International Journal of Control, Automation and Systems 17(8) (2019) 2125-2134
http://dx.doi.org/10.1007/s12555-018-0758-6 ISSN:1598-6446 eISSN:2005-4092
http://www.springer.com/12555

# Fault Detection Method Using Multi-mode Principal Component Analysis Based on Gaussian Mixture Model for Sewage Source Heat Pump System

**Young-Jun Yoo**

**Abstract:** This paper presents an algorithm for fault detection of a sewage heat pump system by designing multi-mode principal component analysis with Gaussian mixture model. If the heat pump system fails, the loss of energy and time is enormous, therefore the fault detection of the system is important. For this purpose, this study proposes a fault detection method using multi-mode principal component analysis with Gaussian mixture model. The data were clustered into multi-mode of Gaussian on principal component subspace. Based on the multi-model, the values of Hotelling's $T^2$ and $SPE$ were calculated and used for the fault detection as indexes that are compared performance with clustering model using $k$-means and $k$-medoids algorithm as well as conventional PCA. Actual data of the sewage heat pump were used to verify the proposed method. The results of the fault detection performance show that the proposed model shows the best performance of fault detection among the conventional, $k$-means, and $k$-medoids PCA models.

**Keywords:** Fault clustering, fault detection, Gaussian mixture model, principal component analysis, sewage source heat pump system.

## 1. INTRODUCTION

A heat pump is a high-efficiency device that transfers heat from a colder area to a hotter area by using mechanical energy. In addition, the electrically driven heat pump is a clean and safe system that operates with air, water, geothermal (ground water) as a heat source, no emission of harmful gas, no risk of explosion. Because of the advantages of the system, many researchers have conducted numerous studies about the heat pump system recently. Baek [1] investigated the feasibility of the waste water usage for a heat pump as a heat source and to obtain engineering data for system design. Funamizu [2] presented characteristics of heat energy in waste-water, reuse plans, and some experiences in Japan. This paper discusses full-scale reuse projects for heating and cooling in the Tokyo metropolitan districts and project for melting snow in Sapporo city. The soft-dirt characteristic of the heat-exchanging pipe in a sewage heat pump system is studied [3, 4]. These studies present the technical and economic analysis of the increase in heat pump temperature in the sewage disposal process. However, the sewage heat pump system faults could be caused such as evaporator/condenser fault by reduction of heat transfer area wastewater sludge, an increment of superheat temperature by expansion valve malfunction, and volumetric efficiency reduction by compressor malfunction [5]. The faults of the heat pump system may cause energy wastage, system unreliability, and shorter equipment life. Therefore, it is necessary to develop a fault detection algorithm of a sewage source heat pump equipment to operate with high efficiency [6].

When a mechanical and chemical system is defective or damaged, the system downtime can be caused, and the social and economic damage resulting from this can be enormous. Therefore, it is important to carry out preventive maintenance before a failure occurs, and most systems currently rely on preventive maintenance to be performed at regular time intervals. However, periodic preventive maintenance is carried out at regular intervals, regardless of whether the system is actually defective or not. Therefore, it is costly to lose unnecessary replacement of normal parts and there is a limit to prevent sudden system failure. In order to prevent these sudden system failures, research through the data-driven approach using statistical methods [7–12] or optimization algorithm [13–18] has been conducted recently such as rotating machinery fields [13, 14, 17, 18]. These kinds of approaches are to deduce the reliability of the system.

A principal component analysis (PCA) is a one of the most widely used the fault monitoring methods of the data-driven method. PCA is a multivariate statistical analysis method that reduces dimensions while maintaining essential information of huge data. This method improves

Manuscript received October 19, 2018; revised December 23, 2018 and January 9, 2019; accepted February 19, 2019. Recommended by Associate Editor Ning Sun under the direction of Editor Young IL Lee.

Young-Jun Yoo is with the School of Electrical Engineering, Pohang University of Science and Technology, San 31, Hyojadong, Namgu, Pohang, Gyungbuk 37673, Korea (e-mail: youdalj@postech.ac.kr).

 

defect detection and diagnostic capability. Studies on the application to the PCA process industry are discussed in [7, 8].

In the HVAC and refrigeration field, a study on fault diagnosis using PCA [20–23] was carried out. The study [20] proposes the method exploits PCA to distinguish anomalies from normal operation variability and a reconstruction-based contribution approach to isolate variables related to faults. Some statistical training data cleaning strategy is presented [21] for PCA-based chiller sensor fault detection, diagnosis and data reconstruction method. In this study, the training data quality can be improved by the presented data-cleaning strategy, finding and removing outliers from the original training data set. Nunzio Cotrufo and Radu Zmeureanu [22] proposed new PCA-based method of soft fault detection and identification for the ongoing commissioning of chillers, which is composed of the three main phases: threshold model training, outliers detection, and variables identification. PCA-R-SVDD method [23] is proposed for the diagnosis of chillers. The proposed method shows significant improvement compared with the traditional methods due to the better fault data distribution and tighter monitoring statistic.

