Smart Diagnosis of Journal Bearing Rotor Systems: Unsupervised Feature Extraction Scheme by Deep Learning

Hyunseok Oh[1], Byung Chul Jeon[2], Joon Ha Jung[3], and Byeng D. Youn[4]

[1,2,3,4] *Department of Mechanical Engineering, Seoul National University, Seoul* 08826, *Republic of Korea*

hsoh52@hotmail.com
puurni@empas.com
reallibero@gmail.com
bdyoun@snu.ac.kr

ABSTRACT

In most approaches for journal bearing rotor system diagnosis, dominant features are manually extracted based on expert's experience and domain knowledge. With the adoption of advanced journal bearings and the limited knowledge about physics-of-failure, the current practice for feature extraction showed limitations for real applications in the power plant industry. To this end, this paper proposes an unsupervised scheme to extract features from correlated vibration signals. First, raw vibration signals from a pair of sensors are preprocessed by generating two-dimensional images. Second, the vibration images are characterized with a HOG (histogram of original gradients) descriptor. Then, deep learning is used to extract relevant features for journal bearing rotor system diagnosis. To demonstrate the validity of the proposed unsupervised-feature-extraction scheme, a case study was conducted with data from the RK4 rotor kit. The results showed that the proposed scheme outperformed existing methods in terms of fault classification accuracy. We anticipate that the proposed scheme is promising as it can minimize the reliance of expert's experience and domain knowledge.

1. INTRODUCTION

Failure of rotors can lead to significant financial loss in engineered systems. To prevent unexpected failure and minimize its downtime, the concept of condition-based maintenance (CBM) is adopted in industrial applications. To realize the CBM, three diagnostic approaches are used: model-based, data-driven, and hybrid. In the data-driven diagnostic approach, the key steps are (1) data acquisition, (2) feature extraction and selection, and (3) fault classification. Among them, the step of feature extraction and selection is known to be most critical and challenging. Therefore, numerous studies addressed how to extract and select features (Ha, Youn, Oh, Han, Jung, & Park, 2016; Lee, Wu, Zhao, Ghaffari, Liao, & Siegel, 2014; Oh, Han, McCluskey, Han, & Youn, 2015; Yang & Kim, 2006).

So far, a number of signal preprocessing techniques were developed to extract features for rotor systems. Some examples are Fourier transform, wavelet transformation, empirical mode decomposition, principal component analysis, and independent component analysis. Excellent studies can be found in the research area of feature extraction. They showed good performance in the rotor system diagnosis. However, a couple of limitations were found in implementing the existing methods in practical applications. First, a considerable amount of expert's experience and domain knowledge is required to extract a number of proper features for the diagnosis of rotor systems. Second, differences in scale and makers warrant unique features for the diagnosis of a particular system, although an identical type of systems such as motors, pumps, and compressors is considered. It can be time-consuming and labor-intensive to customize existing features for a similar system. To overcome the limitations, a better method is required to extract features in the diagnostics of rotor systems.

According to LeCun, Bengio, and Hinton (2015), high-level features in complicated data can be extracted using a general-purpose learning procedures in an unsupervised manner, i.e., deep learning. With the deep learning, a couple of breakthroughs were reported in the area of automatic speech recognition, computer vision, and language translation. Therefore, the deep learning may have a potential for smart diagnosis of rotor systems. There are only a limited number of publications about fault diagnosis using deep learning. Most of them was published in the recent years. A multi-

sensor health diagnosis approach was proposed with the use of the deep belief network (Tamilselvan & Wang, 2013). A hierarchical diagnosis network was constructed based on deep learning (Gan, Wang, & Zhu, 2016). The potentials of deep neural networks were studied to capture fault characteristics and diagnose rotating machinery (Jia, Lei, Lin, Zhou, & Lu, 2016). For rolling element bearing diagnostics, the deep belief network was optimized with the particle swarm algorithm (Shao, Jiang, Zhang, & Niu, 2015). A stacking denoising autoencoders was proposed to detect anomalies in gas turbine combustors with temperature measurement data (Yan & Yi, 2015). The concept of deep belief networks and symbolic time series analysis was proposed to detect combustion instability from flame images (Sarkar, Lore, Sarkar, Ramanan, Chakravarthy, Phoha, & Ray, 2015). In the existing deep learning publications for fault diagnosis, none of them considered correlated signals from a pair of sensors for engineered system diagnostics. The correlated signals are commonly acquired from mission- and safety-critical systems such as steam turbines.

