Scientia Iranica A (2022) 29(6), 2789–2802

SCIENTIA IRANICA

Sharif University of Technology
Scientia Iranica
Transactions A: Civil Engineering
http://scientiairanica.sharif.edu

Identifying damage location under statistical pattern recognition by new feature extraction and feature analysis methods

Y.A. Feizi[a], M.K. Sharbatdar[b,*], R. Mahjoub[a], and M. Raftari[a]

a. *Department of Civil Engineering, Khorramabad Branch, Islamic Azad University, Khorramabad, Iran.*
b. *Faculty of Civil Engineering, Semnan University, Semnan, Iran.*

Received 18 May 2020; received in revised form 21 August 2021; accepted 6 December 2021

KEYWORDS
Structural;
Health monitoring;
Damage localization;
Pattern recognition;
Convolution;
Component analysis;
Statistical distance.

Abstract. The main objective of this article is to identify the location of damage by a new feature extraction technique and to propose some efficient feature analysis tools as statistical distance measures. The proposed algorithm of feature extraction relies on a combination of the well-known Principal Component Analysis (PCA) and convolution strategy. After extracting the features from raw vibration signals of undamaged and damaged conditions, those are applied to the proposed feature analysis approaches called coefficient of variation, Fisher criterion, Fano factor, and relative reliability index, all of which are formulated by using statistical moments of the features extracted from the PCA-convolution algorithm. To localize damage, the sensor location with the distance value exceeding a certain threshold limit is identified as the damaged area. The main innovations of this research are to present a new hybrid technique of feature extraction suitable for Structural Health Monitoring (SHM) applications and four effective statistical measures for feature analysis and damage identification. The performance and reliability of the proposed methods are verified by a four-story shear-building model and a benchmark concrete beam. Results demonstrate that the approaches presented here can substantially identify the location of damage using the features extracted from the proposed PCA-convolution algorithm.

1. Introduction

Evaluating the health of civil engineering structures has now received significant attention due to their importance in transportation systems, social life, economic, etc. Most of them are needed to be monitored and maintained in an effort to prevent catastrophic events caused by damage occurrence, natural disasters, aging, and material deterioration. For these reasons, Structural Health Monitoring (SHM) has emerged to assist civil engineers in assessing the health and safety of important civil structures and detecting any possible structural damage [1,2]. On this basis, it is necessary to deploy civil structures by various kinds of sensors [3], measure raw vibration signals, construct Finite Element (FE) or numerical models, update the constructed numerical models, and apply model-based or data-based methods for SHM [4]. Damage can be defined as intentional or unintentional changes in geometry, boundary conditions, and material properties leading to adverse alterations in the behavior and responses of a structure [2]. These changes appear as cracks in concrete elements and broken welds in the

*. *Corresponding author.*
E-mail addresses: enyousefali@yahoo.com (Y.A. Feizi); msharbatdar@semnan.ac.ir (M.K. Sharbatdar); r_mahjoob@yahoo.com (R. Mahjoub); Raftari_m@yahoo.com (M. Raftari)

doi: 10.24200/sci.2021.55991.4504

steel connections, failure, and fatigue. All of them may cause undesirable stresses, inappropriate vibrations, failure, and collapse in the structure. To avoid such adverse events, the process of SHM is categorized into four main steps: (i) Early damage detection, (ii) damage localization, (iii) damage type recognition, and (iv) damage severity estimation.

The first step intends to initially alarms the occurrence of damage or safe condition. When the structure suffers from damage, one attempts to identify the damaged area of the structure. Once the location of damage is identified, the type of damage (cracks, failure, etc.) can be recognized. Finally, the severity of damage is estimated in order to either repair the damaged area or replace it. An important note is that as the mentioned steps increase, the complexity and difficulty of SHM methods increase, as well.

The model-based methods need to have FE models of civil structures and their structural properties, i.e., mass, damping, and stiffness. Due to discrepancies between the numerical and real models of structures, model updating [5–7] is mandatory for the model-based strategy. The central idea of SHM via this strategy lies in the fact that the numerical (updated) and real models of the structure are considered as undamaged and current states. Hence, it is attempted to use information of both models to define a damage equation as an inverse problem and solve it via various mathematical and optimization techniques [8–11]. By contrast, the data-based methods only utilize measured vibration signals without any FE modeling or updating procedures, and any data transformation from raw time domain into frequency or modal domains. It needs to clarify that although these methods are highly suitable for early damage detection, damage localization, and sometimes damage type recognition, their main drawbacks are related to the damage severity estimation. For these steps, the model-based methods can play important roles in accurately estimating the severity of damage [2]. The other important note regarding the data-based methods for the first three steps of SHM is that the procedures of early damage detection and localization are further prevalent and there are a few researches on recognizing the type of damage.

Most of the data-based techniques used in early damage detection and damage localization are implemented by statistical pattern recognition paradigm under four main steps: (i) operational evaluation, (ii) sensing and data acquisition, (iii) feature extraction, and (iv) feature analysis [2,12]. Feature extraction aims to extract meaningful information from the measured vibration data that should be correlated with damage, known as a Damage-Sensitive Feature (DSF) [13]. Time series analysis [14–16], time-frequency signal analysis [17–19], and Principal Component Analysis (PCA) [20–23] are widely used and effective methods for feature extraction. Feature analysis is a decision-making procedure that utilizes the DSFs of undamaged and damaged conditions extracted from vibration signals in order to analyze them for early damage detection and damage localization. This process can be performed in two strategies: (i) a direct comparison of the DSFs and (ii) training a machine-learning model. In the first strategy, the DSFs of the undamaged and damaged states are directly compared via statistical metrics without learning any model. For this strategy, statistical distance measures are the most common approaches. Depending the type and size of the DSFs, there are some efficient distances for both early damage detection and damage localization such as Mahalanobis distance [24–26], Kullback-Leibler divergence [27–29], correlation distance measures [19,30], classical and robust multidimensional scaling [31,32], spectral distances [33,34], etc.

The second strategy relies on training a machine-learning model via some DSFs of the undamaged state, serving as training data. Once the machine-learning model has been trained, the remaining DSFs of the undamaged state (validation data) as well as all DSFs of the damaged condition, all of which are considered to generate test data, are fed into the trained model to make a decision in terms of early damage detection and damage localization [35]. In general, any machine-learning model can be developed by the concepts of supervised learning, requiring fully labeled data, i.e., the DSFs of both the undamaged and damaged states for training the model [36,37], or unsupervised learning, which can be established by only partially labeled data (i.e., the DSFs of the only undamaged state are necessary to learn the model of interest and the labels of the DSFs of the damaged condition are unknown) [38]. Although both strategies are suitable for early damage detection and damage identification, the utilization of the direct statistical distance is benefited by simplicity and computational efficiency compared to the machine learning-based strategy, especially when the extracted features exhibit proper sensitivity to damage.

On the other hand, most of the machine-learning methods undertake the process of early damage detection in order to understand whether damage is present throughout the whole structure (the first level of SHM), particularly in a long-term manner. Distance-based novelty detection [38–40], clustering [41–43], and artificial neural networks [44,45] are the widely used machine learning methods based on the concept of unsupervised learning for early damage detection. In contrast, the use of direct statistical measures is often suitable for damage identification, particularly in a short-term manner. However, the preliminary step of this process is to apply effective and efficient DSFs. The effective features mean that those should be sensitive to damage and proper for damage identification. Moreover, the efficient features make sense to extract information from vibration signals that do not lead to a time-consuming process or rigorous parameter estimation. Another reason for having an appropriate result of damage identification is to use statistical distance measures with a high rate of detection along with their simplicity and computational efficiency. Therefore, the main objective of this article is to propose effective methods for identifying damage on the basis of statistical pattern recognition. In this regard, a new approach as a combination of the well-known PCA and convolution technique, called PCA-convolution, is proposed to extract the DSFs. In this algorithm, one attempts to project the matrix of raw vibration signals onto principal components and to utilize them in the algorithm of convolution instead of raw vibration data. Hence, the convolution of a pair of principal components regarding the undamaged and damaged conditions is computed as a new DSF. Additionally, this article presents some effective statistical distance measures, called Coefficient of Variation (CV), Fisher criterion, Fano factor, and relative reliability index for damage localization. Accordingly, the sensor location concerned with the largest distance quantity is identified as the damage area. To verify the accuracy and capability of the proposed methods, two numerical models including a four-story shear building and a benchmark concrete beam are considered and studied. Results show that the proposed distance measures with the aid of the DSFs extracted from the proposed PCA-convolution algorithm are accurately able to identify the damage location. Furthermore, it is seen that this algorithm can extract reliable and sensitive features from raw vibration signals through a simple but effective algorithm.