Application of traditional process monitoring methods based on the assumption that the process has only one stable operation region may cause false alarms when the process is operated under another steady-state nominal operating mode. This is because of different modes of the process that has different statistical properties such as mean value, variance, and correlation between variables. The multi-mode PCA modeling using Gaussian mixture model [9, 10] or k-nearest neighbor algorithm [11, 12] are presented to cover multiple operation modes with different statistical properties.

Likewise, sewage source heat pump system may not have one feature of mean value, variance, and correlation. The paper [6] proposes a fault detection algorithm using PCA of one mean and variance. In this paper, the PCA-based fault detection method has been proposed for the sewage source heat pump system. Some actual operational data of the sewage source heat pump unit were collected and a PCA model is designed. The square prediction error (SPE) and Hotelling's $T^2$ are used to detect faults. However, the accuracy of the fault detection decreases even though, Hotelling's $T^2$ exceeds its limit at some operation point.

This paper proposes a fault detection algorithm based on multi-mode PCA for the sewage heat pump system. Most of the multi-mode PCA is modeled in 2-dimensional principal component subspace [9–12]. However, if the proportion of accumulation of the eigenvalue of the first and second principal components is not sufficient, the clustered data could not describe the trend of the original data. To extend the availability of the multi-mode PCA, this paper clusters the data using Gaussian mixture model in 3-dimensional principal component (PC) subspace that includes the third PC component. Using the Gaussian mixture model, the data are clustered recursively with the parameters of the mean and standard deviation according to the number of specified mode. Base on the clustered data group, the PCA model can be designed. When test data come in the Gaussian mixture model, the proposed algorithm decides which group the data belongs to, calculates Hotelling's $T^2$ and *SPE* for that group, and uses them for fault detection. The faulty state is determined by the proposed multi-mode clustering model without any previously defined set-point information. The proposed model is verified with sewage heat pump data and compared performance with clustering model using $k$-means and $k$-medoids algorithm as well as conventional PCA. The $T^2$ chart and the *SPE* chart show that the proposed model shows the best performance of fault detection among the conventional, $k$-means, and $k$-medoids PCA models.

## 2. PRELIMINARY

### 2.1. Principal component analysis

PCA is a multivariate statistical analysis method that reduces dimensions while maintaining essential information of huge data. This method reduces the dimensionality of the data and analyzes them by newly setting the principal component (PC) in the order that the variable's covariance is the largest.

The measured data matrix $X \in \mathbb{R}^{n\times p}$, which has $n$ observations (samples) of $p$ measurement variables, can be decomposed to produce loading vectors corresponding to the largest singular values in order to capture the variations of the variables optimally. PCA determines a set of principal component loading matrix $V \in \mathbb{R}^{n\times p}$ and diagonal matrix $\Lambda \in \mathbb{R}^{p\times p}$, ordered by the amount of variance explained in the loading vector directions and solving an eigenvalue decomposition of the sample covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ as

$$\Sigma = \frac{1}{n-1}X^TX = V\Lambda V^T, \tag{1}$$

where $\Lambda = diag(\lambda_1, \lambda_2, \cdot, \lambda_p)$ and $\lambda_1, \lambda_2, \cdot, \lambda_p$ are arranged in the descending order.

A PCA model is usually built from a few PCs. These principal components are the results of decomposing a data matrix $X$ using PCA as follows:

$$X = TV^T + E = \sum_{i=1}^{k} t_i v_i^T + E, \tag{2}$$

where $k$ is the number of principal component loading vectors, $T = [t_1, t_2, \cdots, t_k] \in \mathbb{R}^{n\times k}$ is a principal component scores, $V = [v_1, v_2, \cdots, v_k] \in \mathbb{R}^{p\times k}$ is the loadings to be estimated, $E \in \mathbb{R}^{n\times p}$ is the residual term, $v_i \in \mathbb{R}^p$ is a principal component loading vector and $t_i \in \mathbb{R}^n$ is score vector of the PCA model.

### 2.2. Hotelling's $T^2$ calculation

Hotelling's $T^2$ statistic represents the difference between the point at which the original data set is projected onto the PC and the PC center point. If the scores exceed the confidence limit of Hotelling's $T^2$ at $i$-th sample point, an abnormality of the system behavior could be detected. The $T^2$ statistic for the lower-dimensional space can be calculated for each new observation $x \in \mathbb{R}^p$ by:

$$T^2 = x^T P(\Sigma_a)^{-2} P^T x, \tag{3}$$

where $\Sigma_a$ represents the non-negative real eigenvalues corresponding to the $P$ principal components.
The upper confidence limit for $T^2$ is obtained using the F-distribution:

$$T^2_{a,n,\alpha} = \frac{a(n-1)}{(n-a)} F_{a,n-a,\alpha}, \tag{4}$$

where $n$ is the number of samples in the data set, $a$ represents the number of principal components and $\alpha$ shows the level of significance.