This paper propose an unsupervised scheme to extract features from correlated vibration signals for rotor systems. The remaining sections are organized as follows. Section 2 reviews the theoretical background of deep belief network. Section 3 presents the proposed scheme for smart rotor system diagnosis. Section 4 shows a case study with the journal bearing rotor system. Section 5 concludes the paper with future work.

2. A BRIEF OVERVIEW OF DEEP BELIEF NETWORK

A breakthrough in deep learning was initiated by Hinton, Osindero, and Teh (2006). The key idea of deep learning is to train a hierarchy of features one level at a time, which is referred to as greedy layer-wise unsupervised pretraining. At each iteration, one layer of weights and biases is added to a deep architecture. A set of layers is built after multiple iterations. The iteration process is called pretraining. A pretrained deep architecture can be used to build a classifier.

To form the deep architecture, multiple layers should be stacked in a proper manner. Although there is no consensus on which type of layers should be stacked and how to stack them, several approaches were proposed (Bengio, Courville, & Vincent, 2013): (1) stacking pretrained restricted Boltzmann machines (RBMs) into a deep belief network (DBN), (2) stacking autoencoders (or RBMs) into a deep autoencoder, and (3) constructing a free energy function iteratively. In this section, we overview the first approach since it is widely accepted in the research community of deep learning.

2.1. Pretraining RBM

The RBM consists of one visible layer and one hidden layer (Smolensky, 1986). The energy function of the RBM is:

$$\varepsilon^{RBM}(v,h) = -v^T W h - b^T v - d^T h \tag{1}$$

where v and h are the visible and hidden vectors (a series of binary state of 0 or 1); W is the matrix that encodes the interactions between the visible and hidden nodes; b and d are the vectors that present the visible self-connections and the hidden self-connections (i.e., biases), respectively.

The probability that the RBM assigns to a visible vector is calculated by summing energy over all hidden vectors:

$$P(v) = \frac{1}{Z}\sum_{h} e^{-E(v,h)}, \qquad Z = \sum_{v,h} e^{-E(v,h)} \tag{2}$$

where Z is the partition function that normalizes the probability, $P(v)$. It is worth noting that the probability of having a particular training vector can be increased by adjusting the model parameters (i.e., weights and biases) to lower the energy of the training vector, whereas the probability can be decreased by adjusting them to increase the energy.

The model parameters of the RBM can be calculated with stochastic maximum log-likelihood. The derivative of the log probability of a training vector with respect to a weight is:

$$\frac{\partial \log p(v)}{\partial w_{ij}} = \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}} \tag{3}$$

where subscripts, i and j, indicate the i^{th} visible and j^{th} hidden nodes, respectively; the angle brackets are the expectation of the dot product of the i^{th} visible and j^{th} hidden node binary states. Therefore, the optimization can be stopped when the second term in Eq. (3) converges to the first term. The weight can be adjusted:

$$\Delta w_{ij} = \varepsilon\left(\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}}\right) \tag{4}$$

where ε is the learning rate.

The first term in Eq. (4) is straightforward to calculate when a training vector is given. However, calculating the second term in Eq. (4) is computationally expensive since it requires numerous iterations with an infinite number of samples. To address this, a training method called contrastive divergence (CD) is proposed by Hinton (Hinton 1999):

$$\Delta w_{ij} = \varepsilon\left(\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{reconstruction}}\right) \tag{5}$$

With a given training vector, a single (or a few) iteration of alternating Gibbs sampling can approximate the second term in Eq. (4) without sacrificing accuracy while reducing computational cost. The identical training method can be used to find another model parameters of the RBM, i.e., biases for visible and hidden nodes.

2.2. DBN Construction

A DBN is comprised of multiple RBMs. As mentioned in Section 2.1, a single RBM consists of two layers: one visible layer and one hidden layer. The weights and biases associated with the two layers (i.e., model parameters) can be

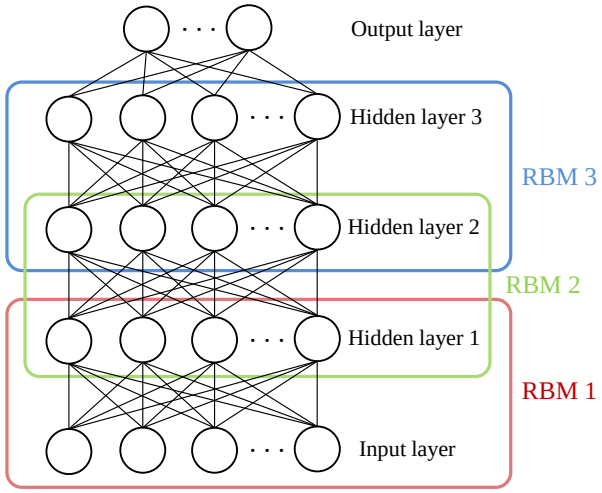


Figure 1. Multilayered architecture for deep belief network.

