

PCA Classification of vibration signals in WSN based oil pipeline monitoring system

Waleed F. Shareef

Nasheed F. Mossa

University of Technology-Iraq, Control and System Department,
Baghdad, Iraq

60026@uotechnology.com

nasheed707@gmail.com

Received : 12\12\2018

Revised ://

Accepted : 26\12\2018

Available online : 25 /1/2019

DOI: 10.29304/jqcm.2019.11.1.469

Abstract:

Using wireless sensor network technology in structure health monitoring applications results in generating large amount of data. To sift through this data and extract useful information an extensive data analysis should be applied. In this paper, a Wireless Sensor Network (WSNs) is proposed for the oil pipeline monitoring system with proposed method for event detection and classification. The method depends on the Principal Component Analysis (PCA). It applied to features extracted from vibration signals of the monitored pipeline. These vibration signals are collected while applying damage events (knocking and drilling) to the oil pipeline. PCA is applied to features extracted from both time domain and frequency domain. The results manifest that this method is able to detect the existence of damage and also to distinguish between the different levels of harmful events applied to the pipeline.

Keywords: WSN, Structural health monitoring, Oil pipeline.

1. Introduction

With the fast evolution of oil, manufacture oil transportation has become more important in economic expansion. In some cases, the most serious damage to transport networks is artificial damage such as ramming, drilling, steel pipe knocking, vehicle movement, etc. to steal oil. At present, the prevailing methods of judging damage are the process of observation of the difference between the pump station and the other, when the pressure is below the required level, then the guess is the existence of damage in oil pipe. Determine the damage area, depending on the human teams which in turn exploring along the pipe path [1]. Monitor the sensitivity of these methods are minimal and are obtained exclusively warning after the occurrence of the damage. The immediate problem to be solved in the detection and prevention interventions by pipeline damage or damage potential, but without causing any damage. This field currently is an active research area. Fang Wang et al., [2] suggested the method of detecting the leakage of the pipeline using statistical features in the time domain of acoustic sensors. These features are taken from the normal (not leakage) sample signals. The vector size of the extracted features is reduced with the PCA method. The model will be feeding Support Vector Data Description (SVDD) trainer features acoustic signal in real time. If the output function of a positive decision, it means that the input signal is abnormal otherwise, the signal is normal. In [3] authors offer the advantage of extracting the vibration signal detected by fiber optic along the crude oil pipeline and the Independent Component Analysis (ICA) warning system. It can be used for fiber-optic distributed along the pipeline to get the vibration signal as well as to determine the leak site and third-party intervention. ICA is applied to separate optical fiber vibration signals from each other, and the relevant signal is extracted. In [4], the authors offer the advantage of the extraction and integration of multiple sensor data in the system of monitoring and pre-warning to secure the pipeline based on multi-seismic sensors. The seismic signals are processed and extracted the features in respective modules. Empirical Mode Decomposition (EMD) was used to analyze signals.

Test devices are composed of many seismic sensors and units of data acquisition and processing. The unit displays three typical target signals respectively, namely individual walking, car moving, and manual drilling. In [5], researchers provide an analysis of the effectiveness of statistical time domain features in determining the vibration signal error. The authors attempted to use the time domain feature to determine the characteristics of mechanical failure of the engine induction. Algorithms are used to determine the features to improve accuracy and reduce the burden of the arithmetic.

In this paper, we studied and developed a novel pipeline security monitoring and early warning system by using WSNs based on PCA. This system consists of data acquisition units that propagate along the pipeline. These sensors and units collect and analysis vibration signals resulting from different targets. The vibration signals are analyzed to obtain useful features. Usually, raw signals are not adequate to identify the existence of a damage; therefore, damage features are extracted from the time domain, frequency domain, or time-frequency domain analysis [7].

2. Architecture of oil pipeline monitoring system

The proposed oil pipeline monitoring system based on WSN illustrated in Figure 1. The system consists of two sensor nodes mounted at the end of 2m carbon steel oil pipeline, and one base station connected to the computer [8]. Base Station node (BS node) or “Monitoring Node”: consists of MCU unit and ZigBee module connected via USB serial cable to the PC. End Points node (EP node) or Vibration Sensing Nodes: These nodes sense the vibration and send their readings to the monitoring node. Each one of these nodes consists of MCU unit, ZigBee module, vibration sensor (accelerometer) and Direct Current (DC) power supply. The distance between sensors is 1.8 m. The pipe mounted on one stand in each end. In this work, a real time vibration signals resulting from the damaging activity on pipe are captured by the accelerometer of each node, and transmitted wirelessly through base station to the computer. The position of (X-Y-Z) axis of the installed accelerometer for each node on the pipe shown in Figure 2.