**Remark 1:** SPE or Q statistics shows the variability of the residual subspace (RS). This can also be used as an indicator. SPE values and calculation formulas can be found in [7–12].

### 2.3. Process flow of conventional PCA fault detection

The process using conventional PCA (Fig. 1) is described for the fault detection. In order to make a PCA model, pre-processing of model data should be performed first. It is necessary to determine which variables to use, remove outliers, and normalize the data to design PCA modeling. PCA model is designed through the process of selecting the number of PCs so that the cumulative value of eigenvalue can represent the characteristics of the data and determining the control limit. The method of monitoring the fault is to calculate the value of $T^2$ by normalized test data with the designed PCA model. In this method, it is determined a fault operation depending on whether the value of $T^2$ is exceeded or not.

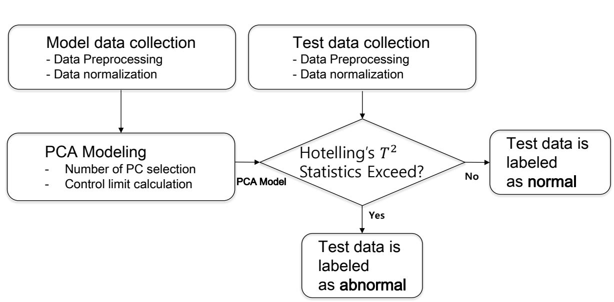


Fig. 1. Flowchart of fault detection with the conventional PCA.

## 3. PROPOSED METHOD

Conventional PCA-based fault monitoring methods are assumed that the monitoring variables are normally distributed with single mean and covariance. However, for real cases of the monitoring process, the variables have a more complex distribution because of the nonlinearity or the dynamics of the system and multi-operation mode. In order to deal with these complex probability distributions, this section provides a method to perform more accurate fault monitoring.

### 3.1. Gaussian mixture clustering

Gaussian mixture clustering is an algorithm that finds the most optimal Gaussian mixture model (GMM) of the mean and covariance if the number of clustering groups is decided. In other words, when the data are assumed to be k-Gaussian, GMM algorithm finds k-group Gaussian model that has values of mean and covariance which could explain the most data well.

Among the most famous algorithm to solve the GMM is Expectation Maximization (EM) algorithm. In expectation-step (E-step), the value of the latent variable is found with the highest "expectation" when $\mu$, $\Sigma$, $\pi$ are currently given. In maximization-step (M-step), newly estimated latent variable is used to calculate its value to maximize $\mu$, $\Sigma$, $\pi$. The EM algorithm is repeated E-step and M-step are alternately performed until the likelihood value is converged.

In order to solve this problem using the EM algorithm, GMM will introduce a latent variable $z \in \mathbb{R}^k$. The j-th variable of $z$, $z_j$ is a binary random variable and satisfies following conditions that $\mathbf{p}(z_j = 1) = \pi_j$, $\Sigma_j z_j = 1$ and $\Sigma_j \pi_j = 1$.
The marginal probability of $z$ can be computed as follows:

$$\mathbf{p}(z) = \Pi_j \pi_j^{z_j}, \tag{5}$$

and the conditional distribution of the specified data $x$ can be expressed as

$$\mathbf{p}(x \mid z) = \Pi_j N(x \mid \mu_j, \Sigma_j)^{z_j}. \tag{6}$$

Therefore, the joint distribution could be formulated and marginalized as

$$\mathbf{p}(x) = \Sigma_z \mathbf{p}(x,z) = \Sigma_z \mathbf{p}(z)\mathbf{p}(x \mid z) = \Sigma_z \pi_j N(x \mid \mu_j \Sigma_j). \tag{7}$$

Based on the above results, it is possible to apply the EM algorithm. First, $\mathbf{p}(z_j = 1 \mid x)$ is calculated in E-step. This is the process of calculating probability or posterior belonging to the cluster of $j$ for each sample data. This value can easily be calculated via the Bayes' rule as follows:

$$\mathbf{p}(z_j = 1 \mid x) = \frac{\mathbf{p}(z_j = 1)\mathbf{p}(x \mid z_j = 1)}{\Sigma_j^k \mathbf{p}(z_j = 1)\mathbf{p}(x \mid z_j = 1)}$$

$$=\frac{\pi_j N(x \mid \mu_j, \Sigma_j)}{\Sigma_j^k \pi_j N(x \mid \mu_j, \Sigma_j)}. \tag{8}$$