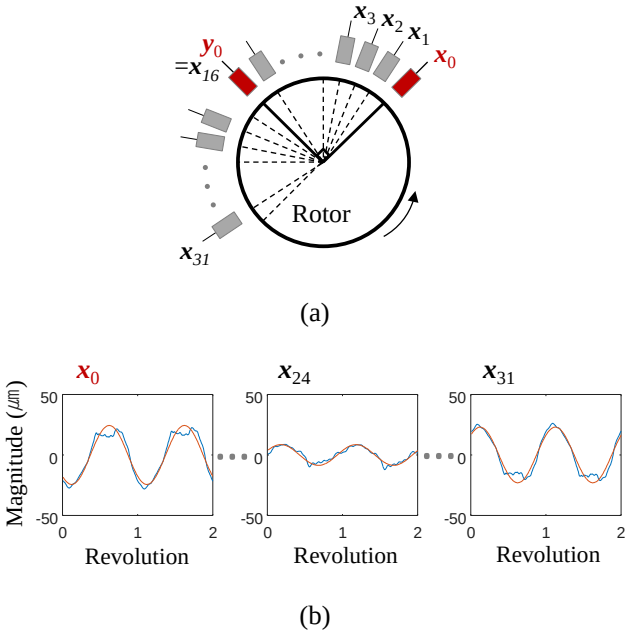


Figure 2. Virtual sensor signals (a) virtual sensor placement and (b) vibration signals generated by the ODR technique (indicated by the blue color): the 1× frequency components are superimposed on the individual vibration signals (indicated by the red color).

determined by training. Then, another RBM is stacked on top of the previous RBM and its model parameters can be determined by the training. Stacking and training of additional RBMs can be repeated in an unsupervised manner. Finally, an output layer is stacked on the top of the stacked RBMs. The whole model parameters can be fine-tuned by the conventional training method such as back-propagation in a supervised manner. As shown in Figure 1, for example, three RBMs are stacked to build a DBN. (1) The first and second layers are the visible and hidden layers of RBM 1, respectively. Model parameters of RBM 1 are determined in an unsupervised manner. (2) The second and third layers become the visible and hidden layers of RBM 2, respectively. Model parameters of RBM 2 are determined in the same manner. (3) The third and fourth layers correspond to the visible and hidden layers of RBM 3, respectively. Model parameters of RBM 3 are determined in the same manner. (4) The fifth layer is added on top of the stacked RBMs to construct the DBN. The whole model parameters are fine-tuned with the back-propagation method.

Section 2 overviewed the DBN architecture and presented how to pretrain the DBN, which is theoretical background of the proposed method in this paper. In Section 3, for smart diagnostics of rotor systems, we present a novel scheme that learns vibration images of correlated signals in an unsupervised manner.

3. PROPOSED SCHEME FOR UNSUPERVISED FEATURE EXTRACTION WITH CORRELATED VIBRATION SIGNALS

Section 3.1 presents how to visualize correlated x- and y-axis vibration signals on two dimensional (2D) images. An omnidirectional regeneration (ODR) technique is employed to generate multiple vibration signals from virtual sensors. Section 3.2 shows an image processing technique, i.e., HOG, to characterize the ODR images. Section 3.3 describes the unsupervised feature extraction scheme with the ODR-based HOG descriptor and deep belief networks.

3.1. Image Generation for Correlated Vibration Signals

The procedures for generating vibration images are as follows. First, raw vibration signals are acquired from actual sensors. Second, additional vibration signals are produced from virtual sensors by the ODR technique. Third, among the produced vibration signals, a reference vibration signal is determined. Fourth, vibration signals are aligned in accordance with the reference vibration signal. And the aligned vibration signals are assembled to produce a single image. The final step is to normalize the magnitude of the vibration in the image. The steps one to five are repeated to generate multiple images using a set of vibration signals acquired at another sampling sequences. We presented each step with details as follows.