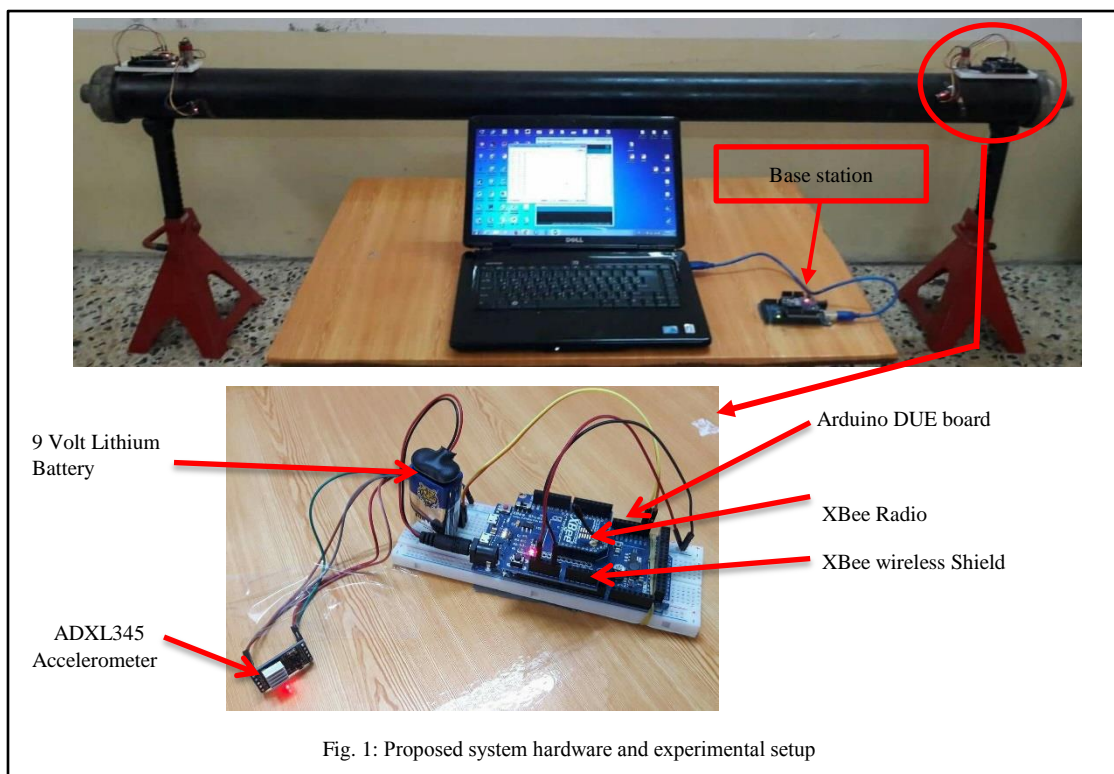


Fig. 1: Proposed system hardware and experimental setup

3. Test scenarios

For the purpose of this paper we subjected the pipe to both knocking and drilling events. These events were applied using a hammer and a handheld electrical drill. To achieve consistent readings, measurements of the two sensor nodes were averaged before sending the data to the base station. Figure 3 shows the real-time vibration data captured by the accelerometer without any damaging event (healthy state of the pipeline). It was observed that signal amplitude varied from one axis to another. This is due to the position on which the accelerometer is installed on the pipeline, as we illustrated in Figure 2. The values of the amplitude of health status on the X-axis is 0.15 G and in Y-axis is -0.07 G, while in the Z-axis was shifted 1 G due to earth gravity. These values are the basic premise that distinguishes the health status of the pipeline, which we can compare the remaining damage events with them to know the nature and type of damage that can be exposed to the pipeline. Figure 4 illustrates the vibration data of knocking the pipe four times in average time equal to approximately (7 seconds).

For each case event four data sets were captured, for the purpose of creating a basic database for each case which can be relied upon as input to the PCA system analysis for diagnosis of the damage events. As shown in Figure 4, one notes that the largest value of amplitude was on the X-axis,

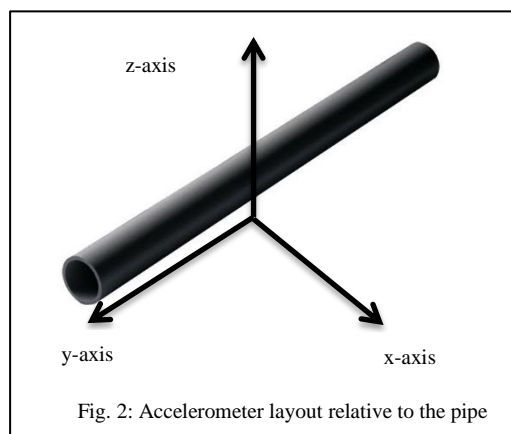


Fig. 2: Accelerometer layout relative to the pipe

where it reached $\pm 2G$ and contains more information than the rest of the other axes. In the Y-axis, the amplitude values were approximately between -1.7 and 1.7, in addition that the information of signals is less than the X-axis. Lastly, the values of Z-axis amplitude reached 2G, which are the minimum values from the other axes. It is clear that the effect of the knock case is very clear on the X-axis than the others.

Figure 5 shows that the process of drilling the pipe three times in total average time equal to (35 seconds). The amplitude values at the pipe-drilling attempt were between $\pm 1G$ for X-axis and Y-axis, while in Z-axis were equal to 2 G. It is clear that the amplitude values in all three axes are convergent.

This is due to the fact that the vibration from the drilling process is very high, which results in the impact of the accelerometer on all axes and regardless of the axis direction relative to pipeline. The process of trying to drilling the pipe was done in different areas along the pipe and also vertically on the pipeline.