Next, when $z$ is fixed, the remaining parameters can be calculated as

$$\mu_j = \frac{\Sigma_i \mathbf{p}(z_{ij}=1 \mid x) x_i}{\Sigma_i \mathbf{p}(z_j = 1 \mid x)}, \tag{9}$$

$$\Sigma_j = \frac{\Sigma_i \mathbf{p}(z_{ij}=1 \mid x)(x_i-\mu_j)(x_i-\mu_j)^T}{\Sigma_i \mathbf{p}(z_j=1 \mid x)}, \tag{10}$$

$$\pi_j = \frac{\Sigma i \mathbf{p}(z_j = 1 \mid x)}{N}. \tag{11}$$

The performance measure of GMM is log likelihood function. That is, the objective of the GMM clustering is to find a parameter that maximizes the probability of the data $x \in \mathbb{R}^n$ of a given parameter. Log likelihood is defined by ln $\mathbf{p}(x \mid \theta)$, where $\mathbf{p}(x \mid \theta)$ is the component density given parameter $\theta$. The parameter $\theta$ will be found is composed of the mean $\mu_j$, covariance $\Sigma_j$ of each Gaussian, and the probability $\pi_j$ that each sample data belongs to each Gaussian. Therefore, if the multi-Gaussian distribution of $x_i$ of given $\mu_j$, $\Sigma_j$ is defined as $N(x_i \mid \mu_j, \Sigma j)$, and the latent variable $z$ is introduced, the log likelihood function can be written as

$$\ln \mathbf{p}(x \mid \pi, \mu, \Sigma) = \sum_i^n \ln \sum_j^k \pi_j N(x_i \mid \mu_j, \Sigma_j). \tag{12}$$

In summary, the likelyhood function (12) is used to update the parameters (9), (10), and (11) in the M-step.

In this paper, the projected data on PC subspace is segmented into mean and covariance for each group, cluster them and design accurate stochastic modeling for various operations and distributions using GMM with the EM algorithm.

### 3.2. Hotelling's $T^2$ or SPE calculation with mode decision

In order to calculate $T^2$ of multi-mode PCA, it is necessary to confirm what data set belongs to the corresponding Gaussian model. The mode of the Gaussian model is decided by the Mahalanobis distance corresponding sample data, sub-mode mean $\mu_{sub_j}$ and sub-mode covariance $\Sigma_{sub_j}$. The mode that is the smallest Mahalanobis distance among the sub-mode is selected (Fig. 2) to the mode of the $i$-th data point $x_i$. This can be formulated as follows:

$$m_{x_i} = \underset{j \in [1,2,\cdots,m]}{\arg\min} \sqrt{(P^T x_i - \mu_{sub_j})^T \Sigma_{sub_j} (P^T x_i - \mu_{sub_j})}. \tag{13}$$

With the $i$-th sample mode $m_{x_i}$, the Hotelling's $T^2$ or the square prediction error is calculated.

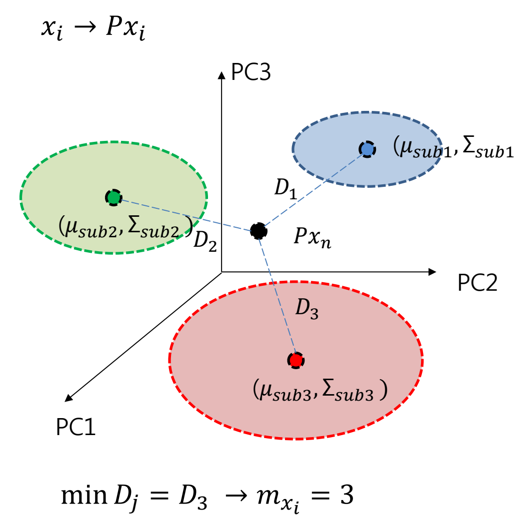


Fig. 2. Description of the sub-mode selection.

**Remark 2:** The proposed PCA modeling is achieved in the following steps. First, data are projected to PC subspace. In this step, the reduction in data dimension achieved through PCA. Second, with projected data are clustered by Gaussian mixture model. This method could eliminate the difficulty of the determination of GMM by [9]. The number of parameters in a mixture model with $m$ local Gaussian models increases by $2m(p+1)$ when a new variable is added to $p$ observed variables. Moreover, the number of parameters increases by $p^2+p$ when additional local Gaussian model is added to mixture model.

**Remark 3:** The pseudo-code of the proposed clustering algorithm is shown in Fig. 3. The input data $X \in \mathbb{R}^{n\times p}$ is clustered into the modes. This algorithm finds the parameters $\theta_m$ using Expectation and Maximization step until the convergence of the parameter. After convergence, the parameters are used to find the mode of the data samples and calculate the performance index of $T^2$ or $SPE$.

