Damage Localization Using Pattern Recognition Techniques and Statistical Dissimilarity Analysis

HAROLD S. RODRIGUEZ, DIEGO A. TIBADUIZA BURGOS, FRANCESC POZO MONTERO and JERSSON X. LEÓN

ABSTRACT

The increasing expansion of standards and regulations aimed to guarantee the safe operation of different types of structures and materials ---which are extensively used in buildings, airplanes, transformers, machinery, etc.--- has driven the development and implementation of systems intended to supervise the state of the structure so critical failures may be predicted and avoided. Nevertheless many of those systems are expensive and require the structure to be inactive in order to evaluate its state, so it is frequently necessary to use external equipment and specialized technician teams, producing therefore a high interest in many sectors of industry, military and academic fields on supporting structural health monitoring (SHM) projects to develop new alternative systems. Among the proposals, the ones that includes integrated sensor networks systems are becoming increasingly popular due to their potential to perform the structure health evaluation even during these equipment are still in use, which results in a significant reduction in the economical and productive impact that other systems cause and the possibility to introduce new evaluation techniques that exploit the sensor and programming flexibility. This paper introduces a SHM system aimed to locate damages on metallic and composite material structures. For this purpose, a methodology is applied in this work that combines statistical analysis of dissimilarity with pattern recognition on the data recollected from the piezoelectric sensors network distributed on the structure surface. This methodology is tested in an aluminum plate instrumented with eight piezoelectric sensors and some masses added to the structure in order to emulate changes in the structure due to a damage. Results shows that it is possible to locate all damages.

Harold S. Rodríguez Lozada, Departamento de Ingeniería eléctrica y electrónica. Universidad Nacional de Colombia. Bogotá, Colombia. hsrodriguezl@unal.edu.co.

Diego Alexander Tibaduiza Burgos, Departamento de Ingeniería Eléctrica y Electrónica. Universidad Nacional de Colombia, Bogotá-Colombia. dtibaduizab@unal.edu.co.

Francesc Pozo Montero, Control, Modeling, Identification, and Applications (CoDAlab), Departament de Matemàtiques, Escola d'Enginyeria de Barcelona Est (EEBE), Universitat Politècnica de Catalunya (UPC), Campus Diagonal-Besòs (CDB), Eduard Maristany, 16, Barcelona 08019, Spain

Jersson Xavier León, Departamento de Ingeniería Mecánica y Mecatrónica Universidad Nacional de Colombia, Bogotá-Colombia. jxleonm@unal.edu.co

INTRODUCTION

The increasing interest for the development of structural health monitoring systems has been focused its efforts in the last decade on the use of integrated systems. These integrated systems use distributed sensor grids all along the structure so the monitoring does not require to stop the structure operation in order to generate a diagnosis. This kind of systems arose as an alternative to classic monitoring equipment, where monitoring task is expensive because this job could require several days to be completed according to the structure size, its complexity, the number of persons involved in the inspection and the kind and number of external systems used. One of the common inconvenient with classic inspection methods is associated to the stop of the structure operation which leads to considerable economic losses [1,2]. Among the most notable classic monitoring systems limitations, one can consider the necessity to keep the structure idle and isolated, in order to avoid disturbances that are introduced to the structure and to take measures while a group of technicians uses their sensor modules. The said limitation actually is a requirement to perform the analysis and complete the process of measurement, which is a disadvantage that increases the costs of the analysis by impeding the use of the elements under test. On the other hand, the implementation of a sensor grid on the surface or in-between layers of the structure is used to acquire data of its state, this gives the chance to get the data at any moment even if the structure is being used at that moment. Still, it is necessary to use a configuration that optimizes the number of sensors and its distribution on the grid, so it is possible to perform such analysis that allows it to detect, locate and evaluate the damage on the structure using data analysis and signal processing techniques. Although there are many systems that have integrated these kinds of methodologies, its development is still in process or its use have been focused on very specific applications, which is why there is still the necessity to investigate about its possible uses and the development of the related elements required for its operation [5]. This paper seeks to contribute to the solution of the problem described especially focusing on the damage localization task, this is performed by integrating a sensor network and algorithms of pattern recognition that allow the development of a methodology suitable for its use on structures made of metal or composite materials. This work is done in such a way that allows to specially develop the system for its use on structures made out of materials that are common

on aircraft structures, in this specific case this is tested on an aluminum plate instrumented with a piezoelectric sensor network which have some artificial damages in different positions. The organization of this work includes a theoretical background, a description of the damage localization methodology, the experimental setup, the results section and, finally, the conclusions.

