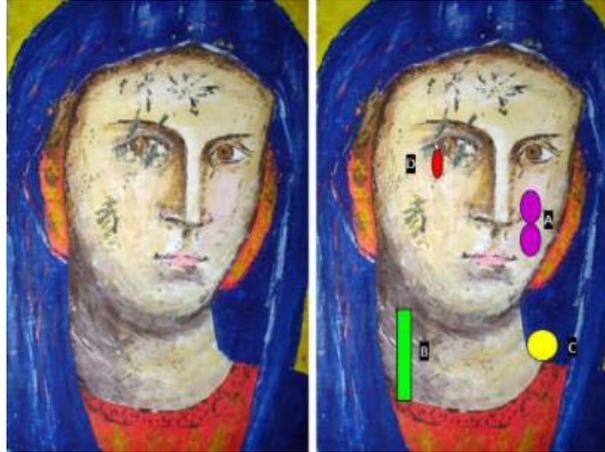


Fig .3: Panel painting and defect locations

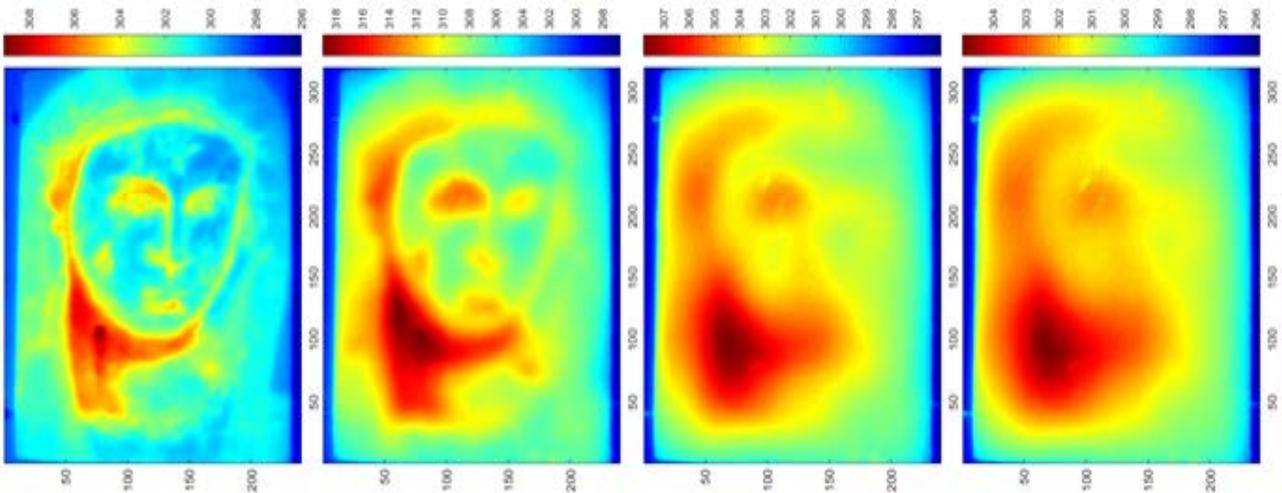


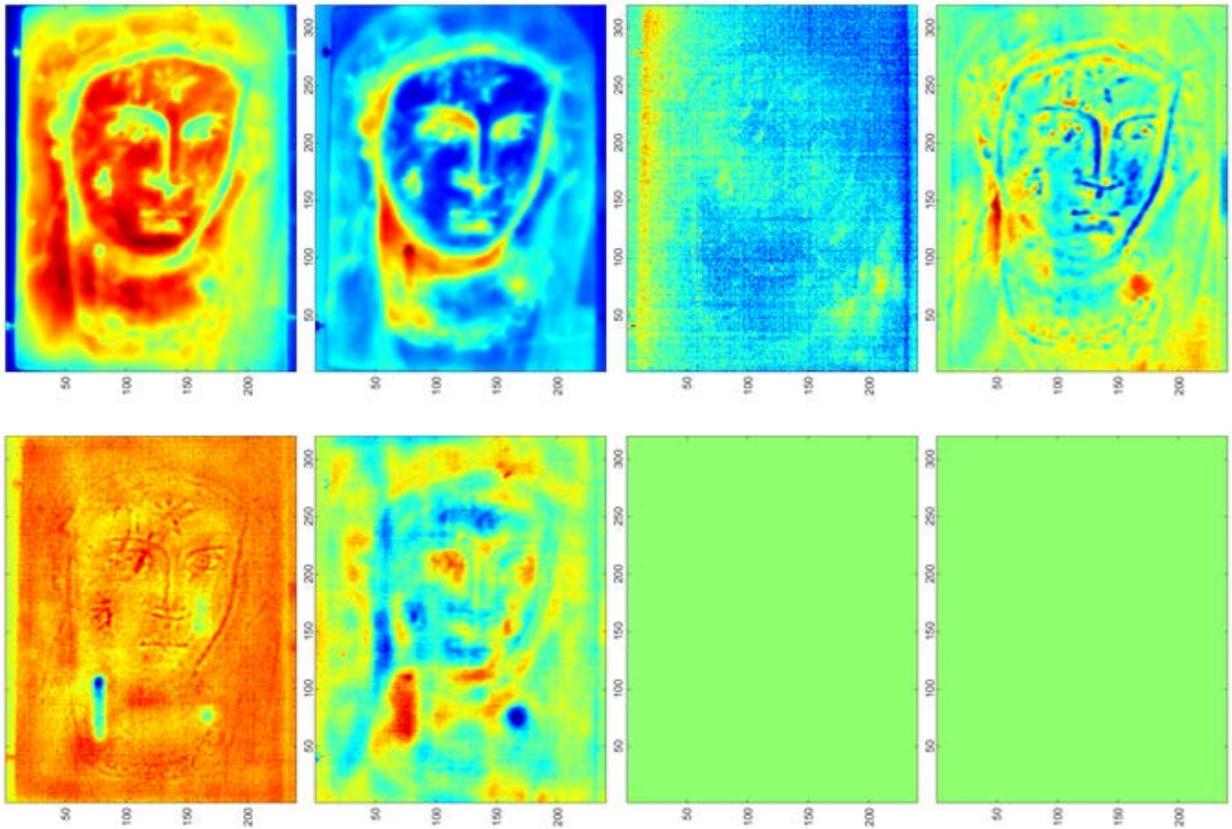
Fig. 4: Thermal images collected at the 20th, 100th, 200th, and 250th seconds from the beginning of the heating phase