In rotor systems, a pair of sensors is mounted on the housing of bearings. For example, gap sensors are placed along with the predefined x- and y-axis directions perpendicular to the

rotor central axis. Signals from the two sensors may not be sufficient to diagnose the rotor system since anomalous behaviors of the rotor (e.g., unbalance) can occur in any direction that is not aligned with the sensing directions. To address this issue, the idea of adding virtual sensor signals given actual sensor signals, so-called ODR technique, was proposed by Jung, Jeon, Youn, Kim, & Kim (2015). The ODR signals from the i^{th} virtual sensors, x_i and y_i, are defined (i=1, 2, …, N):

$$x_i = [\cos(i \times \Delta\theta)]x_0 - [\sin(i \times \Delta\theta)]y_0 \quad (6)$$

$$y_i = [\sin(i \times \Delta\theta)]x_0 + [\cos(i \times \Delta\theta)]y_0 \quad (7)$$

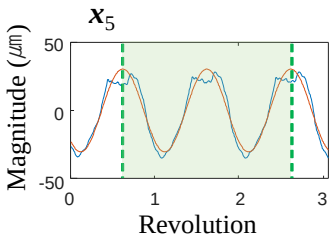
where x_0 and y_0 are the vectors that consist of vibration signals from actual x- and y-axis sensors, respectively; N is the maximum number of the ODR signals to be generated; and $\Delta\theta$ is the increment of the rotation angle. Figure 2(a) illustrates multiple virtual sensors with the two actual sensors (x_0 and y_0). Figure 2(b) shows vibration signals generated by the ODR technique to be used to generate a single image.

With the N number of signals from the virtual sensors, we can generate an image that represents the state of the rotor at a particular sampling sequence. Among the set of N vibration signals, we should select a vibration signal as a reference that is assembled into the first row of the image. In this study, the reference is determined to be the vibration signal that contains the largest magnitude at the first harmonic frequency (1×) of the rotating speed. This is reasonable since an anomalous behavior can be best observed at the 1× frequency component with the largest magnitude.

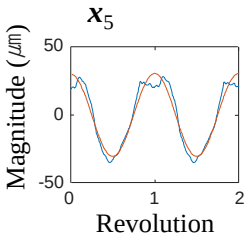
The reference vibration signal is assembled to the first row of the image. However, we need to consider the consistency between images for training of a DBN. Therefore, the first data point in the reference vibration is determined to be that with the maximum magnitude. For example, as shown in Figure 3(a), the first data point of the original reference vibration signal is not the data point with the maximum magnitude. For the purpose of consistency between the images, we defined the reference vibration signal that starts from the data point with the maximum magnitude as shown in Figure 3(b). Other vibration signals are prepared in accordance with the synchronization rule. Then, an image is assembled with from top to bottom. Suppose a vibration signal, x_j, is a reference signal. Then, x_{j+1}, x_{j+2}, x_{j+3}, …, x_j, are assembled into the 2nd, 3rd, 4th, …, N^{th} rows of the image. Figure 4 shows an image generated with correlated vibration signals.

3.2. ODR Image Processing with HOG Descriptor

The histogram of oriented gradients (HOG) is a descriptor that is used to detect objects in computer vision and image processing. A combination of the HOG and SVM was proposed by Dalal and Triggs (2005) to detect human, especially, pedestrians on the image with a large range of poses and backgrounds. Since the introduction, the use of the HOG descriptor has been extended to extract the characteristic of shape-based objects such animals and vehicles. In this study, we employed the HOG descriptor to improve the quality of the input images provided to the DBN for rotor system diagnostics.



(a)



(b)

Figure 3. Vibration signal: (a) reference and (b) phase synchronized ones.

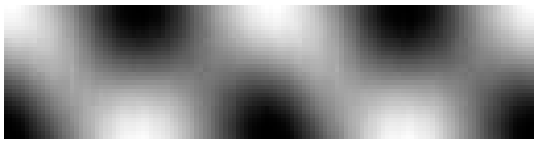


Figure 4. Image of correlated vibration signals measured for two revolutions of the rotor.

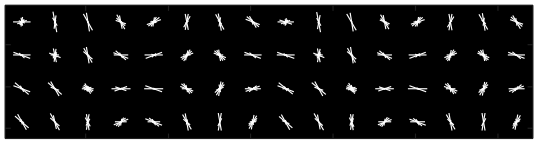


Figure 5. Vibration image after HOG processing.