THEORETICAL BACKGROUND

The necessity of reduction on the amount of data to be processed had been the object of many other projects that have implemented different kind of techniques, some of them have given a special treatment to signal acquisition and transformation in such a way it is more suitable for a complex analysis. In the pre-processing step is where data reduction is performed. This mean that data is represented in such a way that the information it contains is practically the same but its size or dimension is reduced. This is achieved even when the data under analysis has another representation and a different number of dimensions. This procedure can be divided into two different phases of the system, the first one is the pre-processing stage: the stage that prepares the data for it to have the form and characteristics that are needed. The second phase is the analysis of the data through techniques that transform the data until it reaches a suitable form for its treatment, and this includes shape, dimension number and format changes. Data analysis can be performed through the manipulation of the sources and nature of the data, so it is necessary to develop models to represent physical behaviors and analyze how a disturbance in the structure changes the input. Otherwise, it is possible to acquire a large amount of data and use it to represent different states of the structural health, where the analysis, in this case, would be done using pattern changes in the data and avoiding the complexity of a physical model. Once the hardware features of the system have been defined and the data processing techniques have been implemented into a whole algorithm there is still the necessity to find an index to measure the difference between different states, by this may be possible to associate the index value to the areas of a structure and perform the damage localization. This section is aimed to provide some basic definitions about pre-processing techniques and dissimilarity index used in the methodology used in this work.

1. Blind source separation

Blind source separation (BSS) is an algebraic method widely used on biology, medicine, and engineering to perform signal analysis on phenomena, especially when multiple sources are represented in one single signal or behavior under study. Its utility is given by the ability to get information from a complex signal with multiple sources mixed, so it can be represented in a much more suitable form to be analyzed through the components identified in it [6]. This task requires taking a dataset x that contains a group of measures taken from the phenome under observation, which is part of a bigger and more complex system so its analysis is difficult due the mixed signal contained in every sample [8]. Said difficulty is the main utility of BSS given that the complex signals dataset on the input is transformed into a components matrix, this has a smaller size and a set of simpler signals called components that represents the characteristic variations of the original dataset. The analyzed system is modeled in

matrix format seen bellow so it is represented the matrix x conformed for the samples taken from the system under test, which actually is the only variable we know. There also is the matrix s that represents every single signal involved in the mix that conforms the samples in x, so their relationship is given by a mixing matrix A [6], which will be the target of this procedure as in equation (1).

$$As = x \qquad (1)$$

The mixing matrix describes the mapping process that takes the sources matrix to become into the samples matrix, it is conformed for a group of unknown coefficients that allows the transition of the signals from a R^N domain conformed by the N signal vectors in s to the R^P domain given by the P time signals samples that conforms the matrix x. To obtain the matrix s is necessary to find the corresponding A^{-1} matrix that allows the transition from x to the sources matrix s, so it is necessary to approximate the coefficients of the inverse mixing matrix B through different methodologies that depend on the origin of the data and the expected characteristics of the outcome dataset [2]. Matrix B calculation is performed by a cross product between the matrices seen in the equation (2); these represent different aspects of the data according to the characteristics expected for the signal matrix. In this respect, the whitening matrix W is aimed to guarantee the orthogonality between the source signals so the vectors that conform matrix s are uncorrelated, on the other hand, the rotation matrix J is in charge of correcting the found sources position.

$$B = JxW \qquad (2)$$

The procedure followed to obtain the value of B matrix coefficients including the assumptions necessary to obtain that results affects the behavior of the matrix obtained. Therefore, this source matrix could be composed of signals that represent one primary aspect of the samples taken or even these may represent the components which resulted in the matrix [6]. Among the many techniques utilized for BSS, there is one characterized by its simplicity and the potential to resume a large and complex dataset into a few representative components, this is PCA (principal components analysis) which decomposes the sample dataset into some components that contain the principal information [5,9]. This technique calculates the covariance matrix so it can represent the variation of the sample matrix, from this each principal component will be given by the eigenvectors from the covariance matrix and its variance representation will be represented for the correspondent eigenvalues.