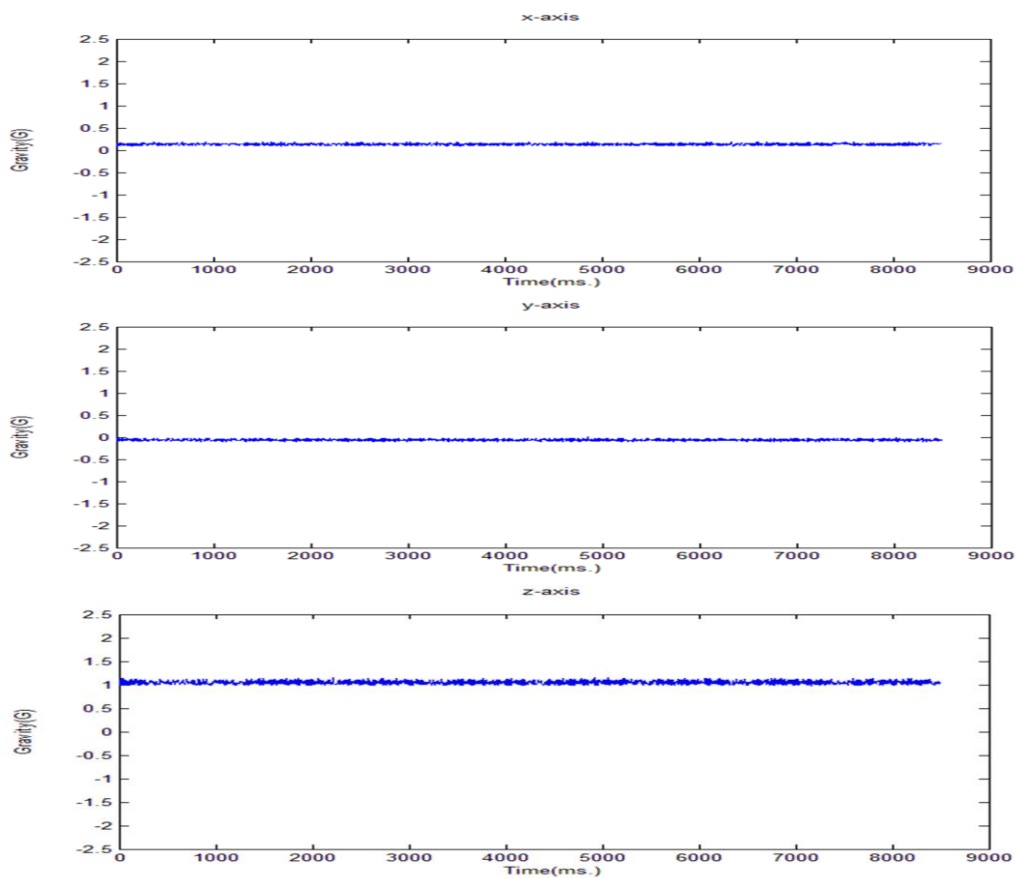


Fig. 3: The signals of the X, Y and Z in time-domain for normal case

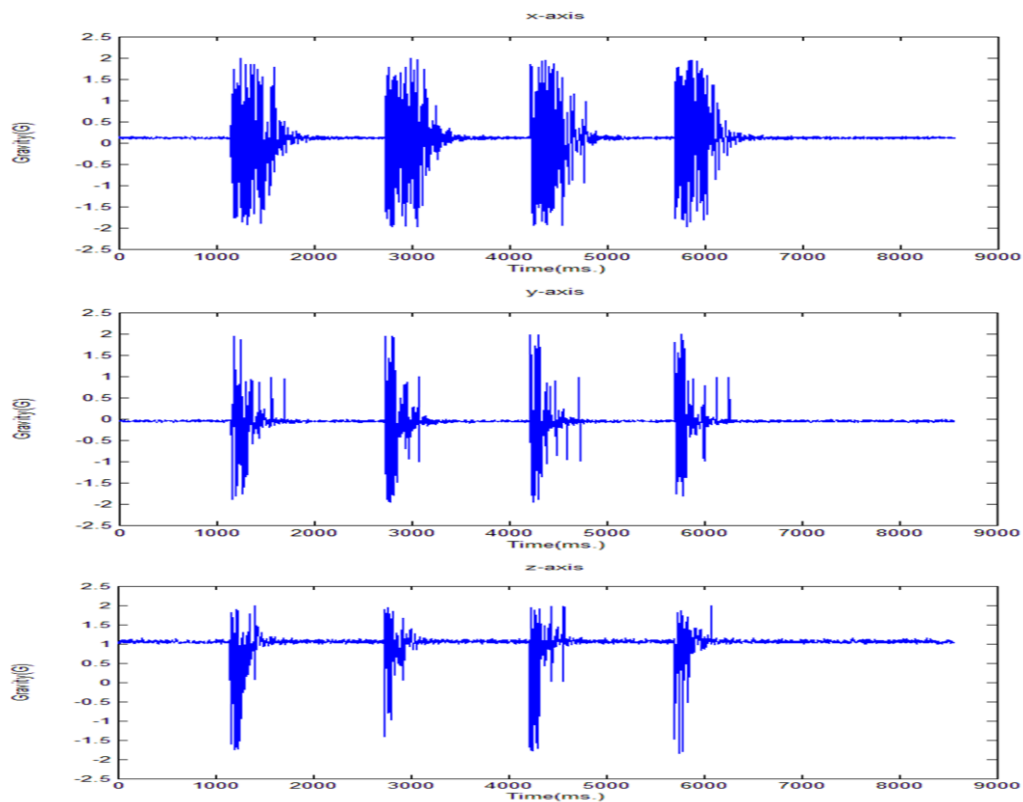


Fig. 4: The signals of the X, Y and Z in time-domain for knocking event case

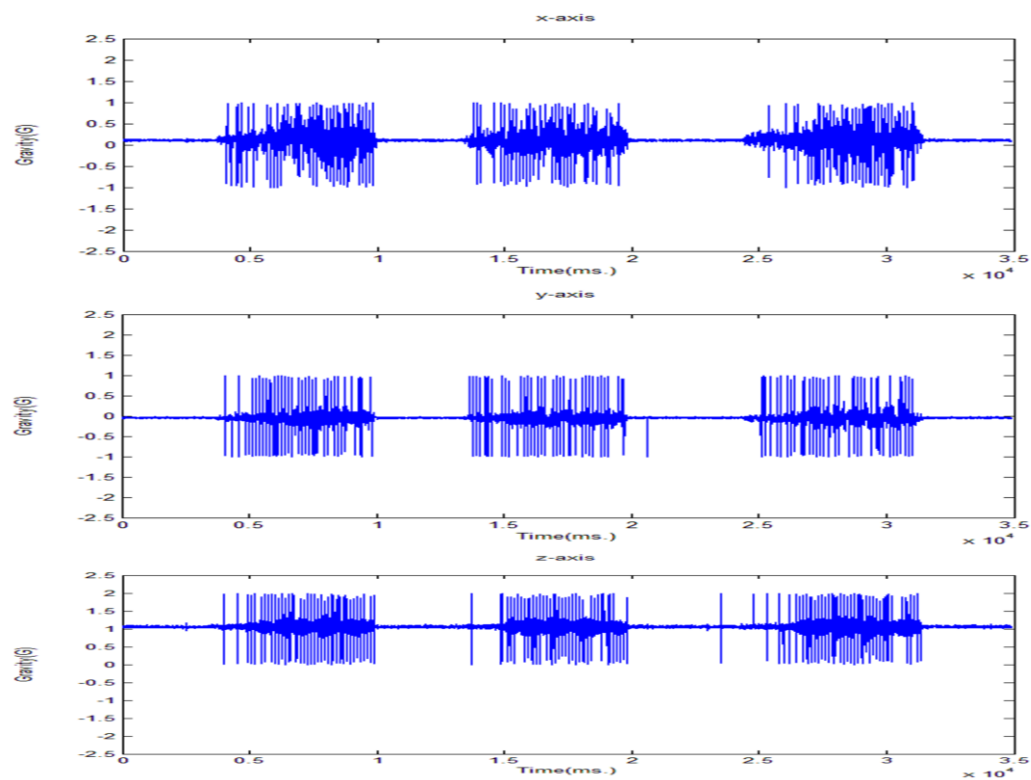


Fig. 5: The signals of the X, Y and Z in time-domain for drilling event case

4. Peak to Peak(pk-pk)

Peak-to-peak refers to the difference between the highest positive and the lowest negative amplitude in a waveform.

4.2. Analysis in frequency domain

Frequency domain representation is the assessment of the strength of various frequency components (the power spectrum) of a time domain signal [6]. FFT is considered as more effective and efficient diagnosis technique to obtain the Fourier Transform of discretized time signals. This signal is considered for a finite time called the “frame” or “time window”, which is then digitized and stored for feature extraction [6].

To detect the error, FFT can be used effectively, a known algorithm and a useful technique for signal analysis. Therefore, most researchers focused on FFT to detect damage in pipes [12]. Also for structural health monitoring(SHM) and in frequency domain analysis of acceleration measurements [14-16]. FFT is an improved algorithm for the performance of a Discrete Fourier Transform (DFT), which is an effective technique for static signal analysis [16]. DFT is defined by Equation (5). In this paper we propose to applying FFT algorithm on real time measurements data set, for the reasons mentions above and for optimizing the system. Once the FFT implementation complete, we calculate frequency domain features which we abbreviate them in other section.

$$X(K) = \frac{1}{N} \sum_{n=0}^{N-1} X(n) W_N^{nk} \quad \text{for } 0 \leq k \leq N-1 \quad \text{and } 0 \leq n \leq N \quad (5)$$