The key idea of the HOG descriptor is that the shape of an object in local areas can be characterized by a series of histograms of local intensity gradients (LIGs). To implement the idea, an image should be divided into small spatial regions (cells). Then, for each cell, a histogram is drawn with the counts of the number of LIGs. For rotor system diagnostics, the processing of ODR images with the HOG descriptor includes six steps. First, the ODR image is sectioned into multiple cells with n (width) by n (height) pixels. Second, a spatial gradient (i.e. magnitude and orientation) for every

pixel is computed. Third, a histogram is built for each cell by "weight-voting" the gradients. Counts in each bin of the histogram is the number of gradients in a particular range of the direction. Fourth, the counts of the histograms generated for a number of cells is normalized to account for changes in illumination and contrast. Fifth, the normalized histograms are concatenated to produce a series of histograms for a block. Last, the counts of the concatenated histograms are normalized between blocks. An example of the HOG descriptor is prepared with the image in Figure 4. LIGs are shown with gradients for each cell as shown in Figure 5. The final outcome from the HOG descriptor is the concatenated histogram of the cell gradients as shown in Figure 6.

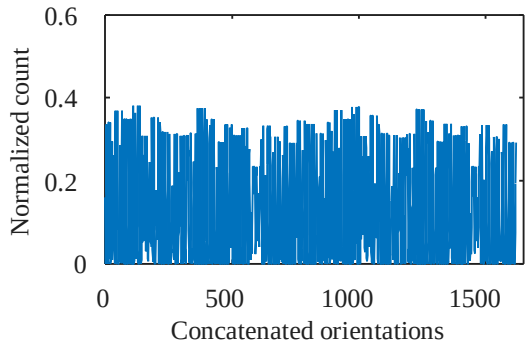
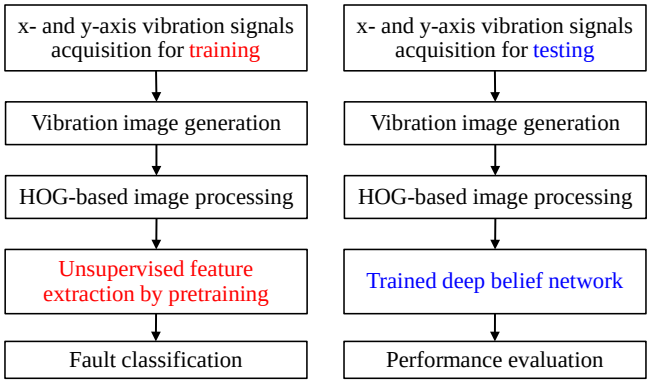


Figure 6. Concatenated histogram.



(a) (b)

Figure 7. Procedures for rotor system diagnostics: (a) training and (b) testing.

3.3. Unsupervised Feature Extraction Scheme with ODR-based HOG Descriptor and DBN

The ODR-based HOG descriptor is provided as inputs of the DBN. The number of hidden layers of the DBN should be sufficient to capture the characteristics of the ODR-based HOG descriptor. The number of nodes of each hidden layer should be determined by the user. Other parameters such as learning rate, batch size, and training-validation-testing data ratio should be also determined. Currently, there is no solid guideline for the determination of the aforementioned parameters. After pretraining of the DBN with training data as explained in Section 2.1, a DBN is constructed. The DBN can perform clustering of given data.

The pretrained DBN can be used for system diagnostics with the supervised learning as specified in Section 2.2. Figure 7 shows the procedures to diagnose rotor systems. It combines the proposed unsupervised feature extraction scheme with the supervised diagnostic method. First, vibration data are collected from rotor systems. For the purpose of training a deep belief network, data from multiple sources such as a testbed in the laboratory and an actual rotor system in the field can be used. The collected vibration data are grouped into two sets including training and test data. Then, vibration images are generated with multiple vibration signals from virtual sensors by the ODR technique. The images are processed by the HOG descriptor to capture the characteristic of the images. For high-level feature generation, unsupervised learning by building a deep belief network is conducted. A large amount of field data without any label can be incorporated in the unsupervised learning since a training data set does not require any label that indicates the condition of the rotor system. Finally, the pretrained deep belief network is fine-tuned in a supervised manner. We evaluated the performance of the proposed diagnostic scheme for rotor system with the test data.

4. CASE STUDY: JOURNAL BEARING ROTOR SYSTEM DIAGNOSIS

This section presents a case study for the diagnostics of journal bearing rotor systems. Section 4.1 describes the data used for the case study. Section 4.2 shows the vibration imaging and HOG descriptor with the data. A deep belief network is used to extract features in an unsupervised manner. The performance of the proposed unsupervised feature extraction scheme is evaluated with the journal bearing rotor systems. The diagnostic accuracy is compared with those from conventional feature extraction methods.