▶ Input: Given measured data $X \in \mathbb{R}^{n\times p}$ for each instance $x_i \in \mathbb{R}^p$, the correspondent mode $m \in \{0,1,\ldots,k\}$

▶ Output: the mode $m$ of the each $x_i \in \mathbb{R}^p$

▶ Algorithm:

Do:
- Initialize Parameters $\theta_m = \{\mu_j\,(9), \Sigma_j\,(10), \pi_j\,(11)\}$
- Repeat until convergence
  : Expectation step
    - compute Expectation using (8)
  : Maximization step
    - update the parameters, $\mu_j$ (9), $\Sigma_j$ (10), $\pi_j$ (11) using log likelihood function (12)
- Save parameters $\mu_j$, $\Sigma_j$, $\pi_j$ with mode $k$
- Compute the mode decision each $x_i$ using (13)

Fig. 3. Pseudo code of the proposed clustering algorithm.

### 3.3. Process flow of multi-mode PCA fault detection

Fig. 4 illustrates flowchart of the proposed fault detection method using multi-mode PCA using GMM. Similar to the conventional PCA method, model-data are collected and normalized with the mean and the standard deviation of the global PCA model and projected into the global PC subspace. The difference of the model between the conventional and the proposed is that the number of modes should be set to cluster multi-mode PCA using GMM. After designing the multi-mode PCA model, test-data are also normalized and matched the specified mode by (13). After mode of the test data point is decided, the Hotelling's $T^2$ or $SPE$ value are then calculated with the matched mode. If Hotelling's $T^2$ or $SPE$ value exceeds its limit, fault operation could be detected, if not, normal operation is decided. In the next chapter, the proposed fault detection algorithm is described using multi-mode PCA method by checking the value of $T^2$ or $SPE$ for sewage source heat pump system. The proposed scheme is compared to conventional PCA as well as other clustering methods such as $k$-means and $k$-medoids.

## 4. MONITORING RESULTS FOR SEWAGE SOURCE HEAT PUMP SYSTEM

### 4.1. Sewage source heat pump system description

The heat pump system in [6] used for validation for fault detection and monitoring is shown in Fig. 5. It receives heat from the sewage stage at the intermediate water side and supplies heat to composite side through the heat pump via phase change of R22 refrigerant. The heat pump system consists of three heat exchangers. The heat exchanger of sewage/intermediate water transfers the heat of the sewage to the intermediate water. The transferred heat is absorbed by the heat exchanger while the refrigerant vaporizes in the heat pump evaporator. The R22 refrigerant is liquefied in the heat exchanger of the condenser

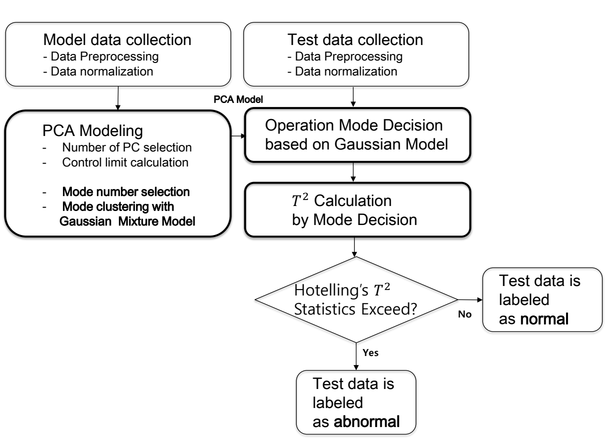


Fig. 4. Flowchart of a fault detection with the proposed multi-mode PCA.

Table 1. Measured data details.

| Data | Description | Unit |
|---|---|---|
| T1 | Evaporator water inlet temperature | °C |
| P1 | Evaporator water inlet pressure | Mpa |
| T2 | Evaporator water outlet temperature | °C |
| P2 | Evaporator water outlet pressure | Mpa |
| T3 | Condenser water inlet temperature | °C |
| P3 | Condenser water inlet temperature | Mpa |
| T4 | Condenser water outlet temperature | °C |
| P4 | Condenser water outlet temperature | Mpa |
| T5 | Compressor gas temperature | °C |
| P5 | Compressor high temperature | Mpa |
| P6 | Compressor low temperature | Mpa |
| I | Compressor current | A |

side to release heat to the composite side to provide the user with a warm heat source. Twelve measurement parameters were selected to detect the fault of the system, shown in Table 1 which shows the details of the measurements. In addition, Fig. 5 shows the measurements' location in the system.

As Fig. 5 shown, the main equipment of this system includes sewage/intermediate heat exchanger, evaporator, condenser, compressor and thermal expansion valve. Also, the system data are shown as the measured temperature (Fig. 6(a)(b)(c)), pressure (Fig. 6(d)(e)(f)), and compressor current (Fig. 6(g)). The measurement point of each variable is shown in the diagram of the sewage heat pump (Fig. 5). The sampling rate was measured once every 2 hours, and data were collected in December 2016. Except for some points, it can be confirmed that the system operated stably.