The process described above results in a feature vector that contains the obtained principal components, this is used to get the data projection of the considered datasets to the components domain [3,10]. Therefore, the result of a PCA algorithm is given by the equation (3), where X is the sample dataset and P^T is the transposed feature vector that allows the projection of the data to the components domain as a score matrix [4,5].

$$T = XP^T \qquad (3)$$

2. Dissimilarity Index

Dissimilarity is a quantification of how different two groups are, this is according to its behavior measured through different parameters and metrics, so the variation of a dataset is used to calculate a numeric value to appreciate how it is different group respect to a previously defined reference. This is useful when a pattern analysis is performed by considering that the main interest is to appreciate how a variation in the

data contributes to a particular output, hence the characterization of a system and its behavior given an input may be evaluated and predicted.

There are multiple parameters that contribute to the value of dissimilarity between two groups, still, we may consider two of them to resume the difference simply by taking the covariance and the distance of the group under evaluation taking to other as a reference. Said parameters effect is considered when calculating the Mahalanobis distance through the equation below, this considers the covariance matrix C and the distance between the predefined group references in a single expression, therefore, resumes variations in position and shape.

$$d_{ij}^2 = \left(K_i - K_j\right)^T C^{-1}\left(K_i - K_j\right) \tag{4}$$

Another method to consider is the Euclidean distance, this is very similar to the Mahalanobis distance [7], but only considers the separation distance between the data center using the equation bellow.

$$d_{ij}^2 = \left(K_i - K_j\right)^T\left(K_i - K_j\right) \tag{5}$$

DAMAGE LOCALIZATION METHODOLOGY

Figure 1 summarizes the steps of the damage localization methodology. Every single block describes a specific function in the system when it is processing the data, which takes different transformation from its acquisition to the final diagnosis of the damage location. In general, input data from the acquisition system are organized and pre-processed in the same way to data from healthy structure (train) and structure with damages (test). After that, PCA model is calculated to define the pattern and data from different damages of the structure is projected into the PCA model.
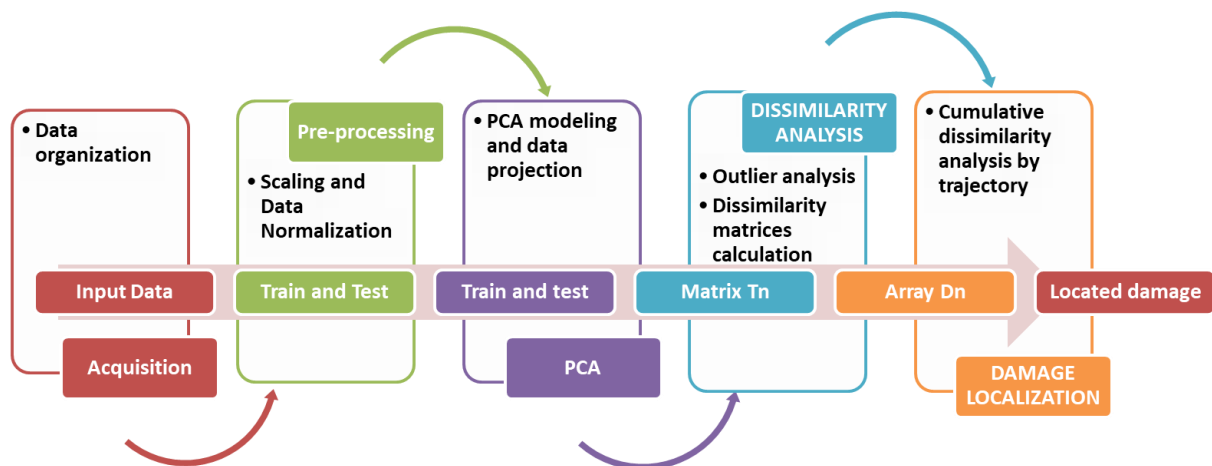

Figure 1. Block Diagram of the damage localization methodology

Dissimilarity is calculated as it was explained in the theoretical background and finally a cumulative dissimilarity analysis is obtained by the different trajectories between the PZTs in the sensor network.