Where:

X(n) is the discrete time signal

N is the sampling period

n and k are the discrete time and frequency indexes, respectively.

The transformation kernel W_N^{nk} is given by Equation (6).

$$W_N^{nk} = \cos\left(\frac{2nk\pi}{N}\right) + j \sin\left(\frac{2nk\pi}{N}\right) \quad (6)$$

Figures 6, 7 and 8 shows the FFT results for one set of data, we notice that the power amplitude values different from one case to others, such as: the normal case power amplitude approximately equal to zero. Because of this case indicate the health state of the pipe (without applied damage events).

The effect of drilling case was on the z-axis greater than the others axis, where the maximum power amplitude equal to 150 at 50 Hz and in x-axis equal to 105 at 9 Hz then in y-axis equal to 55 at 159 Hz. The reason of the minimum power amplitude was in y-axis, as we mentioned the accelerometer installation in method that the y-axis in parallel to the pipe, while the z-axis perpendicular on the pipe. And the other reason that the electric drill used towards z-axis (in a vertical direction on pipe). The knocking case was record its maximum amplitude value in z-axis where equal to 107 in 2 Hz then in x-axis equal to 80 at 60 Hz and in y-axis 47 at 29 Hz.

Generally, the difference of power amplitude values at three axis for the knock cases and there effects on these axis, was not similar to effect of drilling case on the axis (in other word the effect of knock case not be maximum on z-axis and then x-axis and minimum effect on y-axis as in drilling case). Because of the knocking cases performed in different regions on the pipe. Once on vertical direction, the other once on horizontal direction, and once in arbitrary location and so on.