2. Mixture feature extraction

2.1. Principal Component Analysis (PCA)

PCA is a multivariate statistical process that is used to convert a set of correlated variables into a set of values of linearly uncorrelated variables called principal components. Mathematically, PCA is defined as an orthogonal linear transformation that transforms data into a new coordinate system [20,22]. Assume that $\mathbf{X} \in \Re^{n\times m}$ is an original data matrix containing information from m variables (sensors) and n measured vibration data points. Before applying PCA, it is necessary to carry out a standardization process on the data matrix to remove the differences between the ranges of variables. Once the variables are standardized, the matrix of covariance related to the vibration measurements is defined as follows:

$$\mathbf{C}_X = \frac{1}{m-1}\mathbf{X}^T\mathbf{X}, \tag{1}$$

$$\mathbf{C}_X\tilde{\mathbf{P}} = \tilde{\mathbf{P}}\Lambda. \tag{2}$$

In these equations, $\mathbf{C}_X \in \Re^{m\times m}$ is a square symmetric that represents the matrix of covariance of the original matrix $\mathbf{X}$. The covariance matrix measures the linear relationship degree within the original data set among all possible pairs of variables. Meanwhile, the eigenvectors of $\mathbf{C}_X$ are the columns of $\widetilde{\mathbf{P}}$ and the eigenvalues are the diagonal terms of $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$. Of note, the eigenvector with the highest eigenvalue takes into account as the first principal component of the data set; therefore, the eigenvectors corresponding to the columns of matrix $\widetilde{\mathbf{P}}$ are arranged on the basis of the eigenvalues in descending order. In such a way, a new matrix $\mathbf{P}$ (i.e., $\widetilde{\mathbf{P}}$ sorted and reduced) can be established as the PCA model. In fact, $\mathbf{P} \in \Re^{m\times k}$ is a linear transformation matrix, which is used to convert the correlated data matrix $\mathbf{X}$ into an uncorrelated matrix $\mathbf{T} \in \Re^{n\times k}$ in the following form:

$$\mathbf{T} = \mathbf{X}\mathbf{P}, \tag{3}$$

with a view to obtain the uncorrelated matrix, it is important to choose a matrix of transformation named as $\mathbf{P}$ such that the covariance of the new data matrix $\mathbf{T}$ is diagonal, that is:

$$\mathbf{C}_T = \frac{1}{n-1}\mathbf{T}^T\mathbf{T}. \tag{4}$$

Substituting Eq. (3) into Eq. (4), one can write:

$$\mathbf{C}_T = \frac{1}{n-1}\mathbf{P}^T\mathbf{X}^T\mathbf{X}\mathbf{P} = \mathbf{P}^T\mathbf{C}_X\mathbf{P}. \tag{5}$$

The variance of each column vector in matrix $\mathbf{T}$ can be expressed in the form:

$$\sigma^2_{\mathbf{t}_j} = \frac{1}{n-1}\mathbf{t}_j^T\mathbf{t}_j = \frac{1}{n-1}\left(\mathbf{X}\mathbf{p}_j\right)^T\left(\mathbf{X}\mathbf{p}_j\right) = \mathbf{p}_j^T\mathbf{C}_X\mathbf{p}_j = \lambda_j, \tag{6}$$

where $\mathbf{t}_j$ and $\mathbf{p}_j$ are the jth vectors of matrices $\mathbf{T}$ and $\mathbf{P}$, respectively. Furthermore, the covariance is:

$$\sigma^2_{\mathbf{t}_j,\mathbf{t}_k} = \frac{1}{n-1}\mathbf{t}_j^T\mathbf{t}_k = \frac{1}{n-1}\left(\mathbf{X}\mathbf{p}_j\right)^T\left(\mathbf{X}\mathbf{p}_k\right) = \mathbf{p}_j^T\mathbf{C}_X\mathbf{p}_k = \lambda_j\mathbf{p}_j^T\mathbf{p}_k = 0. \tag{7}$$

As a result, each vector of the transformed data matrix $\mathbf{T}$ is uncorrelated and its variances is given by the covariance matrix eigenvalues $\mathbf{C}_X$ of the original matrix. In the full dimension case, the transformation process is invertible since $\mathbf{P}\mathbf{P}^T = \mathbf{I}$, where $\mathbf{I}$ is the unity matrix; thus, the original data matrix can be recovered as $\mathbf{X} = \mathbf{T}\mathbf{P}^T$. By considering $\mathbf{T}$, it is not possible to recover the original matrix in a complete manner; however, this matrix can be projected back onto the original m-dimensional domain and another matrix can be obtained in the following form:

$$\overline{\mathbf{X}} = \mathbf{T}\mathbf{P}^T, \tag{8}$$

in which,

$$\overline{\mathbf{X}} = \mathbf{X} - \mathbf{E}, \tag{9}$$

$$\mathbf{E} = \mathbf{X}\left(\mathbf{I} - \mathbf{P}\mathbf{P}^T\right). \tag{10}$$

In this equation, $\overline{\mathbf{X}}$ denotes the projection of the matrix $\mathbf{X}$ onto the selected k principal components and $\mathbf{E}$ is the projection onto the residual left components.

2.2. Convolution

In signal processing, the convolution of two signals, $\mathbf{u}$ and $\mathbf{v}$, measures their similarity under the points as $\mathbf{u}$ slides across $\mathbf{v}$ [46]. From a mathematical viewpoint, the convolution is a mathematical operation on two functions, producing a third function. In the time-domain signals, the convolution of two signals involves integration for the continuous signals and summing for the discrete signals, where one of them is shifted [47]. On this basis, the general form of convolution of two signals, $\mathbf{u} \in \Re^r$ and $\mathbf{v} \in \Re^s$, is given by:

$$\mathbf{c} = \sum_j \mathbf{u}(j) \cdot \mathbf{v}(q - j + 1), \tag{11}$$

where $j = 1, 2, \ldots, r; r$ and s are the sizes of the signal vectors $\mathbf{u}$ and $\mathbf{v}$, respectively. In this equation, $\mathbf{c}$ is the vector of convolution with the length of q, where $q = r+s-1$. Suppose that $\mathbf{u}$ and $\mathbf{v}$ are two vectors (e.g., two signals) from two different conditions. According to the convolution theory, if there is no difference between the two conditions, the convolution of these vectors indicates a full similarity between them; otherwise, a clear dissimilarity can be observed from the convolution vector $\mathbf{c}$. Even though one can directly apply the convolution method to detect and/or locate damage by finding the dissimilarity between the two vibration signals in the two different structural states, the success significantly depends on the quality of the vibration signals. This means that the presence of any irrelevant information in the signals to damage, e.g., noise, signal variability caused by environmental changes, etc., causes erroneous results of damage detection and localization. In such cases, the probability of the occurrence of false alarm and/or false detection errors increases seriously. Accordingly, the direct use of the convolution algorithm on the raw vibration signal is not sufficiently applicable to either feature extraction or feature analysis. For these reasons, this article proposes the capability of this algorithm (i.e., as a non-parametric approach) with the aid of the PCA (i.e., a parametric approach) to develop a more effective method for feature extraction.