For comparison, both FA and PCA were used to process the thermographic data, whose results are shown in **Fig. 5** and **Fig. 6**, respectively. It is clear that both methods reveal the existence of defects A, B, and C. Defect D is too deep to be discovered by AIRT. Therefore, neither FA nor PCA discover this defect. Thanks to factor rotation, FA better separates the defects from the backgrounds and concentrates all detected defects in one loading image, i.e. the fifth loading, facilitating the identification of the defects. Although defects A, B, and C can also be identified in the PCA results, the separation of the defect information and backgrounds is not good, while different defect features distribute in different loading images.

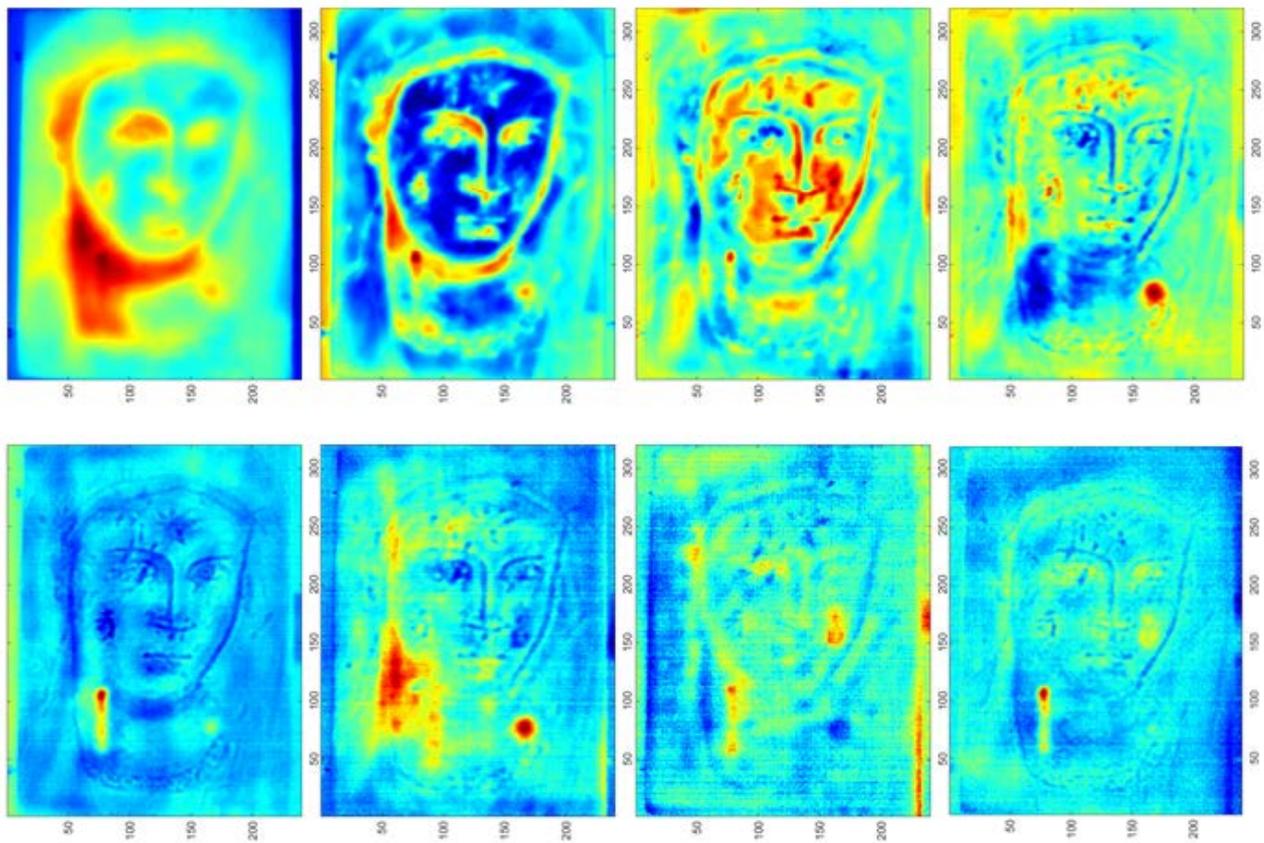
Signal-to-noise ratios (SNRs) [11] were calculated for both FA and PCA to compare these two methods quantitatively. The mathematical expression of SNR is as follows.