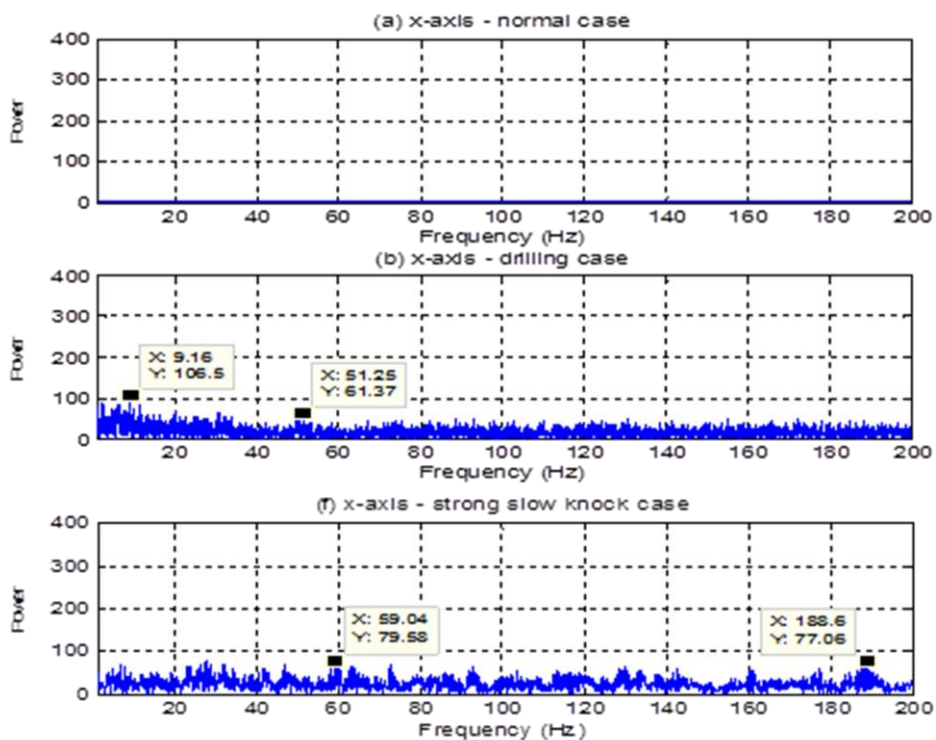


Fig. 6: FFT for x-axis of (a) normal case, (b) drilling case, (c) knocking case.