2.3. PCA-convolution algorithm

Assume that $\mathbf{X}$ and $\mathbf{Y} \in \Re^{n\times m}$ are the vibration data matrices in the undamaged and damaged conditions. Based on the fundamental principle of the PCA technique, one can transform these matrices into two uncorrelated data matrices, $\mathbf{T}_x$ and $\mathbf{T}_y$, by using the linear transformation matrices $\mathbf{P}_x$ and $\mathbf{P}_y$ as follows:

$$\mathbf{T}_x = \mathbf{X}\mathbf{P}_x, \tag{12}$$

$$\mathbf{T}_y = \mathbf{Y}\mathbf{P}_y. \tag{13}$$

In the following, two new data matrices in the original coordinate can be obtained as:

$$\overline{\mathbf{X}} = \mathbf{T}_x\mathbf{P}_x^T, \tag{14}$$

$$\overline{\mathbf{Y}} = \mathbf{T}_y\mathbf{P}_y^T, \tag{15}$$

where $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$ are obtained by projecting the original data matrices $\mathbf{X}$ and $\mathbf{Y}$ onto the principal components. Now, the matrices $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$ are applied to compute the convolution between their column vectors. In order to locate damage via the proposed statistical measures, one needs to calculate a convolution vector between the undamaged state and itself, $\mathbf{c}_x$, as well as another convolution vector between the undamaged and damaged conditions, $\mathbf{c}_y$. These vectors are formulated in the following forms:

$$\mathbf{c}_x = \sum_j \overline{\mathbf{x}}_i(j) \cdot \overline{\mathbf{x}}_i(q - j + 1), \tag{16}$$

$$\mathbf{c}_y = \sum_j \overline{\mathbf{x}}_i(j) \cdot \overline{\mathbf{y}}_i(q - j + 1), \tag{17}$$

where $i = 1, 2, \ldots, m$; $\overline{\mathbf{x}}_i$ and $\overline{\mathbf{y}}_i$ are the i^{th} vector (sensor) of $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$, respectively. Note that the vectors $\mathbf{c}_x$ and $\mathbf{c}_y$ serve here as the new DSFs regarding the undamaged-only and undamaged and damaged conditions. Although these are the convolution vectors, there are some advantages that make the proposed feature extraction method more suitable for damage identification. First, noise in measured vibration signals is filtered out by the PCA method. Therefore, one can handle the drawback of directly using the convolution technique for the raw data. Second, as the convolution technique indicates the overlap between two signals (i.e., vectors), the convolution vectors obtained from $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$ can increase the rate of detectability and localizability.

3. Statistical distance measures

Upon extracting the DSFs from the proposed mixture approach, those should be applied to some distance metrics, i.e., CV, Fisher criterion, Fano factor, and relative reliability index for damage localization. In the following, there measures are briefly described and formulized:

- *Coefficient of variation*
  In statistics and probability theory, the CV, which indicates the ratio of the standard deviation to the mean, is a statistical measure of the dispersion of a probability distribution or a random variable around the mean. In this article, this criterion is employed as a distance measure to identify the damage location based on the DSFs obtained from the proposed PCA-convolution algorithm. Given the convolution vector $\mathbf{c}_y$, the CV is expressed as follows:

$$CV_i = \left(\frac{\sigma_y}{\mu_y}\right)_i, \tag{18}$$

  where $i = 1, 2, \ldots, m; \sigma_y$ and μ_y denote the standard deviation and mean values of $\mathbf{c}_y$, respectively.

- *Fisher criterion*
  The Fisher criterion or Fisher's linear discriminant is a classification method that projects high-dimensional data onto a line. This projection maximizes the dissimilarity between the means of the two classes while minimizing the variance within each class. Considering $\mathbf{c}_y$ and $\mathbf{c}_x$ as the DSFs of the training and testing data sets, Fisher criterion (J) for identifying damage location is formulated as:

$$J_i = \left(\frac{(\mu_y - \mu_x)^2}{\sigma_y^2 + \sigma_x^2}\right)_i, \tag{19}$$

  where σ_x and μ_x are the standard deviation and mean values of the convolution vector $\mathbf{c}_x$, respectively.

- *Fano factor*
  In statistics, the Fano factor is a value of the dispersion of a probability distribution or a random variable at a specific time window. This measure highly resembles the CV except using the variance of data samples instead of their standard deviation considered in the CV. Hence, it is possible to extend the general formulation of the Fano factor based on the convolution vectors for identifying the damage location. Considering the vectors $\mathbf{c}_y$ and $\mathbf{c}_x$, a developed Fano factor for these sets can simply be written using their variance and mean values. To achieve meaningful results regarding the damage localization problem, a direct difference between the Fano factors of the vectors $\mathbf{c}_y$ and $\mathbf{c}_x$ is computed as $\left(\frac{\sigma_x^2}{\mu_x} - \frac{\sigma_y^2}{\mu_y}\right)$ which can be rewritten as follows:

$$F_i = \left(\frac{\sigma_x^2\mu_y - \mu_x\sigma_y^2}{\mu_x\mu_y}\right)_i. \tag{20}$$

- *Relative reliability index.*
  The reliability index is a useful indicator to compute the failure probability in the structural reliability analysis. This measure is based on the ratio of the mean value to the standard deviation, in which case one can understand that the reliability index is the inverse of the CV. However, this index cannot directly be applied to the problem of damage localization due to its reverse situation with respect to the CV, which has been proposed to locate damage. To deal with the limitation of applying this index to the problem of interest, a relative error between the reliability index between the vectors $\mathbf{c}_y$ and $\mathbf{c}_x$ is computed and designated as a new statistical measure R as follows:

$$R_i = \left|\left(\frac{\mu_x\sigma_y - \mu_y\sigma_x}{\sigma_y\mu_x}\right)_i\right|. \tag{21}$$

Regarding this equation, since it is possible to determine the negative value of R, one should utilize the absolute operator. An important property of the proposed statistical measures is that each of them gives a positive scalar value at each sensor location. Having considered m sensors optimally installed on the structure, four m-dimensional vectors of the statistical measures can be derived. In each of these vectors, the sensor location with the largest amount of that statistical measure is identified as the location of damage.

4. Numerical examples

4.1. A four-story shear building frame

In order to demonstrate the accuracy and ability of the proposed methods, a simple four-story shear-building model is constructed, as depicted in Figure 1. It is a discrete dynamic system with four Degrees Of Freedoms (DOFs) so that each floor includes one DOF in the horizontal coordinate. Suppose that four accelerometers (S1–S4) are mounted on the model to measure the acceleration time history at each DOF. The structural characteristics of the model including mass and stiffness are represented in Table 1. The classical damping is an appropriate idealization. Furthermore, Rayleigh damping model is utilized to construct the damage matrix using 5% damping ratio for all modes. The state-space method is employed to implement dynamic time-domain analysis and measure the acceleration time histories from the simulated sensors. In order to excite the model, four different Gaussian white noise signals are applied to the points across the sensors for simulating ambient vibration. As a sample, Figure 2 shows the excitation and acceleration response signals at the location of Sensor 4 in the undamaged condition.

To simulate damage, it is assumed that an additional concentrated mass is inserted in the third story.

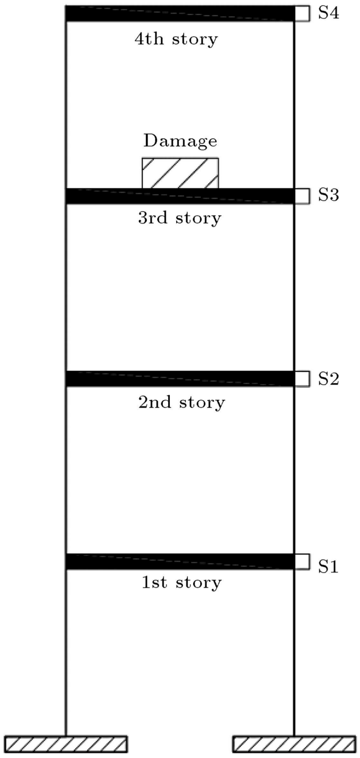

Figure 1. The four-story shear-building model.

Table 1. The structural properties of the shear-building model.