First, in the 73rd sample, the temperature of the condenser water out drops sharply from 42 °C to 4.175°C. The second operation fault point is indicated by the 253rd sample, which can be checked in condenser temperature (Fig. 6(a)) and pressure (Fig. 6(d)). The temperature drops by 4 °C in water in/out and the pressure increases by 5.8% at 0.8 MPa for water in. In the case of water out, it increases sharply by 3.97% from 0.78 MPa. For the third fault point, the gas temperature (Fig. 6(c)) sharply decreases at the 349th sample. In the following section, the proposed anomaly detection method is compared with the conventional PCA-based anomaly detection method as well as other clustering method based on PCA for the mentioned data.

### 4.2. PCA modeling with the proposed method

In order to model both the conventional PCA and the proposed multi-mode PCA, the number of PCs must be chosen. Three PCs were able to express the trend of 12 data (400 samples) in Section 4.1 by 77.3%, and reduced the dimension of data on three PCs.

The number of GMM models was determined to be three. This is because the number of GMMs with a minimum average Hotelling's $T^2$ value is most representative of the probability of a normal state when a model is created using given data. As shown in Table 2, the average value of Hotelling's $T^2$ is the smallest (1.9189) when the number of Gaussian mixture models is three.

**Remark 4:** The minimum average of $T^2$ vaule shows data distribution is close to gaussian models mean with small covariance, therefore, the gaussian model with minimum average of $T^2$ vaule shows the more fittable the statistical bounds (95% or 99% confidence rate) with given data. This could be shown in the result of the $T^2$ value (Fig. 7) with various numbers of the Gaussian mixture model. The red box regions show the fault points that is judged incorrectly using Gaussian mixture model (red- (a) model number 1, (b) model number 2,(c) model number 4 (d) model number 5).

In the three-dimensional representation of diminished data, Hotelling's $T^2$ limit when applying conventional PCA was visualized in Fig. 8a. The inside of the black ellipsoid indicates the 95 % confidence interval (1.96 $\sigma$) and the outside indicates the 99% confidence interval (2.58 $\sigma$). As one can see in Fig. 8(a), several points exceed the control limit.

Fig. 8(b) shows the result of applying the proposed multi-mode PCA. The data are clustered into three modes using GMM. The means and covariances of the three modes are the following values:

$$\mu_{sub_1} = \begin{bmatrix} -3.4052, & 0.6909, & 0.1822 \end{bmatrix},$$

$$\Sigma_{sub_1} = \begin{bmatrix} 3.1418 & 1.1626 & 0.2074 \\ 1.1626 & 2.6238 & -0.5225 \\ 0.2074 & -0.5225 & 1.9548 \end{bmatrix},$$

$$\mu_{sub_2} = \begin{bmatrix} 0.6795, & 0.2893, & -0.4674 \end{bmatrix},$$

$$\Sigma_{sub_2} = \begin{bmatrix} 3.4851 & 0.2019 & -0.0360 \\ 0.2019 & 1.7552 & 0.0781 \\ -0.0360 & 0.0781 & 1.1995 \end{bmatrix},$$

$$\mu_{sub_3} = \begin{bmatrix} 1.6383, & -0.1903, & 0.2089 \end{bmatrix},$$

$$\Sigma_{sub_3} = \begin{bmatrix} 2.0000 & 0.6850 & 0.1776 \\ 0.6850 & 1.7832 & 0.1842 \\ 0.1776 & 0.1842 & 1.3753 \end{bmatrix}.$$

For proposed multi-mode PCA, 95% confidence intervals were visualized in PCA in Fig. 8(b) like conventional PCA in Fig. 8(a). The PC scores can be confirmed that many points fall within the control limit (Fig. 8(b)) whereas several points are out of the confidence limit for conventional PCA (Fig. 8(a)). In addition, the scores can be clustered by the color of which mode belongs (mode-1: blue, mode-2: red, mode-3: green).

Furthermore, two clustering algorithms of $k$-means (Fig. 8(c)) and $k$-medoids (Fig. 8(d)) are applied to the same data. Comparing the Fig. 8b,c,d of each algorithm has some yellow highlighted points that belong to the difference the group. As a result, the mean and covariance matrix of each group are different, and the shape of the confidence rate bounds (Fig. 8(b), (c), (d) ellipsoid) are different. The fault detection index ($T^2$ and $SPE$) values of the proposed multi-mode PCA using GMM will be compared and analyzed in Section 4.3.