EXPERIMENTAL SETUP AND RESULTS

To excite the structure, one PZT in the sensor network is used as actuator and a BURST signal is applied. Rest of the sensors collect the propagated signals trough a digital scope. This configuration is repeated until each PZT in the network works as actuator. This setup is used to acquire signals from the healthy structure (TRAIN MATRIX) and the structure with the damage in different positions (TEST MATRIX) as in Figure 6. Data organization is shown in Figure 2.

T1(1)	S2	S3	S4	...
T1(2)	S2	S3	S4	...
T1(3)	S2	S2	S4	...
T1(4)	.	.	.	.

T1(1:20)	S2	S3	S4	...
T1(21:40)	D1,2	D1,3	D1,4	...
T1(41:60)	D2,2	D2,3	D2,4	...
T1(...)	.	.	.	.

a.   b.

Figure 2. Data organization a) TRAIN MATRIX and b) TEST MATRIX

Once the two types of matrices are organized these are used into the pre-processing algorithm to extract its mean and scaling the samples, this aims to achieve a complete preparation for the data to be evaluated. Using BSS techniques it is possible to reduce the data size and facilitate the whole process, this methodology uses PCA to perform the features extraction so the variance in the adjusted dataset is represented through a predefined components number; said procedure leaves as a result a matrix for every piezoelectric in the system.

M1(1:20)	PS1	PS2	PS3	...
M1(21:40)	PD1,1	PD1,2	PD1,3	...
M1(41:60)	PD2,1	PD2,2	PD2,3	...
M1(...)	.	.	.	.

M2(1:20)	PS1	PS2	PS3	...
M2(21:40)	PD1,1	PD1,2	PD1,3	...
M2(41:60)	PD2,1	PD2,2	PD2,3	...
M2(...)	.	.	.	.

Figure 3. Projected data matrices for a given components number

Once the principal components have been extracted and the formation of projections for different structural states results as the obtained in Figure 4 are available to analyze, these represent the damages simulated on different location using a magnet adhered to the surface of the sheet and hence causing a variation of the mass, which is what a damage would normally cause.
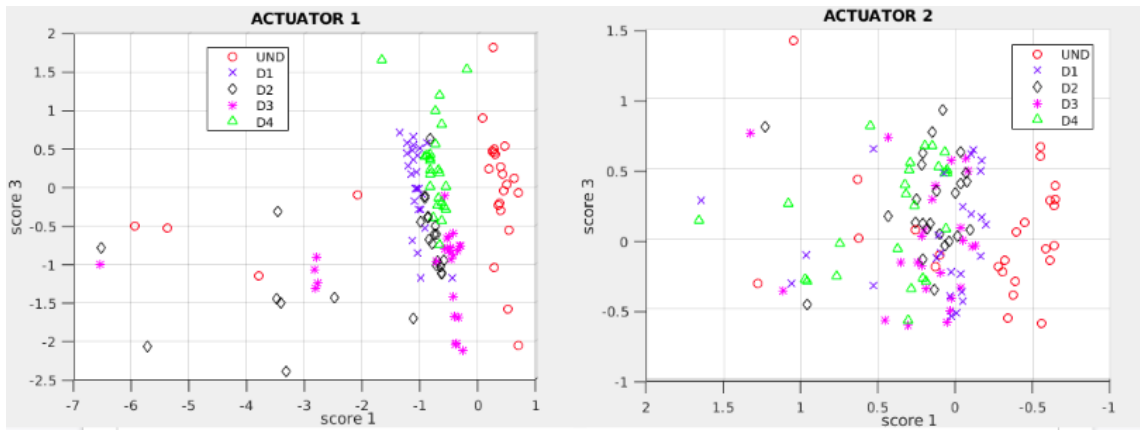
Figure 4. Projected data matrices for a given components number