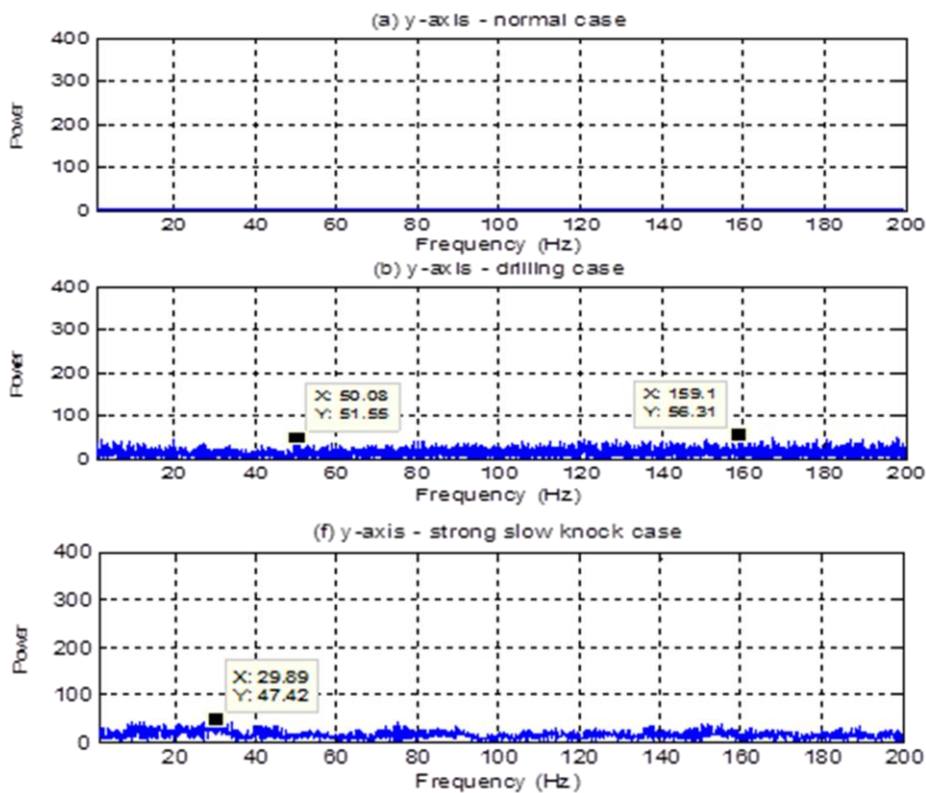


Fig. 7: FFT for y-axis of (a) normal case, (b) drilling case, (c) knocking case

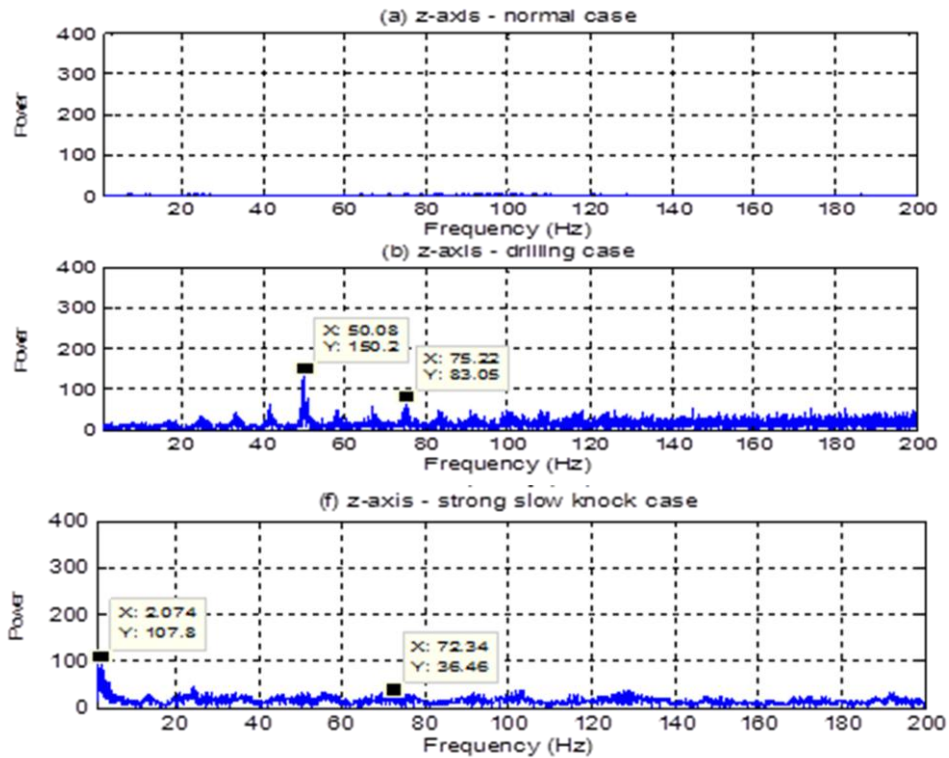


Fig. 8: FFT for z-axis of (a) normal case, (b) drilling case, (c) knocking case

5. Feature extraction in frequency domain

Frequency domain feature extraction is the process of extract useful information from the frequency domain representation of signals. This process is done using the methods discussed in this section. The most frequency-domain features obtained by FFT which considered by researcher are: energy [12], entropy in [11], frequency in [13]. In our proposed system, we propose to use energy and entropy as a features in frequency domain in addition to the time domain features. The energy and entropy are calculated as shown in equations below:

$$Energy = \frac{1}{N} \sum_{j=1}^N (|m_j|)^2 \quad (7)$$

Where:

m_j is FFT component

N is the length of FFT signal

$$Entropy = - \sum_{j=1}^n (p_j * \log(p_j)) \quad (8)$$

Where:

P_j is the probability of signal and must be between 0-1

6. Principal Component Analysis (PCA)

The first step in the analysis of high dimensional data is a dimensionality reduction [17]. Because of two reason; first, it is difficult to interpret multidimensional data sets, and their composition cannot be directly imagined. The second reason is that excessive variables create an empty space and computational problems. PCA is the most useful tool to solve these problems.

PCA is a multivariate statistical technique and powerful tool for analyzing data. It minimizes the feature space dimension by considering the difference of the input data without loss much of information in order to classify patterns in data [18]. The method determines the best projections for the representation of the input data structure. They are selected these projections in a way to enable them to get as much information as possible (i.e. the maximum variation) in a smaller number of dimensions of the region. In order to get the best variation in the data, the data is displayed on a partial space being built by eigenvector of data. In that sense, the eigenvalue corresponding to an eigenvector represents the amount of variance that eigenvector handles.

PCA has been used as a direct method of identifying, classifying and assigning damage as well as an essential step for other methods, and used for face recognition and image compression [17, 18]. PCA used for Structural Health Monitoring (SHM) and has received considerable attention over the past few years [18]. Also, In SHM, the natural frequencies depend not only on the damage but also on environmental conditions, such as temperature and humidity [19]. PCA is used to put this problem in mind because it allows the removal of external factors. As shown in Figure 9 we will explain the steps of implementing PCA algorithm [18].

Step-1 prepare the given data

In this step, the X matrix of size 12×18 (time domain and frequency domain features) was loaded to apply PCA on it.

Step-2 Data standardization

Since physical variables have different amounts and measurements, each data point is standardized. It means subtracting the sample mean from each observation, then dividing by the sample standard deviation. This procedure removes the differences between the scope of variables and giving it the same importance in the data analysis. In addition to the purpose of ensuring that the separation of values and achieving maximum variance between them. When performing this step, anew data matrix was generated (matrix B) in the same dimensions of the original data matrix.

Step-3 The calculation of the covariance matrix (coefficient) & eigenvalues of the covariance matrix (Latent) In this step, the function of PCA in MATLAB was implemented to the matrix B. This function return the coefficient and latent matrix. The result the 18×18 coefficient (loading) matrix and 18×18 eigenvalues (latent).

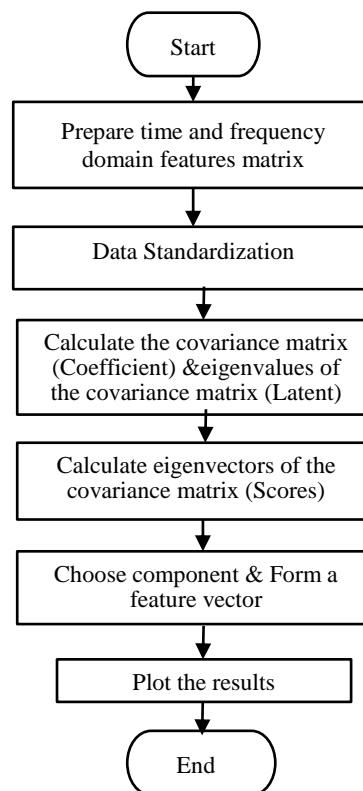


Fig 9: PCA algorithm steps

Step-4 The calculation of eigenvectors of the covariance matrix (Scores) This step, is the calculation of the eigenvectors of the covariance matrix (scores). by multiply the standardized data matrix B by the coefficient (loading) matrix. The result of this step is the eigenvectors of the covariance matrix (scores or in other word called a principal component matrix) with the new higher variance dimension than the original dimension of data 12×18 matrix.

Step-5 Selection of components and the formation of vector features

Here comes the concept of compressing data and reducing the dimensions. If we look at the eigenvectors and eigenvalues values of the previous steps, we will note that the values of eigenvalues are quite different. We notice that the eigenvectors (scores) matrix is arranged in ascending way, so its begin from the lowest significance vector and to the highest significance. In fact, each vector with the higher eigenvalue in latent vector is the principle component of the data set and this gives the components in order of significance.