4.1. Description of Data

The data used in this case study are collected from the journal bearing rotor testbed produced by GE Bently-Nevada (i.e., RK4 rotor kit; see Figure 8). Details of data collection can be found in Jeon, Jung, Youn, Kim, and Bae (2015). This section overviews the data briefly. The rotational speed was 3600 rpm. Two gap sensors measured relative displacements of the shaft in the directions of x- and y- axes with the sampling rate of 8,500 data points per second. Each set of x-directional data contained 200 data points that correspond to two revolutions of the rotor. Another set of y-direction data also contained 200 data points. The x- and y-directional data sets were generated for health conditions of (1) normal, (2) rubbing, (3) misalignment, and (4) oil whirl. This was repeated five times. Table 1 summarizes the RK4 data sets used in this case study.

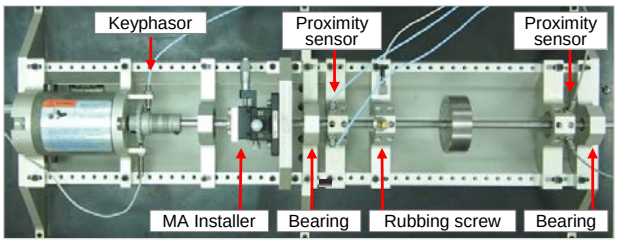
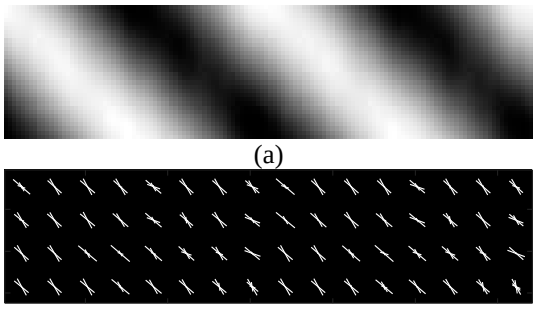


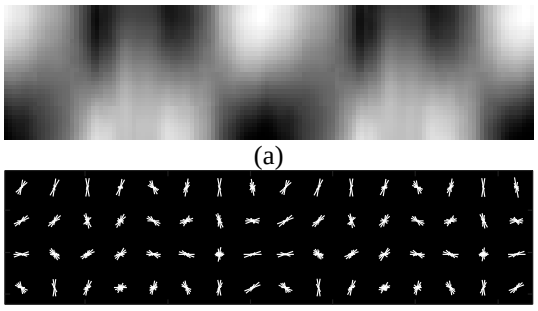
Figure 8. RK4 rotor kit.



(a)

(b)

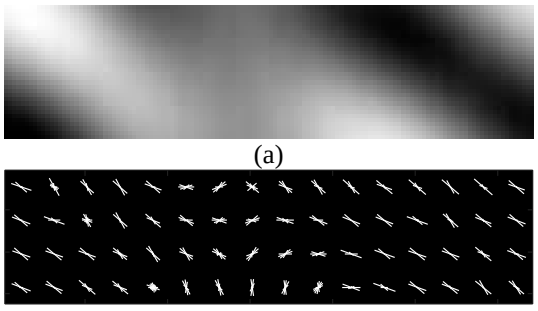
Figure 9. Imaging vibration signals at normal condition: (a) ODR vibration image and (b) ODR vibration image after HOG processing.



(a)

(b)

Figure 10. Imaging vibration signals for rubbing fault: (a) ODR vibration image and (b) ODR vibration image after HOG processing.



(a)

(b)

Figure 11. Imaging vibration signals for oil-whirl fault: (a) ODR vibration image and (b) ODR vibration image after HOG processing.

Table 1. Data points collected from RK4 testbed.