Story no.	Mass (kg)	Stiffness (kN/m^2)
1	4000	1600
2	3000	1400
3	2000	1200
4	1000	1000

Several mass increasing factors including 10, 30, and 50% are allocated to the mass of the third story to simulate the damage pattern. In the same manner, the structural dynamic analysis in the time domain is applied to the damaged cases to measure their acceleration time histories. Based on the statistical pattern recognition paradigm, the measured vibration responses in the undamaged condition (without adding the mass) generate the data matrix $\mathbf{X}$ and the measurements in the three damaged conditions produce the

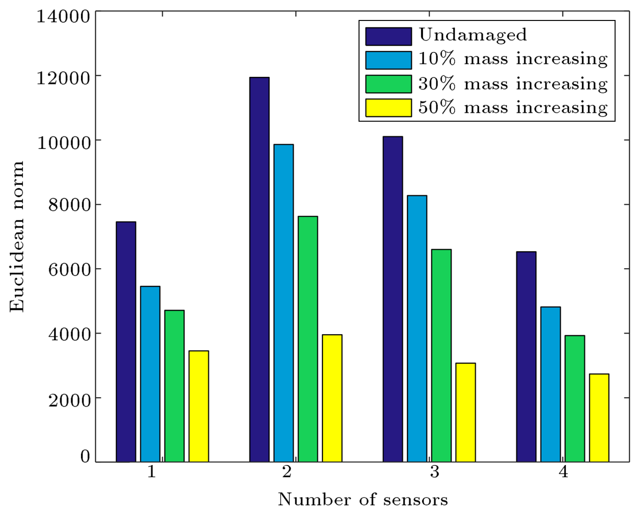

Figure 3. Euclidean norms of the PCA convolution vectors of the training and testing data sets.

data matrix $\mathbf{Y}$. After simulating and measuring the acceleration time-domain responses, the PCA method is applied to convert the original data matrices into the new spaces, $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$. The proposed feature extraction technique is employed to compute the PCA-convolution vectors $\mathbf{c}_x$ and $\mathbf{c}_y$. Figure 3 illustrates the Euclidean norm of these vectors in all sensors of the undamaged and damaged conditions.

As shown in Figure 3, the Euclidean norms of the PCA-convolution vectors are reduced upon increasing the level of damage from the undamaged condition to the highest level of damage (50% mass increasing). It is obvious that the undamaged state has the largest norm, whereas the highest damage scenario gives the smallest norm value. This observation confirms that the proposed feature extraction technique provides the reliable and accurate DSFs due to the sensitivity of the PCA-convolution vectors to damage. Despite this advantage, it cannot properly detect the damage or identify the location of damage. Thus, it is a necessity to apply the proposed distance measures for locating damage. For further investigation, Figure 4 shows the PCA-convolution vectors $\mathbf{c}_x$ and $\mathbf{c}_y$ in Sensor 3 (the location of damage).

As can be observed from Figure 4, the values of PCA-convolution vectors are reduced by increasing the level of damage. By comparing the PCA-convolution

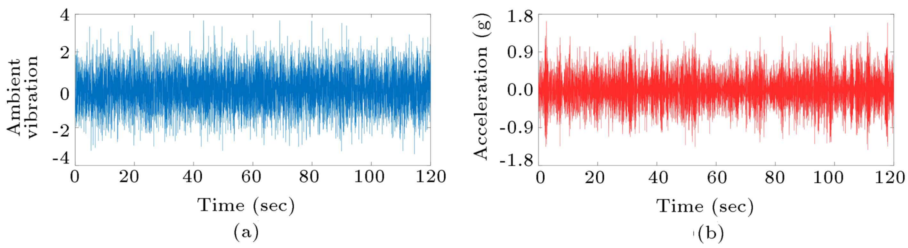

Figure 2. The simulation process at Sensor 4: (a) Ambient vibration and (b) acceleration response.

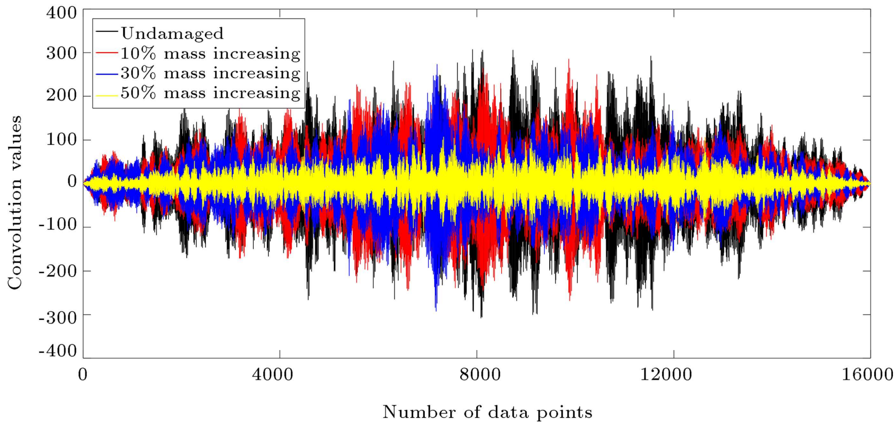

Figure 4. The PCA convolutions of training and testing sets at Sensor 3 (the location of damage).

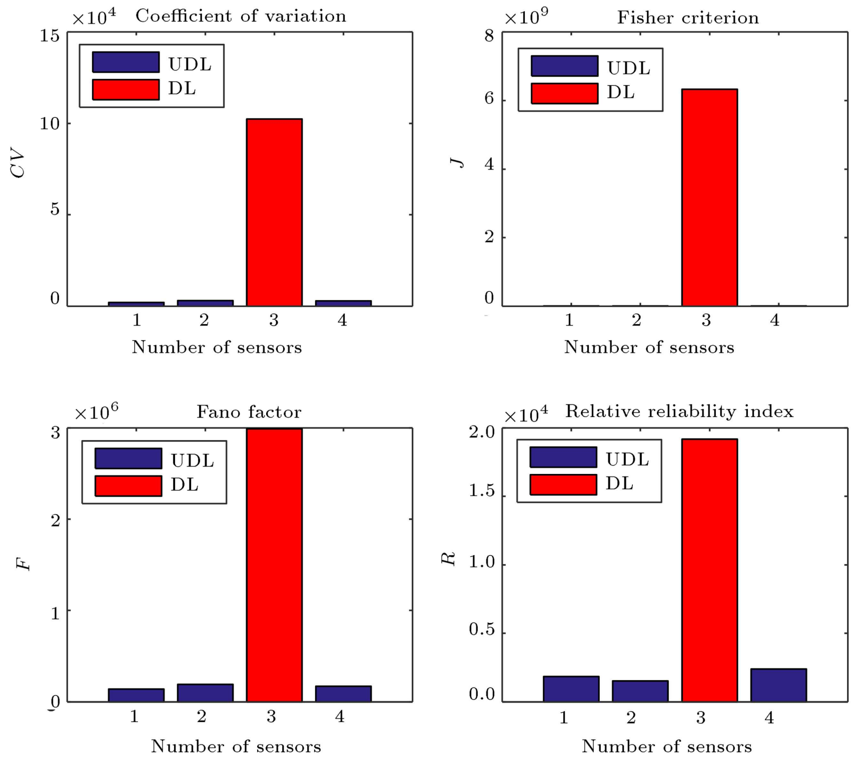

Figure 5. The damage localization in the first damaged case: 10% mass increase.

vectors in the damage cases, it can be suggested that damage leads to a reduction in their values. In such circumstances, the highest level of damage (50% mass increasing) shows the smallest values of the PCA-convolution. To identify the damage location, the statistical moments of the vectors $\mathbf{c}_x$ and $\mathbf{c}_y$ are calculated and used in the proposed statistical distance measures. In this regard, Figures 5–7 display the results of damage localization in Cases 1–3, respectively. In these figures, "UDL" refers to the undamaged location of the model, whereas "DL" means the damaged location. All of the results obtained from these figures demonstrate that Sensor 3 is the location of damage in the shear-building model since the values of the four statistical distance measures in this sensor are larger than the other ones. These observations not only confirm that the PCA-convolution vectors are sensitive to damage, but also prove that all of the statistical measures can successfully identify the location of damage.

4.2. A numerical benchmark concrete beam

To provide further evidence for verifying the proposed methods, a numerical benchmark model [48] is applied. This model is a realistic simulation of the concrete beam, as can be seen in Figure 8. The dimension of the beam features length 5 m, height 0.5 m, and width 0.01 m. It was constructed with four-node linear two-dimensional elements with reduced integration.