These plots can be used to detect damages as many authors have demonstrated. In this case, two scores are only used to build the plots, however, damage localization cannot be approached by this simple method. In the same way increase the number of components is not the best solution in some cases as it is shown in Figure 5.
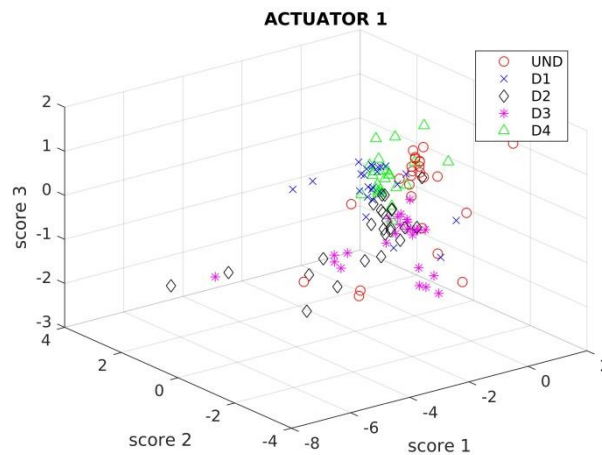

Figure 5. Projected data 3D plot representation for one of the piezoelectric

Given the multidimensional nature of the dataset it was chosen to apply a simple yet effective treatment to perform the calculation of dissimilarity, this consisted on getting the data group center using its mean in every component like coordinates to be used as reference points. The use of those references is given during the separation calculation, this is, when the dissimilarity value is obtained using the distance between groups and center points given by the dispersion of the data. After conducting many several trials with different types of multidimensional distance calculation techniques there were only two whose results were promising, those where Mahalanobis distance and Euclidean distance were used. Both of those techniques where theoretically suitable for the dissimilarity calculation in this application, still though the Euclidean distance probed to be the way to obtain the best results according to the expected variations on the data of every actuator under observation. When the behavior of the waves is considered this seems to have a constant different intensity value at the

moment they reach every sensor on the plate, this changes in a different way according to the position of the damage that is going to be considered. Hence the position of the damage will cause reflections that will lead to introducing disturbances with a considerable amplitude value to the closest sensor, then it is possible to consider after taking a considerable number of samples that the dissimilarity value will be bigger as the damage is closer to the area of the actuator that generated the dataset under evaluation. As a consequence of the analysis above, there can be defined some trajectories on the surface of the plate and the sum of the dissimilarities of the surrounding actuators, hence achieving to join the different dissimilarity values into some cumulative dissimilarities associated to particular trajectories on the plate. This allows to give the dissimilarity analysis a physical meaning and also to transform the measurement obtained from every actuator into a contribution to the adjacent trajectories; this is displayed in Figure 6 where is evaluated a damage defined as damage 2 and it is displayed in a plot with the contribution of every actuator according to the dissimilarity calculated for its dataset.

Finally, Figure 7 shows the results for one of the evaluated damages, in this case, again the damage 2 (located in trajectory 6) was under analysis. When the accumulated dissimilarity is calculated it shows how the bigger values correspond to the trajectories nearby the damage location achieving that way the system objective. Localization can be performed by triangulation methods.
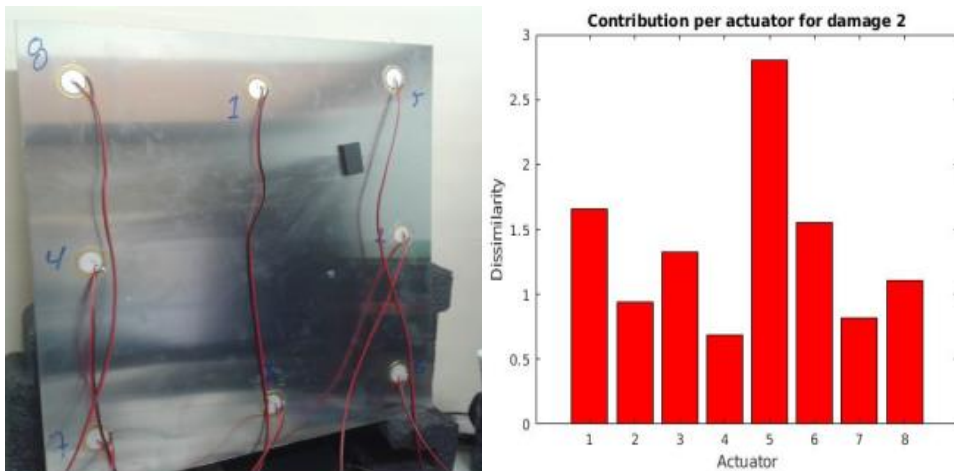