	Normal	Rubbing	Misalignment	Oil whirl
RK4 set 1	200 (x) 200 (y)	200 (x) 200 (y)	200 (x) 200 (y)	200 (x) 200 (y)
RK4 set 2	200 (x) 200 (y)	200 (x) 200 (y)	200 (x) 200 (y)	200 (x) 200 (y)
RK4 set 3	200 (x) 200 (y)	200 (x) 200 (y)	200 (x) 200 (y)	200 (x) 200 (y)
RK4 set 4	200 (x) 200 (y)	200 (x) 200 (y)	200 (x) 200 (y)	200 (x) 200 (y)
RK4 set 5	200 (x) 200 (y)	200 (x) 200 (y)	200 (x) 200 (y)	200 (x) 200 (y)

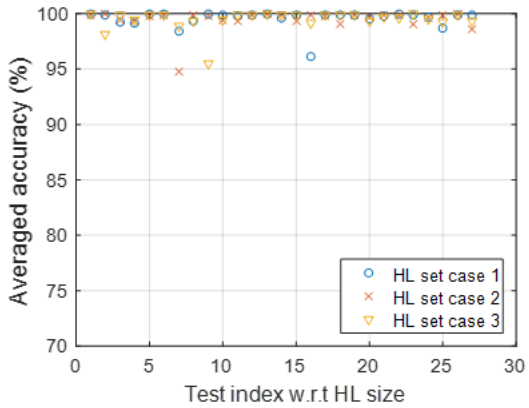


Figure 12. Classification accuracy when the pretrained DBN is combined with the multilayer perceptron classifier.

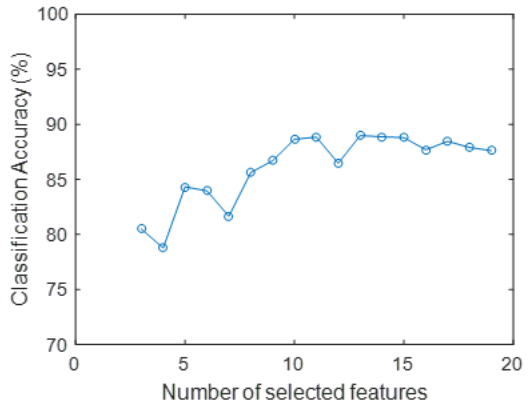


Figure 13. Classification accuracy when time and frequency features are combined with the multilayer perceptron classifier; as the number of features incorporated was increased, the classification accuracy became higher.

4.2. Results and Discussion

The vibration signals acquired from a pair of gap sensors was visualized by the method described in Section 3.1. Then, the

image was processed by the HOG descriptor. Figure 9(a) shows a representative image for the normal health condition, while Figure 9(b) presents an image with gradients that indicate the change of color intensity. In each cell of the image, the gradients were superposed to represent the gradient at individual pixels of the cells. Figure 10 and Figure 11 are the images for the rubbing and oil-whirl faults, respectively.

In the preliminary test, we examined various architecture of the DBN. The details are not shown in this paper since it is out of the scope of this paper. The DBN in this case study had one input, three hidden, and one output layers. The number of nodes for each hidden layer varied from 512 to 2048. A five-fold cross validation technique was employed to reliably evaluate the performance of the DBN.

The classification accuracy with the proposed scheme varied from 95% to 100% as shown in Figure 12. It was corroborated that the variation of the performance was attributed to the number of nodes in the hidden layers of the DBN. We compared the results with the conventional approach with the existing features such as mean, skewness, kurtosis, crest Factor, shape factor, impulse factor, frequency center, RMS at multiple frequencies, (0~0.39x) / 1x, (0.4x~0.49x) / 1x, 0.5x / 1x, (0.51x~0.99x) / 1x, 2x /1x, (3x~5x) / 1x, (3x,5x,7x,9x) / 1x, (1x~10x) / 1x, etc. The extraction of the features required a significant amount time. The selection of efficient features also required considerable efforts. As shown in Figure 13, the MLP classifier with the conventional time and frequency achieved the maximum classification accuracy of 89%, which were lower than the result from the proposed scheme.

5. CONCLUSION

In this study, an unsupervised feature extraction scheme was proposed for smart fault diagnosis of journal bearing rotor systems. The scheme consists of three main steps: (1) vibration image generation, (2) HOG descriptor processing and (3) unsupervised feature extraction by deep learning. First, a pair of vibration signals acquired from a pair of x- and y-axes sensors is visualized by stacking multiple signals from virtual sensors by the omnidirectional regeneration method. Second, the two dimensional vibration images are characterized by the histogram-of-gradients (HOG) descriptor. Finally, features were extracted by deep learning by pretraining the deep belief network (DBN) in an unsupervised manner. We evaluated the performance of the proposed scheme with the data from the RK4 testbed that emulate normal and three faulty conditions (i.e., rubbing, misalignment, and oil-whirl). The fault classification accuracy varied from 95% to 100% depending on the number of nodes in the hidden layers of the DBN. Nevertheless, the accuracy by the proposed method was, on average, 9% higher than that by the fault diagnostic method with the use of a number of existing time and frequency features.