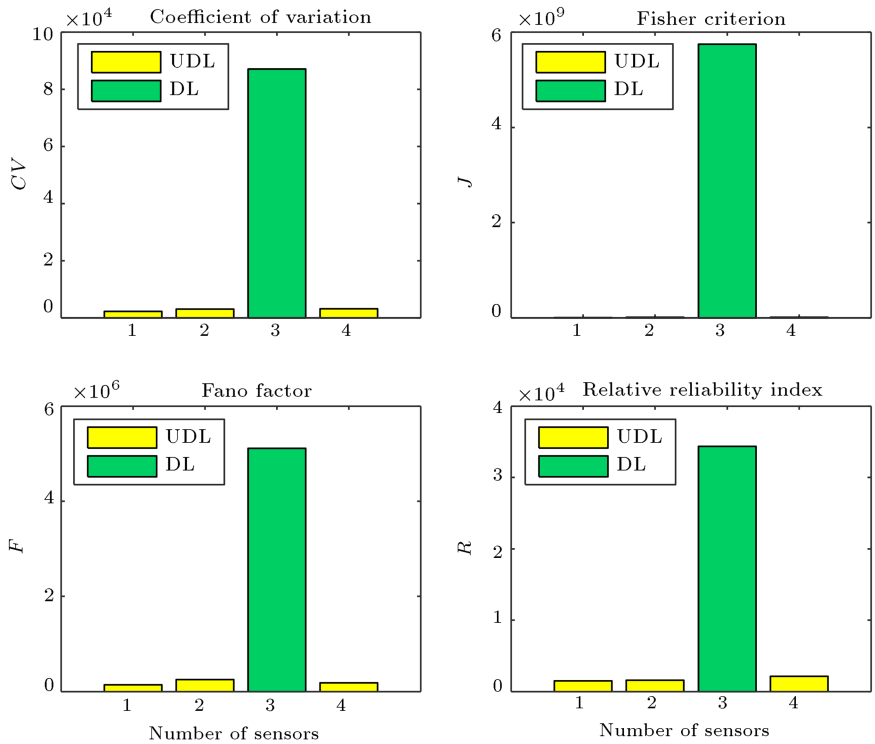

Figure 6. The damage localization in the second damaged case: 30% mass increase.

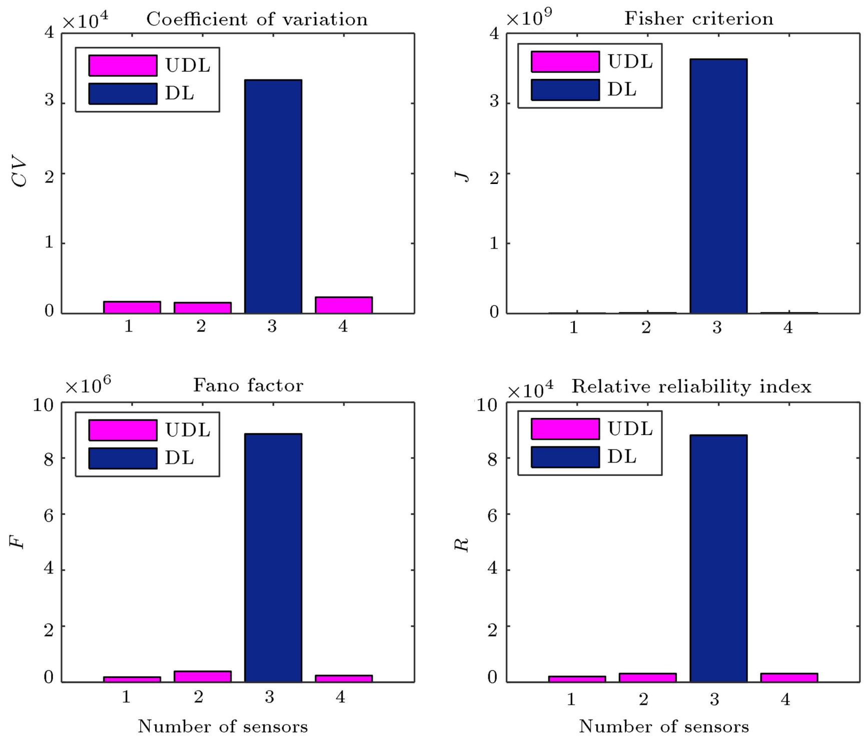

Figure 7. The damage localization in the third damage case: 50% mass increase.

Furthermore, the numerical model of the beam was constructed based on the Euler-Bernoulli theory by presuming that the planes at the ends of the beam remain planes. The total number of sensors is 30 identically spread out at the top and bottom of the beam. The top sensors can be observed in Figure 8. At each sensor location, the acceleration time history was measured in the vertical coordinate.

A uniform transverse random load was applied to the top surface for the excitation of the beam. The load

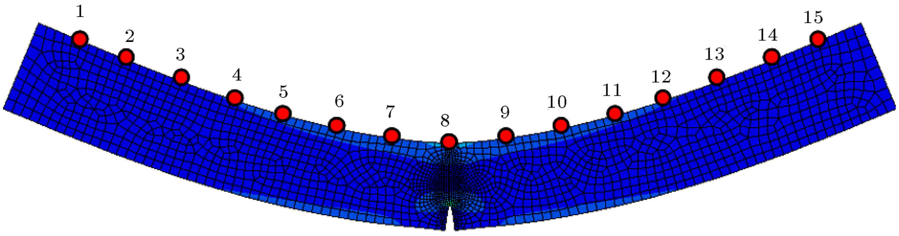

Figure 8. The numerical benchmark model of the concrete beam [48].

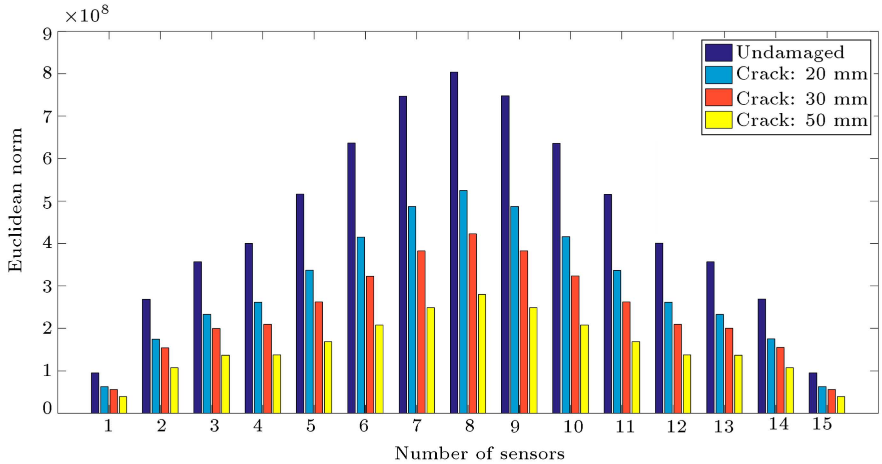

Figure 9. Euclidean norms of the PCA convolution of the training and testing data sets.

histories were lowpass filtered below 1000 Hz, leading to five active modes of the structure. Furthermore, the acceleration time histories were sampled in two seconds with 4001 data points. To simulate damage, a vertical crack was modeled at the bottom of the beam at the location of Sensor 8, as depicted in Figure 8. Such a damage pattern simulates the breathing crack as a more realistic damaged case in many concrete structures, leading to a nonlinear behavior. Several damage patterns along with an undamaged condition (Case 1) were introduced in the numerical beam. These cases consist of different crack lengths including 10, 20, 30, 50, 100, and 150 mm at the middle-span of the beam. In this study, the second (20 mm), third (30 mm) and fourth (50 mm) damage scenarios, i.e., Cases 2–4, are applied to examine the proposed methods for identifying the location of damage.