7. Result and discussion

The result of PCA calculation is 18 principal components (from PCA1 to PCA18), which are time domain and frequency domain features for 3-axis measurements. In order to choose the principal components that separate the damage events clearly, Figure 10 show PCA1 and PCA2. The colored bubbles in the figure above indicate the scenarios of the damage events. Each event we recorded has four data set of measurements, so each four bubbles refer to a cluster of one damage event, where cluster one refers to case, cluster 2 refers to drilling case and cluster 3 refers to knocking case event.

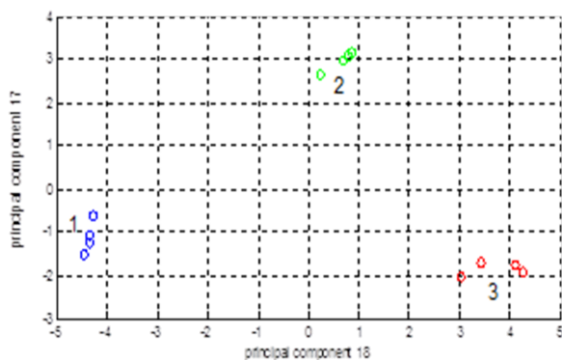


Fig. 12: PCA17 and PCA18

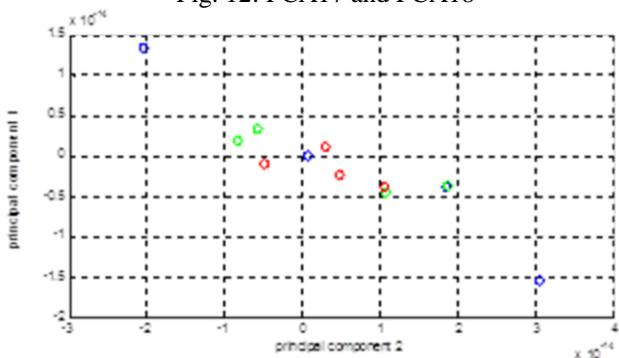


Fig. 10: PCA1 and PCA2

But it is noted that the data are distributed in a manner that cannot be separated or distinguish between them, and gives a state that the primary principal components have very little variance which does not give us the nature of the damage applied on the pipeline. We completed the testing process for others principals components gradually and we choose PCA9 and PCA13, PCA14 and PCA15, also give the same results.

The most significant PCA was PCA18,PCA17 and PCA16 with a percentage approximately equal to 67%, 28% and 3% respectively from the original data, this is make up 98 % from original data. And we can ignore the others PCA which have small variance and don't result to lose much information from original data. And we can ignore the others PCA which have small variance and don't result to lose much information from original data. Figure 11 show PCA16, PCA17 and PCA18.

If we ignore PCA16 which is make up 3% from the original data, the result will be less significant than in Figure 10, this is illustrated in Figure 12. It is clear that the cluster of damaging event are completely separated. This shows that the process of classification and discrimination between scenarios of damaging events can be achieved through our proposed methods .

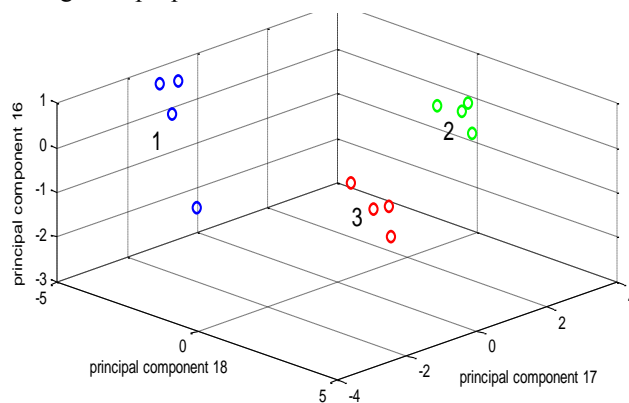


Fig. 11: PCA16,PCA17 and PCA18

8. Conclusions

In this paper, a wireless sensor network in monitoring and pre-warning system for security of oil pipeline based on statistical features and PCA is proposed. The detection model was created with the time domain and frequency domain feature extracted from the vibration signals, which are a result of damage activities on the carbon steel oil pipeline. The vibration signals was captured by accelerometers of wireless sensor nodes along the pipeline. These signals are relayed wirelessly through coordinator node to the PC. In time domain, these features are mean, root mean square, standard deviation and peak-to-peak amplitude. In frequency domain the energy and entropy of the signal spectrum in frequency are used as features. PCA has been applied to these features to analyze the data and reduce the dimension space, and as a result, diagnosis damage on the pipeline is done.

Using PCA with time domain features gives limited detection results comparing to using both time and frequency domain features. PCA was used as a way to identify patterns in data and to express data in a manner that highlighted the similarities and differences between them, in addition to reducing the number of dimensions in data without much loss of information. The results showed that the proposed system alongside with the proposed analysis is able to monitor the healthy state of the pipeline and detect several damaging events.