It is anticipated that the proposed scheme helps minimize the reliance of expert's experience and domain knowledge. Consequently, the time and efforts to develop a diagnostic approach to a particular journal bearing rotor system will be reduced considerably. We suggest some future works. First, an optimal architecture of the deep neural networks needs to be studied. Second, another type of deep learning architectures should be examined. Last, the proposed scheme should be implemented to another type of rotating machines, then, non-rotating machines such as inverters.

ACKNOWLEDGEMENT

This work was supported by the Power Generation & Electricity Delivery Core Technology Program of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea (No. 2012101010001C).

REFERENCES

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 1798-1828. doi:10.1109/tpami.2013.50

Gan, M., Wang, C., & Zhu, C. (2016). Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. *Mechanical Systems and Signal Processing*, vol. 72-73, pp. 92-104. doi:10.1016/j.ymssp.2015.11.014

Ha, J.M., Youn, B.D., Oh, H., Han, B., Jung, Y., & Park, J. (2016). Autocorrelation-based time synchronous averaging for condition monitoring of planetary gearboxes in wind turbines. *Mechanical Systems and Signal Processing*, vol. 70-71, pp. 161-175. doi:10.1016/j.ymssp.2015.09.040

Hinton, G.E. (1999). Produts of experts. *In Proceedings of the Ninth International Conference on Artificial Neural Networks*, September 7-10, Edinburgh, UK.

Hinton, G.E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, vol. 18, pp. 1527-1554. doi:10.1162/neco.2006.18.7.1527

Jeon, B.C., Jung, J.H., Youn, B.D., Kim, Y.-W., & Bae, Y.-C. (2015). Datum unit optimization for robustness of a journal bearing diagnosis system. *International Journal of Precision Engineering and Manufacturing*, vol. 16, 2411-2425. doi:10.1007/s12541-015-0311-y

Jia, F., Lei, Y., Lin, J., Zhou, X., & Lu, N. (2016). Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing*, vol. 72-73, pp. 303-315. doi:10.1016/j.ymssp.2015.10.025

Jung, J.H., Jeon, B.C., Youn, B.D., Kim, D., & Kim, Y. (2015) Omni-directional regeneration (ODR) of gap sensor signal for journal bearing system diagnosis, *In Proceedings of the 7th Annual Conference of the Prognostics and Health Management Society*, October 18-22. Coronado, CA

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning *Nature*, vol. 521, pp. 436-444. doi:10.1038/nature14539

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. *Mechanical Systems and Signal Processing*, vol. 42, pp. 314-334. doi:10.1016/j.ymssp.2013.06.004

Oh, H., Han, B., McCluskey, P., Han, C., & Youn, B.D. (2015). Physics-of-failure, condition monitoring, and prognostics of insulated gate bipolar transistor modules: A review. *IEEE Transactions on Power Electronics*, vol. 30, pp. 2413-2426. doi:10.1109/tpel.2014.2346485

Sarkar, S., Lore, K., Sarkar, S., Ramanan, V., Chakravarthy, S., Phoha, S., & Ray, A. (2015). Early detection of combustion instability from hi-speed flame images via deep learning and symbolic time series analysis, *In Proceedings of the 7th Annual Conference of the Prognostics and Health Management Society*, October 18-22. Coronado, CA

Shao, H., Jiang, H., Zhang, X., & Niu, M. (2015). Rolling bearing fault diagnosis using an optimization deep belief network. *Measurement Science and Technology*, vol. 26, 115002. doi:10.1088/0957-0233/26/11/115002

Smolensky, P. (1986). *Information processing in dynamical systems: foundations of harmony theory vol 1. Parallel distributed processing: explorations in the microstructure of cognition*. MA, USA: MIT Press

Tamilselvan, P., & Wang, P. (2013). Failure diagnosis using deep belief learning based health state classification *Reliability Engineering & System Safety*, vol. 115, pp. 124-135. doi:10.1016/j.ress.2013.02.022

Yan, W., & Yu, L. (2015). On accurate and reliable anomaly detection for gas turbine combustors: a deep learning approach, *In Proceedings of the 7th Annual Conference of the Prognostics and Health Management Society*, October 18-22. Coronado, CA

Yang, B., & Kim, K.J. (2006). Application of Dempster-Shafer theory in fault diagnosis of induction motors using vibration and current signals. *Mechanical Systems and Signal Processing*, vol. 20, pp. 403-420.