The main reason for choosing these cases among the above-mentioned scenarios is related to the similarity of the results of damage localization. It should be mentioned that in all cases, the results of damage identification are reasonable and accurate. However, in order avoid presenting similar and repetitive outputs, the cases with the crack sizes of 20, 30, and 50 mm are used. Unlike Ref. [48], the first two measurements of the acceleration responses in the undamaged (measurements 1-2) and damaged (measurements 11-12) cases are chosen to make the data matrices $\mathbf{X}$ and $\mathbf{Y}$. Accordingly, both matrices consist of 8002 samples (rows) and 15 variables (columns). Note that the data matrix $\mathbf{X}$ belongs to the undamaged condition (Case 1) and there are three types of the matrix $\mathbf{Y}$ for Cases 2–4. Based on the PCA technique, the new data matrices, $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$, are initially obtained by transforming and returning the original matrices $\mathbf{X}$ and $\mathbf{Y}$. Using the column vectors of $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$, one can determine the PCA-convolution vectors $\mathbf{c}_x$ and $\mathbf{c}_y$ at each sensor location.

Figure 9 shows Euclidean norms of the PCA-convolution vectors at all sensors. From this figure, it can be observed that the norm of the proposed DSF is reduced through the occurrence of damage in the beam. Furthermore, increase in the damage severity (the crack size) results in a considerable reduction in the norm of the PCA convolutions. In this regard, the highest level of damage (i.e., the 50 mm breathing crack) has the smallest norm. In contrast, the undamaged condition of the beam gives the largest norm value. All of the obtained results lead to the conclusion that the PCA-convolution values are sensitive to damage and their reduction is indicative of the damage occurrence.

In another result, Figure 10 illustrates the vectors of the PCA convolution vectors between the undamaged and damaged conditions at the location of Sensor 8 (the damage area in the numerical beam [48]). This figure clearly indicates the reduction of the PCAconvolution values resulting from the damage. As can be observed, the damaged case with a crack size of 50 mm (Case 4) has the smallest PCA convolution val-

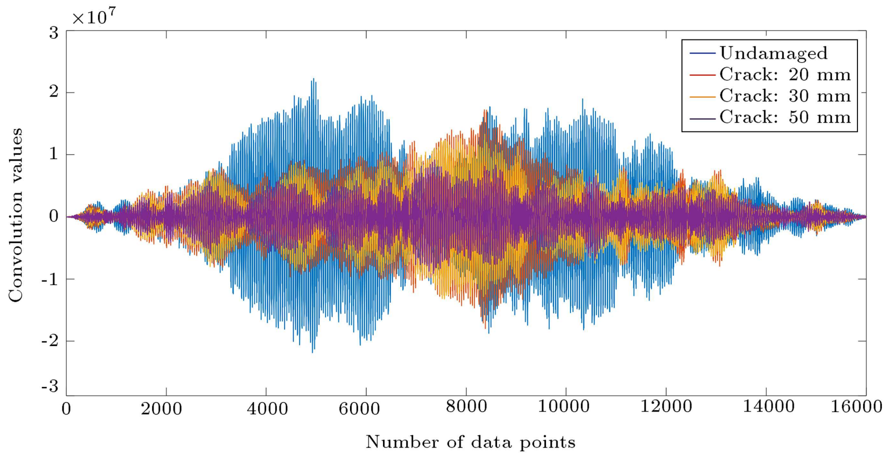

Figure 10. The PCA convolution of training and testing sets at Sensor 8.

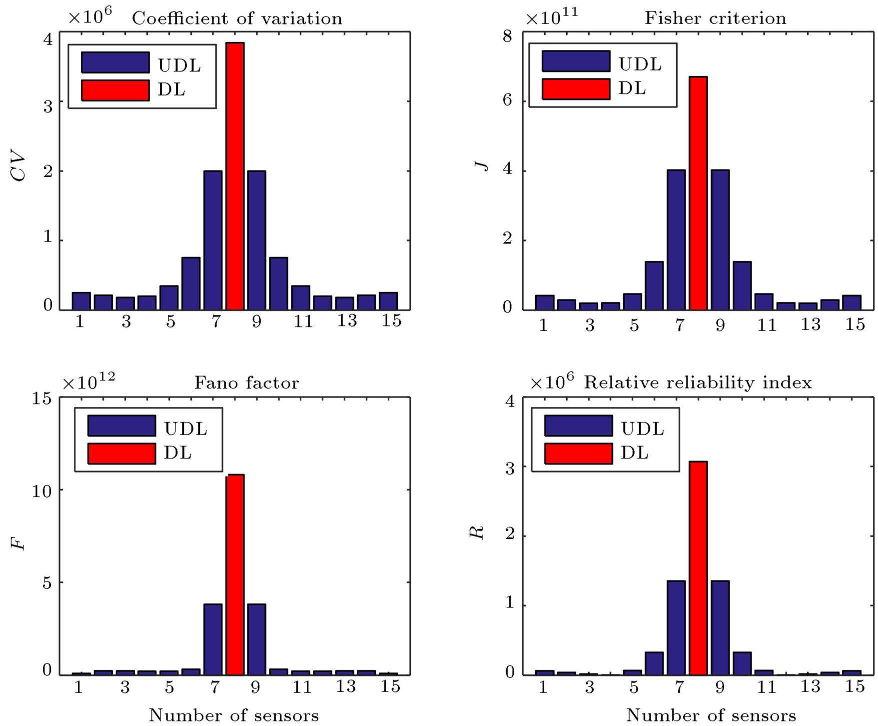

Figure 11. The process of damage localization in the second damage pattern: crack length 20 mm.

ues. The results of the damage localization procedure using the proposed statistical distance measures are shown in Figures 11–13. As can be observed in these figures, the amounts of the distance measures at Sensor 8 (DL) are larger than the other locations (UDL). Hence, the location of this sensor is identified as the damage area of the beam. The obtained results confirm that both of the PCA convolution algorithm and the proposed statistical distance measures are influentially capable of identifying the location of damage.

5. Conclusions

A new feature extraction technique and some efficient statistical distance measures were proposed in this study to identify the damage location. The proposed feature extraction technique was based on the combination of the Principal Component Analysis (PCA) and convolution techniques to extract a new Damage-Sensitive Feature (DSF) by computing the convolution of projecting the measurement data in the undamaged and damaged conditions onto the principal components. The proposed distance measures were the coefficient of variation, Fisher criterion, Fano factor, and relative reliability index. All of them applied the features extracted from the proposed PCA convolution algorithm to identify the damage location. To verify the accuracy and capability of the proposed methods, two numerical models including a four-story shear building and a benchmark concrete beam were used. In

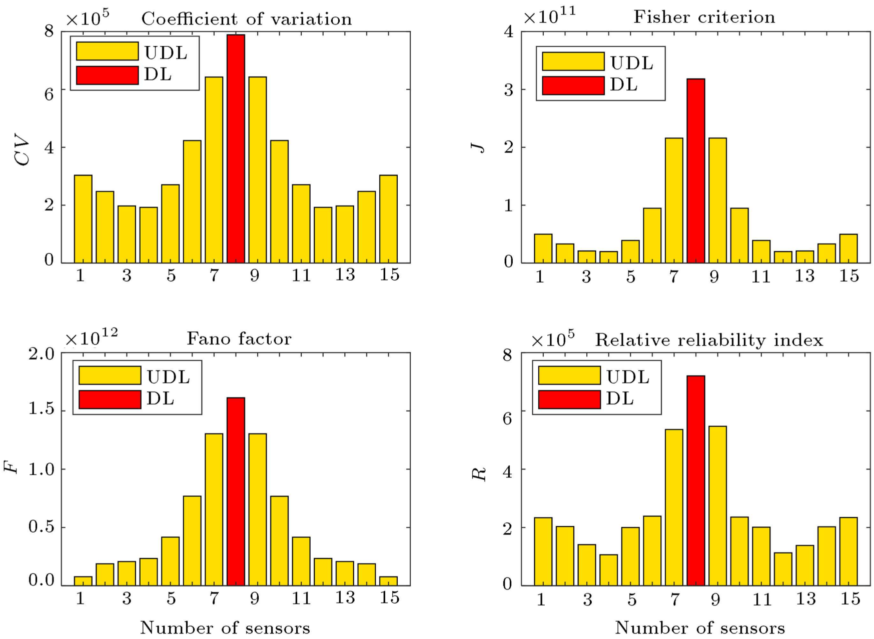

Figure 12. The process of damage localization in the third damage pattern: crack length 30 mm.