### 4.3. Comparison of fault detection performance

In the flowchart (Fig. 4), the value of Hotelling's $T^2$ is a measure of making decisions whether the operation is abnormal or normal. In the case of the conventional PCA (Fig. 9(a) - gray), some interval may be decided as an abnormal operation in red areas (sample interval of the 17−26th, the 82−90th, and the 267−274th) despite the normal operation. The proposed method (Fig. 9(a) - blue) detects as a fault for three samples. The first point is the condenser water out temperature drop point (73rd sample). The second one is the 253rd sample point of condenser temperature falling and pressure rising. The final point is the 349th point that is monitored by gas temperature falling.

Fig. 9(b) shows $T^2$ chart values using various clustering algorithms to compare the performance of the proposed multi-mode PCA to the other clustering algorithm based models. The $T^2$ chart of $k$-medoids (Fig. 9(b) - red) shows that the fault detection model using $k$-means clustering is not suitable as a fault detection model because most of $T^2$ values exceed control limits and are considered faulty in normal operation. The similar trend can be shown in the $k$-medoids models (Fig. 9(b) - green).

Fig. 10 shows $SPE$ values to quantify the performance of the algorithm. The proposed multi-mode PCA (Fig. 10(a) -blue) model detect the faults at the sample of the 73rd (the condenser water out temperature drop), and 349th (gas temperature falling) whereas conventional PCA (Fig. 10(a) -gray) detects incorrectly faults with several points.

Fig. 10(b) shows $SPE$ chart values using various clustering algorithms (Fig. 10(b)- blue (proposed) , red ($k$-means), green ($k$-medoids)) to compare the performance of the proposed multi-mode PCA. The results were not significant differences in SPE values except for the conventional PCA model. This results show that the residual values projected points are all small on each model of the clustering algorithm(10(b)- blue (proposed) , red ($k$-means), green ($k$-medoids)).

## 5. CONCLUSION

This paper presents the fault detection of sewage source heat pump using multi-mode principal component anal-

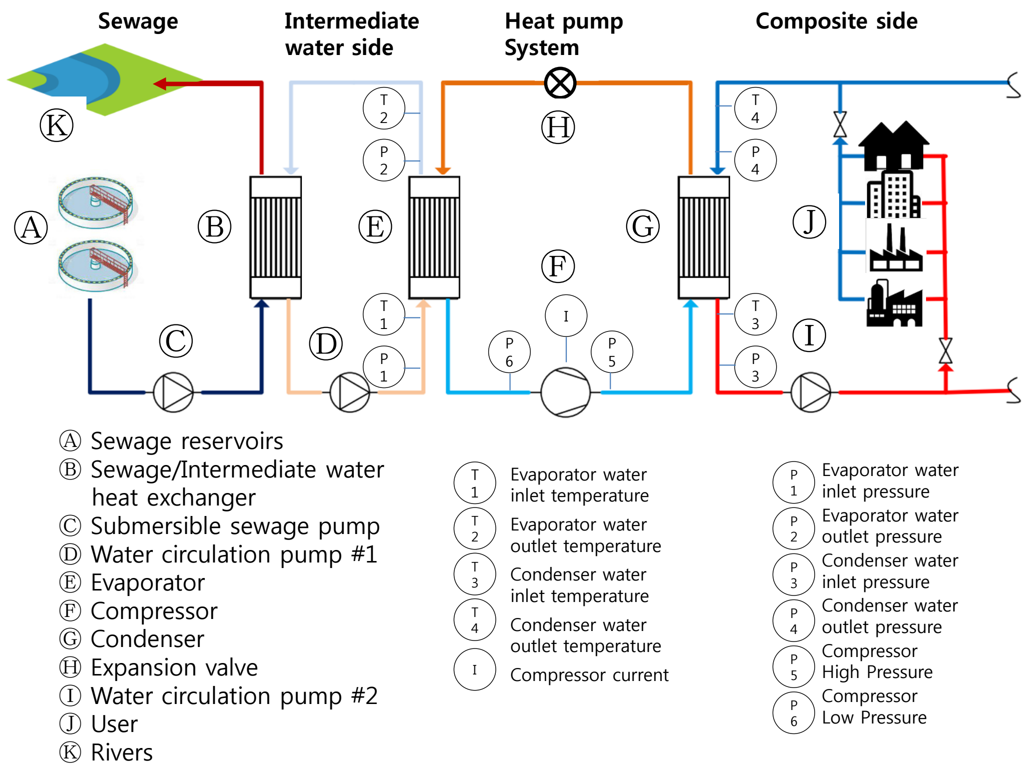


Fig. 5. The sewage source heat pump system.