$$SNR = \frac{|\bar{S} - \bar{N}|}{var(N)} \quad (4)$$

where  $\bar{S}$  is the average signal value, i.e. the average value of the pixels in a defective region,  $\bar{N}$  is the average value of noise, i.e. the average value of the pixels in the intact region,  $var(N)$  is the standard deviation of the pixel values in the intact region. A larger SNR value indicates a better defect detection performance. Herein, the SNR values of FA were calculated based on the fifth loading image that contains most defect information, while the fifth, sixth, and seventh loading images of PCA are all investigated because PCA cannot highlight all defect features in a single loading image. The quantitative comparison results are listed in **Table 1** where the largest SNR of each defect is in bold. In general, FA significantly outperforms PCA, which has larger SNR values of defects A and B. The sixth loading image of PCA shows a larger SNR of defect C. However, the other two defects cannot be identified in this image clearly.



**Fig. 5:** FA results (the first 8 loadings)



**Fig. 6:** PCA results (the first 8 loadings)

**Table 1.** SNR values of FA and PCA

Methods	Defect A	Defect B	Defect C	All defects (A, B, and C)
Fifth FA loading	<b>2.6623</b>	<b>5.0384</b>	2.7195	<b>3.8847</b>
Fifth PCA loading	0.6321	4.1013	1.8046	2.5805
Sixth PCA loading	1.1309	1.8962	<b>4.0147</b>	1.1886
Seventh PCA loading	1.8292	2.5360	2.0604	1.4843

#### 4. Conclusions

In this work, we illustrate that FA is an effective method to extract thermographic data characteristics and reveal defect structures in the tested objects. By conducting factor rotation, FA produces much sparser loading vectors than PCA without subjective parameter selection. As a result, different types of information can be separated, facilitating defect identification. The case study of a panel painting illustrates the feasibility of the proposed method.

#### Acknowledgments

This work was supported in part by Ministry of Science and Technology, ROC, under Grant MOST 108-2221-E-007-068-MY3.

#### REFERENCES

- [1] Shepard S. M. Introduction to active thermography for non-destructive evaluation. *Anti-Corrosion Methods and Materials*; 44: pp. 236-239, 1997.
- [2] Wold S. Esbensen K., Geladi P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*; 2: pp. 37-52, 1987.
- [3] Rajic N. Principal component thermography for flaw contrast enhancement and flaw depth characterisation in composite structures. *Composite Structure*; 58: pp. 521-528, 2002.
- [4] Zou H., Hastie T., Tibshirani R. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*; 15: pp. 265-286, 2006.
- [1] Wu J.-Y., Sfarra S., Yao Y. Sparse principal component thermography for subsurface defect detection in composite products. *IEEE Transactions on Industrial Informatics*; 14: pp. 5594-5600, 2018.
- [6] Yousefi B., Sfarra S., Sarasini F., Castanedo C. I., Maldague X. P. V. Low-rank sparse principal component thermography (sparse-PCT): Comparative assessment on detection of subsurface defects. *Infrared Physics & Technology*; 98: pp. 278-284, 2019.
- [7] Ferguson E., Cox T. Exploratory factor analysis: a users' guide. *International Journal of Selection and Assessment*; 1: pp. 84-94, 1993.
- [8] de Winter J. C. F., Dodou D. Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics*; 39: pp. 695-710, 2012.
- [9] Browne M. W. An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*; 36: 111–150, 2001.
- [10] Kaiser H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*; 23: pp. 187-200, 1958.
- [11] Ibarra-Castanedo C., Piau J.-M., Guilbert S., Avdelidis N. P., Genest M., Bendada A., Maldague X. P. V. Comparative study of active thermography techniques for the nondestructive evaluation of honeycomb structures. *Research in Nondestructive Evaluation*; 20: pp. 1-31, 2009.