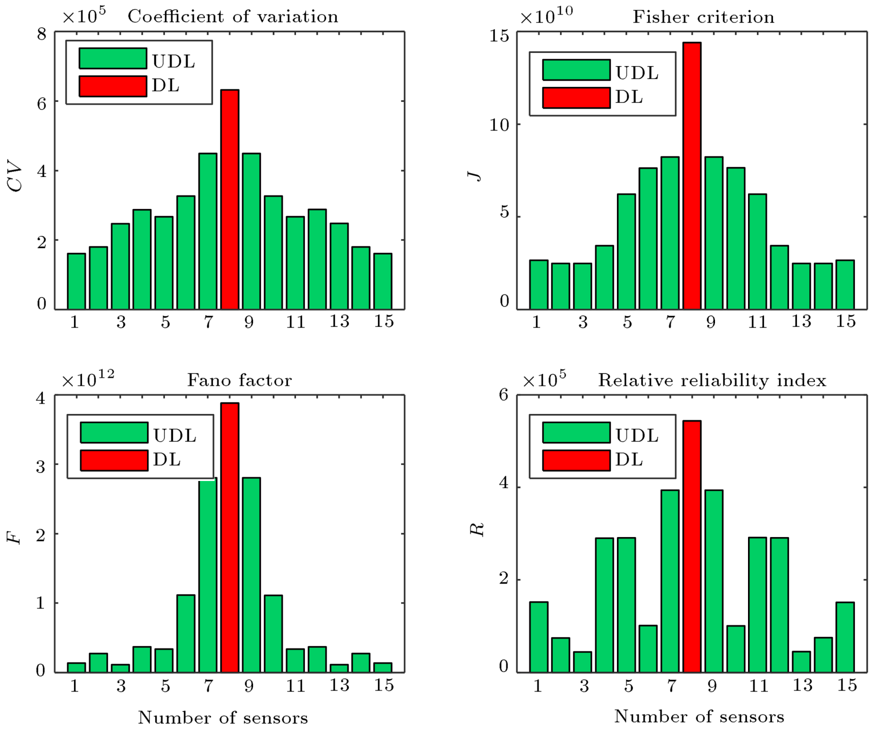

Figure 13. The process of damage localization in the fourth damage pattern: crack length 50 mm.

both models, the numerical results demonstrated that all of the proposed distance measures could identify the location of damage using the DSF extracted from the PCA convolution algorithm. Accordingly, the sensor location concerned with the maximum distance value was identified as the location of damage. Furthermore, the obtained results demonstrated that the proposed DSF was sensitive to damage. For this conclusion, it was observed that the values of the PCA convolution were reduced by increasing the severity of damage. To sum up, it can be concluded that the proposed methods in this study are applicable tools for use in the context of Structural Health Monitoring (SHM). In particular, the proposed feature extraction technique can extract a reliable and sensitive feature from the raw vibration signals through simple but effective algorithms.

Despite the good innovations and results, this study has a few limitations that can be investigated for further studies. In the first limitation, it would be interesting to evaluate the presented methods at least by an experimental example and other types of buildings and civil structures. For the second one, it is desirable in the SHM community to show how the proposed methods, particularly the proposed feature extraction technique, perform well under operational and environmental variability. Finally, it is important to define a threshold boundary for increasing the reliability of damage identification.

Acknowledgement

The authors are fully grateful to Dr. Kullaa for accessing the numerical benchmark concrete beam.

References

1. Qarib, H. and Adeli, H. "Recent advances in health monitoring of civil structures", *Scientia Iranica*, **21**(6), pp. 1733–1742 (2014).
2. Farrar, C.R. and Worden, K., *Structural Health Monitoring: A Machine Learning Perspective*, John Wiley & Sons Ltd (2013).
3. Amezquita-Sanchez, J.P., Valtierra-Rodriguez, M., and Adeli, H. "Wireless smart sensors for monitoring the health condition of civil infrastructure", *Sci. Iranica*, **25**(6), pp. 2913–2925 (2018).
4. Hou, R. and Xia, Y. "Review on the new development of vibration-based damage identification for civil engineering structures: 2010-2019", *J. Sound Vibrat.*, **491**, p. 115741 (2021).
5. Sarmadi, H., Karamodin, A., and Entezami, A. "A new iterative model updating technique based on least squares minimal residual method using measured modal data", *Appl. Math. Model.*, **40**(23), pp. 10323–10341 (2016).
6. Rezaiee-Pajand, M., Sarmadi, H., and Entezami, A. "A hybrid sensitivity function and Lanczos bidiagonalizationTikhonov method for structural model updating: Application to a full-scale bridge structure", *Appl. Math. Model.*, **89**, pp. 860–884 (2021).
7. Rezaiee-Pajand, M., Entezami, A., and Sarmadi, H. "A sensitivity-based finite element model updating based on unconstrained optimization problem and regularized solution methods", *Struct. Contr. Health Monit.*, **27**(5), pp. e2481 (2020).
8. Sarmadi, H., Entezami, A., and Ghalehnovi, M. "On model-based damage detection by an enhanced sensitivity function of modal flexibility and LSMR-Tikhonov method under incomplete noisy modal data", *Eng. Comput.*, **38**(5), pp. 111–127 (2020).
9. Fallahian, S., Joghataie, A., and Kazemi, M.T. "Structural damage detection using time domain responses and teaching-learning-based optimization (TLBO) algorithm", *Sci. Iranica*, **25**(6), pp. 3088–3100 (2018).
10. Toloue, I., Liew, M.S., Harahap, I.S.H., et al. "Damage detection in frame structures using noisy accelerometers and Damage Load Vectors (DLV)", *Sci. Iranica*, **27**(4), pp. 1776–1785 (2020).
11. Dabbagh, H., Ghodrati Amiri, G., and Shaabani, S. "Modal data-based approach to structural damage identification by means of imperialist competitive optimization algorithm", *Sci. Iranica*, **25**(3), pp. 1070–1082 (2018).
12. Amezquita-Sanchez, J. and Adeli, H. "Feature extraction and classification techniques for health monitoring of structures", *Scientia Iranica., Transaction A.*, Civil Engineering, **22**(6), p. 1931 (2015).
13. Amezquita-Sanchez, J.P. and Adeli, H., "Signal processing techniques for vibration-based health monitoring of smart structures", *Archives of Computational Methods in Engineering*, **23**(1), pp. 1–15 (2016).
14. Avendaño-Valencia, L.D. and Fassois, S.D. "Stationary and non-stationary random vibration modelling and analysis for an operating wind turbine", *Mech. Syst. Sig. Process.*, **47**(1–2), pp. 263–285 (2014).
15. Entezami, A., Shariatmadar, H., and Karamodin, A. "Improving feature extraction via time series modeling for structural health monitoring based on unsupervised learning methods", *Sci. Iran., Trans. A, Civ. Eng.*, **27**(3), pp. 1001–1018 (2020).
16. Entezami, A., Shariatmadar, H., and Mariani, S. "Early damage assessment in large-scale structures by innovative statistical pattern recognition methods based on time series modeling and novelty detection", *Advances in Engineering Software*, **150**, pp. 102923 (2020).
17. Vazirizade, S.M., Bakhshi, A., and Bahar, O. "Online nonlinear structural damage detection using Hilbert Huang transform and artificial neural networks", *Sci. Iranica*, **26**(3), pp. 1266–1279 (2019).
18. Noori, M., Wang, H., Altabey, W.A., et al. "A modified wavelet energy rate-based damage identification method for steel bridges", *Sci. Iranica*, **25**(6) Special Issue Dedicated to Professor Goodarz Ahmadi, pp. 3210–3230 (2018).
19. Entezami, A. and Shariatmadar, H. "Damage localization under ambient excitations and non-stationary vibration signals by a new hybrid algorithm for feature extraction and multivariate distance correlation methods", *Structural Health Monitoring*, **18**(2), pp. 347–375 (2019).
20. Tibaduiza, D., Mujica, L., and Rodellar, J. "Damage classification in structural health monitoring using principal component analysis and self-organizing maps", *Structural Control and Health Monitoring*, **20**(10), pp. 1303–1316 (2013).
21. Zhou, Y.-L., Maia, N.M., and Abdel Wahab, M. "Damage detection using transmissibility compressed by principal component analysis enhanced with distance

measure", *J. Vibrat. Control*, **24**(10), pp. 2001–2019 (2018).