References

- [1] Iraqi Ministry of Oil, Oil Pipeline Company, document of specifications of the oil pipelines.
- [2] Fang Wang, Weiguo Lin, Zheng Liu, Shuochen Wu, and Xiaobo Qiu, “*Pipeline Leak Detection by Using Time-Domain Statistical Features*”, IEEE sensors journal, VOL. 00, NO. 0, August 2017.
- [3] Lin Zhu , Zhoumo Zeng, Jingchuan Zhang and Shijiu Jin, “*Feature Extraction of Vibration Signal Detected by Optical Fiber along Crude Oil Pipeline and Forewarning System Based on ICA*”, 978-1-4244-3894-5/09/2009 IEEE.
- [4] Jie-di, Sun, Jin-quan, Zhang and Xiao-jun and Wang, “*Feature Extraction and Multi-sensor Data fusion in Monitoring and Pre-warning System for Security of Pipeline Based on Multi-Seismic Sensors*”, IEEE, 978-1-4673-1278, 2012.
- [5] Nayana B R and P Geethanjali, “*Analysis of Statistical Time Domain Features Effectiveness in Identification of Bearing Faults from Vibration Signal*”, JSEN, IEEE Sensors, 2017.2727638.
- [6] D. Goyal and B. S. Pabla, “*The Vibration Monitoring Methods and Signal Processing Techniques for Structural Health Monitoring: A Review*”, CIMNE, Barcelona, Spain, Springer 2015.
- [7] Jing Tian, Carlos Morillo, Michael H. Azarian, and Michael Pecht, “*Motor Bearing Fault Detection Using Spectral Kurtosis-Based Feature Extraction Coupled With K-Nearest Neighbor Distance Analysis*”, IEEE Transactions on Industrial Electronics, VOL. 63, NO. 3, March 2016.
- [8] Nasheed F. Mossa; Dr. Waleed F. Shareef and Faez F. Shareef ’’ *Design of Oil Pipeline Monitoring System based on Wireless Sensor Network*’’, Accepted for publishing in the Iraqi Journal of Computers, Communications and Control & Systems Engineering (IJCCCE) in 2018.
- [9] Angkoon Phinyomark, Pornchai Phukpattaranont and Chusak Limsakul, “*Investigation long-term effects of feature extraction methods for continuous EMG pattern classification*”, World Scientific, at: <https://www.researchgate.com>, Vol. 11, No. 4 (2012).
- [10] Xiaohang Jin, Mingbo Zhao, Tommy W. S. Chow, and Michael Pecht, “*Motor Bearing Fault Diagnosis Using Trace Ratio Linear Discriminant Analysis*”, 2013 IEEE.
- [11] Kamran Javed, Rafael Gouriveau, Noureddine Zerhouni, and Patrick Nectoux, “*Enabling Health Monitoring Approach Based on Vibration Data for Accurate Prognostics*”, IEEE Transactions on Industrial Electronics, VOL. 62, NO. 1, January 2015. .
- [12] Md. Mizanur Rahman and Mohammad Nasir Uddin, “*Online Unbalanced Rotor Fault Detection of an IM Drive Based on Both Time and Frequency Domain Analyses*”, TIA.2017.2691736, IEEE Transactions on Industry Applications.
- [13] Walteneus Dargie, “*Analysis of Time and Frequency Domain Features of Accelerometer Measurements*”, IEEE, 978-1-4244-4581-3/09/2009.
- [14] Yu-Ching Mo, Ke-Yu Su, Wen-Bin Kang, Liang-Bi Chen, Wan-Jung Chang, and Yun Hui Liu, “*An FFT-based High-Speed Spindle Monitoring System for Analyzing Vibrations*”, 2017 Eleventh International Conference on Sensing Technology (ICST), This work was partially supported by the Ministry of Science and Technology (MOST), Taiwan.
- [15] Zhixian Yan, Vigneshwaran Subbaraju, Dipanjan Chakraborty, Archan Misra and Karl Aberer, “*Energy-Efficient Continuous Activity Recognition on Mobile Phones: An Activity-Adaptive Approach*”, International Symposium on Wearable Computers, 2012.

- [16] Rajeev Sharma, Ram Bilas Pachori and U. Rajendra Acharya, “*Application of Entropy Measures on Intrinsic Mode Functions for the Automated Identification of Focal Electroencephalogram Signals*”, Entropy 2015.
- [17] Lindsay I Smith, “*A tutorial on Principal Components Analysis*”, Department of Computer Science, University of Otago New Zealand, February 26, 2002.
- [18] A. Bellino, A. Fasana, L. Garibaldi and S. Marchesiello, “*PCA-based detection of damage in time-varying systems*”, 2010 Elsevier.
- [19] Eyad. I. Abbas, “*Effect of Eigenfaces Level On The Face Recognition Rate Using Principal Component Analysis*”, Eng. &Tech.Journal, Vol.33,Part (A), No.3, 2015.

تصنيف تحليل المكونات الرئيسية لإشارات الاهتزاز في نظام مراقبة خط أنابيب النفط القائم على شبكات الاستشعار اللاسلكية

وليد فواز شريف
نشيد فيصل موسى
الجامعة التكنولوجية / قسم هندسة السيطرة والنظم

المستخلص :

إن استخدام تقنية شبكة الاستشعار اللاسلكية في هيكلية تطبيقات المراقبة الصحية يؤدي إلى توليد كمية كبيرة من البيانات. لفحص هذه البيانات واستخراج المعلومات المفيدة ، يجب تطبيق تحليل شامل للبيانات. في هذا البحث ، تم اقتراح شبكة مستشعر لاسلكي (WSNs) لنظام مراقبة خط أنابيب النفط بالطريقة المقترحة لكشف الحدث وتصنيفه. تعتمد الطريقة على تحليل المكون الرئيسي (PCA). تم تطبيقه على الميزات المستخرجة من إشارات الاهتزاز الخاصة بمراقبة خط الأنابيب. يتم جمع إشارات الاهتزاز هذه أثناء تطبيق أحداث ضارة (الطرق والحفر) على خط أنابيب النفط. يتم تطبيق PCA على الميزات المستخرجة من النطاق الزمني ومجال التردد. تظهر النتائج أن هذه الطريقة قادرة على اكتشاف وجود الضرر وأيضاً للتمييز بين المستويات المختلفة للأحداث الضارة المطبقة على خط الأنابيب.