Figure 6. Contribution according to the actuator. a) Metal sheet with the simulated damage, b) Contribution per actuator for the simulated damage 2.
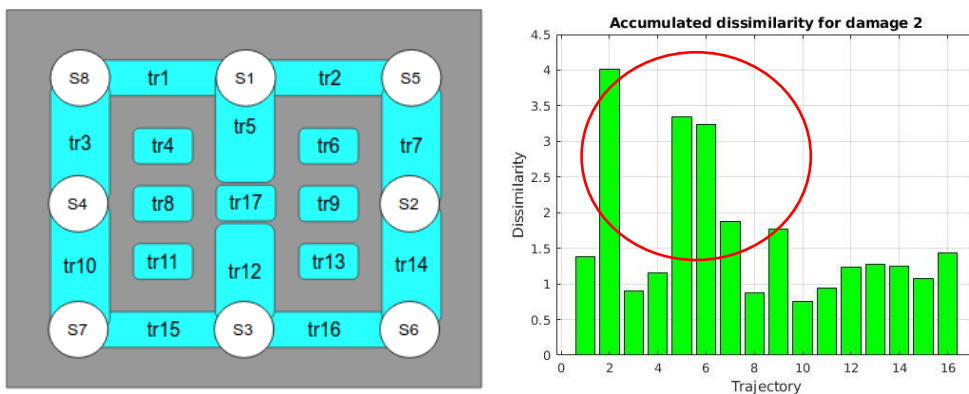

Figure 7. a) Pre defined trajectories on the sheet surface. b) Accumulated dissimilarity on every trajectory for damage 2

CONCLUSIONS

The BSS techniques can be used in a complex analysis like the ones necessary in SHM and adapt the obtained results to many other techniques to define more complex methodologies   as it was shown in this paper. The combination of BSS and dissimilarity analysis showed that it is possible to detect and locate abnormalities in the propagation of the waves between the PZTs in the sensor network. Although more results with different positions of the damage cannot be included in this paper by the constrains in length, similar results were obtained by all the damages in the different possible ways between all sensors.

ACKNOWLEDGMENT

REFERENCES

1. Vitola, J., Pozo, F., Tibaduiza, D. A., and Anaya, M. 2017. "A sensor data fusion system based on k-nearest neighbor pattern classification for structural health monitoring applications". *Sensors*, 17(2), 417.
2. Tibaduiza, D.A., Torres, M.A., Mujica, L.E., Rodellar, J. and Fritzen, C.P. 2012. "A study of two unsupervised data driven statistical methodologies for detecting and classifying damages in structural health monitoring", Mechanical Systems and Signal Processing, 41(1):467–484
3. Abdi, H. ; Williams, L.: Principal Component Analysis. En: WIREs Computational Statistics 2 (2010).
4. Mitsui, Y. ; Kitamura, D. ; Takamichi, S. ; Ono, N. ; Saruwatari, H.: Blind source separation based on independent low-rank matrix analysis with sparse regularization for time-series activity. En: IEEE International Conference on Acoustics, Speech and Signal Processing (2017).
5. Tibaduiza, D. ; Mujica, L. ; Anaya, M. ; Rodellar, J. ; Guemes, A. Principal Component Analysis vs. Independent Component Analysis for Damage Detection. 5th European Workshop on Structural Health Monitoring. 2012.
6. Vázquez, Rita, Romo, Avenilde, Romo-Vázquez, Rebeca, & Trigueros, María. (2016). La separación ciega de fuentes: un puente entre el álgebra lineal y el análisis de señales. Educación matemática, 28(2), 31-57.
7. R. De Maesschalck, D. Jouan-Rimbaud, D.L. Massart,The Mahalanobis distance, Chemometrics and Intelligent Laboratory Systems, Volume 50, Issue 1,2000,Pages 1-18.
8. A. Belouchrani, K. Abed-Meraim, J. -. Cardoso and E. Moulines, "A blind source separation technique using second-order statistics," in *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434-444, Feb. 1997.
9. Anaya, M., Tibaduiza, D. A., and Pozo, F. (2017). Detection and classification of structural changes using artificial immune systems and fuzzy clustering. International Journal of Bio-Inspired Computation, 9(1): 35-52.
10. Tibaduiza, D., Torres-Arredondo, M. Á., Vitola, J., Anaya, M., and Pozo, F. (2018). A damage classification approach for structural health monitoring using machine learning. *Complexity*, 2018.