22. Silva, M., Santos, A., Santos, R., et al. "Deep principal component analysis: An enhanced approach for structural damage identification", *Struct. Health Monit.*, **18**(5-6), pp. 1444–1463 (2019).

23. Rezvani, K., N.M.M., M., and Sabour, M.H. "A Comparison of some methods for structural damage detection", *Sci. Iranica*, **25**(3), pp. 1312–1322 (2018)

24. Sarmadi, H., Entezami, A., Saeedi Razavi, B., et al. "Ensemble learning-based structural health monitoring by Mahalanobis distance metrics", *Struct. Contr. Health Monit.*, **28**(2), pp. e2663 (2021).

25. Bao, C., Hao, H., and Li, Z.-X., "Integrated ARMA model method for damage detection of subsea pipeline system", *Engineering Structures*, **48**, pp. 176–192 (2013).

26. Sarmadi, H., Entezami, A., and Daneshvar Khorram, M. "Energy-based damage localization under ambient vibration and non-stationary signals by ensemble empirical mode decomposition and Mahalanobis-squared distance", *J. Vibrat. Control*, **26**(11–12), pp. 1012–1027 (2020).

27. Entezami, A., Shariatmadar, H., and Mariani, S. "Fast unsupervised learning methods for structural health monitoring with large vibration data from dense sensor networks", *Structural Health Monitoring*, **19**(6), pp. 1685–1710 (2020).

28. Entezami, A., Sarmadi, H., Behkamal, B., et al. "Big data analytics and structural health monitoring: A statistical pattern recognition-based approach", *Sensors*, **20**(8), p. 2328 (2020).

29. Daneshvar, M.H., Gharighoran, A., Zareei, S.A., et al. "Structural health monitoring using high-dimensional features from time series modeling by innovative hybrid distance-based methods", *J. Civ. Struct. Health Monit.*, **11**, pp. 537–557 (2021).

30. Hoell, S. and Omenzetter, P. "Improved damage detectability in a wind turbine blade by optimal selection of vibration signal correlation coefficients", *Structural Health Monitoring*, **15**(6), pp. 685–705 (2016).

31. Entezami, A., Sarmadi, H., Behkamal, B., et al. "Health monitoring of large-scale civil structures: An approach based on data partitioning and classical multidimensional scaling", *Sensors*, **21**(5), pp. 1646 (2021).

32. Entezami, A., Sarmadi, H., Salar, M., et al. "A novel data-driven method for structural health monitoring under ambient vibration and high dimensional features by robust multidimensional scaling", *Struct. Health Monit.*, **14**(5), pp. 101–131 (2021).

33. Liu, Y., Wang, X., Lin, J., et al. "An adaptive grinding chatter detection method considering the chatter frequency shift characteristic", *Mech. Syst. Sig. Process.*, **142**, pp. 106672 (2020).

34. Mei, Q. and Gül, M. "A crowdsourcing-based methodology using smartphones for bridge health monitoring", *Struct. Health Monit.*, **18**(5-6), pp. 1602–1619 (2019).

35. Amezquita-Sancheza, J.P., Valtierra-Rodriguez, M., and Adeli, H. "Machine learning in structural engineering", *Sci. Iranica*, **27**(6), pp. 2645–2656 (2020).

36. Sarmadi, H. and Entezami, A. "Application of supervised learning to validation of damage detection", *Arch. Appl. Mech.*, **91**(1), pp. 393–410 (2021).

37. Entezami, A., Shariatmadar, H., and Sarmadi, H. "Condition assessment of civil structures for structural health monitoring using supervised learning classification methods", *Iranian Journal of Science and Technology Transactions of Civil Engineering*, **44**, pp. 51–66 (2020).

38. Sarmadi, H. and Yuen, K.-V. "Early damage detection by an innovative unsupervised learning method based on kernel null space and peak-over-threshold", *Comput. Aided Civ. Inf.*, **36**(9), pp. 1150–1167 (2021).

39. Ozdagli, A.I. and Koutsoukos, X. "Machine learning based novelty detection using modal analysis", *Computer-Aided Civil and Infrastructure Engineering*, **34**(12), pp. 1119–1140 (2019).

40. Sarmadi, H. and Karamodin, A. "A novel anomaly detection method based on adaptive Mahalanobis-squared distance and one-class kNN rule for structural health monitoring under environmental effects", *Mech. Syst. Sig. Process.*, **140**, pp. 106495 (2020).

41. Entezami, A., Sarmadi, H., and Saeedi Razavi, B. "An innovative hybrid strategy for structural health monitoring by modal flexibility and clustering methods", *J. Civ. Struct. Health Monit.*, **10**(5), pp. 845–859 (2020).

42. Alamdari, M.M., Rakotoarivelo, T., and Khoa, N.L.D. "A spectral-based clustering for structural health monitoring of the Sydney Harbour Bridge", *Mech. Syst. Sig. Process.*, **87**, pp. 384–400 (2017).

43. Sarmadi, H., Entezami, A., Salar, M., et al. "Bridge health monitoring in environmental variability by new clustering and threshold estimation methods", *J. Civ. Struct. Health Monit.*, **11**(3), pp. 629–644 (2021).

44. Amezquita-Sanchez, J.P., Valtierra-Rodriguez, M., Aldwaik, M., et al. "Neurocomputing in civil infrastructure", *Sci. Iranica*, **23**(6), pp. 2417–2428 (2016).

45. Entezami, A., Sarmadi, H., and Mariani, S. "An unsupervised learning approach for early damage detection by time series analysis and deep neural network to deal with output-only (Big) data", *Eng. Proc.*, **2**(1), p. 17 (2020).

46. Rader, C.M. "Discrete convolutions via mersenne transrorms", *IEEE Transactions on Computers*, **21**(12), pp. 1269–1273 (1972).

47. Selesnick, I.W. and Burrus, C.S. "Fast convolution and filtering", In *CRC Press*, pp. 8–2 (1998).

48. Kullaa, J., Santaoja, K., and Eymery, A. "In vibration-based structural health monitoring of a simulated

beam with a breathing crack, Key Engineering Materials", *Trans. Tech. Publications*, **569**, pp. 1093–1100 (2013).

Biographies

Yousef Ali Feizi obtained his MS degree from Semnan University, Semnan, Iran, and his is currently PhD Candidate at the Department of Civil Engineering, Khorramabad Branch, Islamic Azad University, Khorramabad, Iran. He has authored 8 journal and conference papers. He has also supervised MS degree theses. His research interests include structural health monitoring, structural damage detection, and data mining.

Mohammad Kazem Sharbatdar obtained his MS degree from Amirkabir University, Tehran, Iran and his PhD degree from Ottawa University in Canada. He is currently a Professor at the Faculty of Civil Engineering at Semnan University, Semnan, Iran. He has authored 6 books, more than 110 ISI and ISC journal papers and more than 150 conference papers. He has 5 patents. He has also supervised numerous MS and PhD degree theses.

Reza Mahjoub obtained his PhD degree at the University Technology in Malaysia. He is currently an Assistant Professor at the Faculty of Engineering, Islamic Azad University (IAU), Khorramabad branch. He has produced more than 50 papers in index and non-index journals including conference papers. He has also supervised numerous MS and PhD degree theses.

Mahdi Raftari obtained his PhD degree at the University Technology in Malaysia. He is currently an Assistant Professor at the Faculty of Engineering, Islamic Azad University (IAU), Khorramabad branch, Iran. He is the Dean of Engineering Faculty. He is an active person in research where he has produced more than 50 papers in index and non-index journals including conference papers. He is also a member of several national and international organizations and professional bodies.