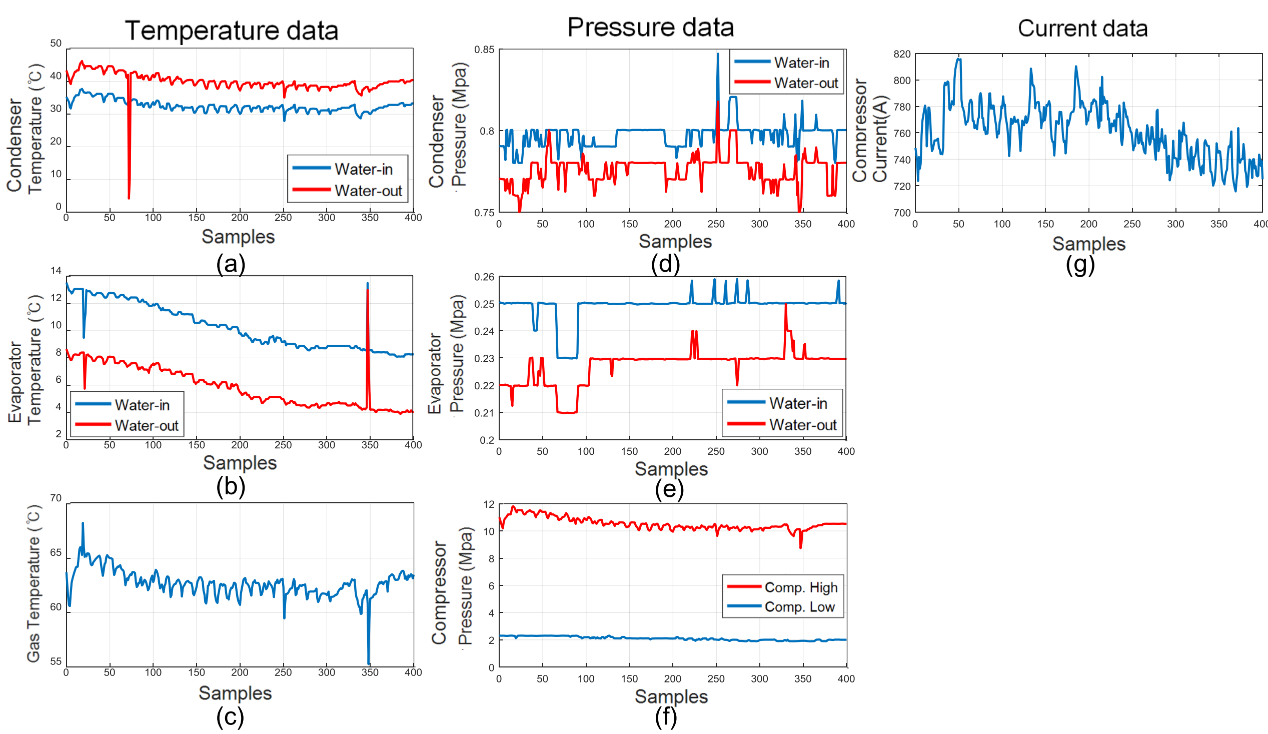


Fig. 6. Data of the heat pump system: (a) temperature data of the condenser (b) temperature data of the evaporator (c) temperature data of the gas (d) pressure data of the condenser (e) pressure data of the evaporator (f) pressure data of the compressor (g) current data of the compressor.

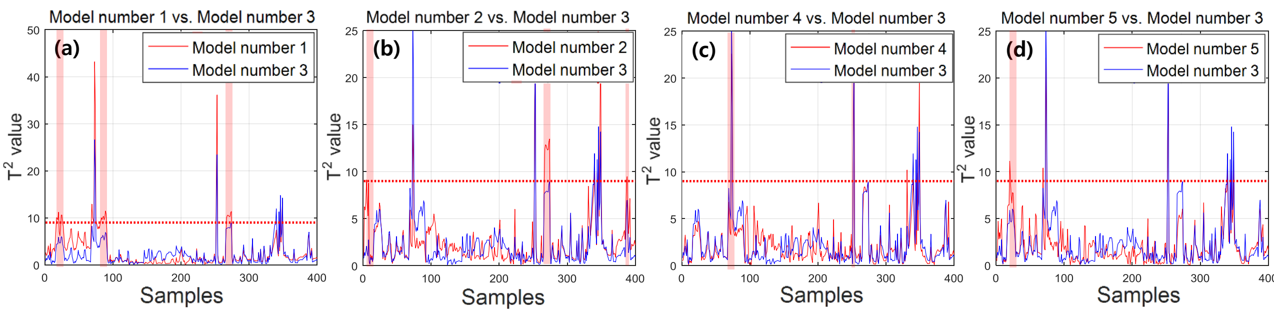


Fig. 7. $T^2$ chart according to the number of GMM: (a) model number 1 vs. 3, (b) model number 2 vs. 3,(c) model number 4 vs. 3 (d) model number 5 vs. 3.

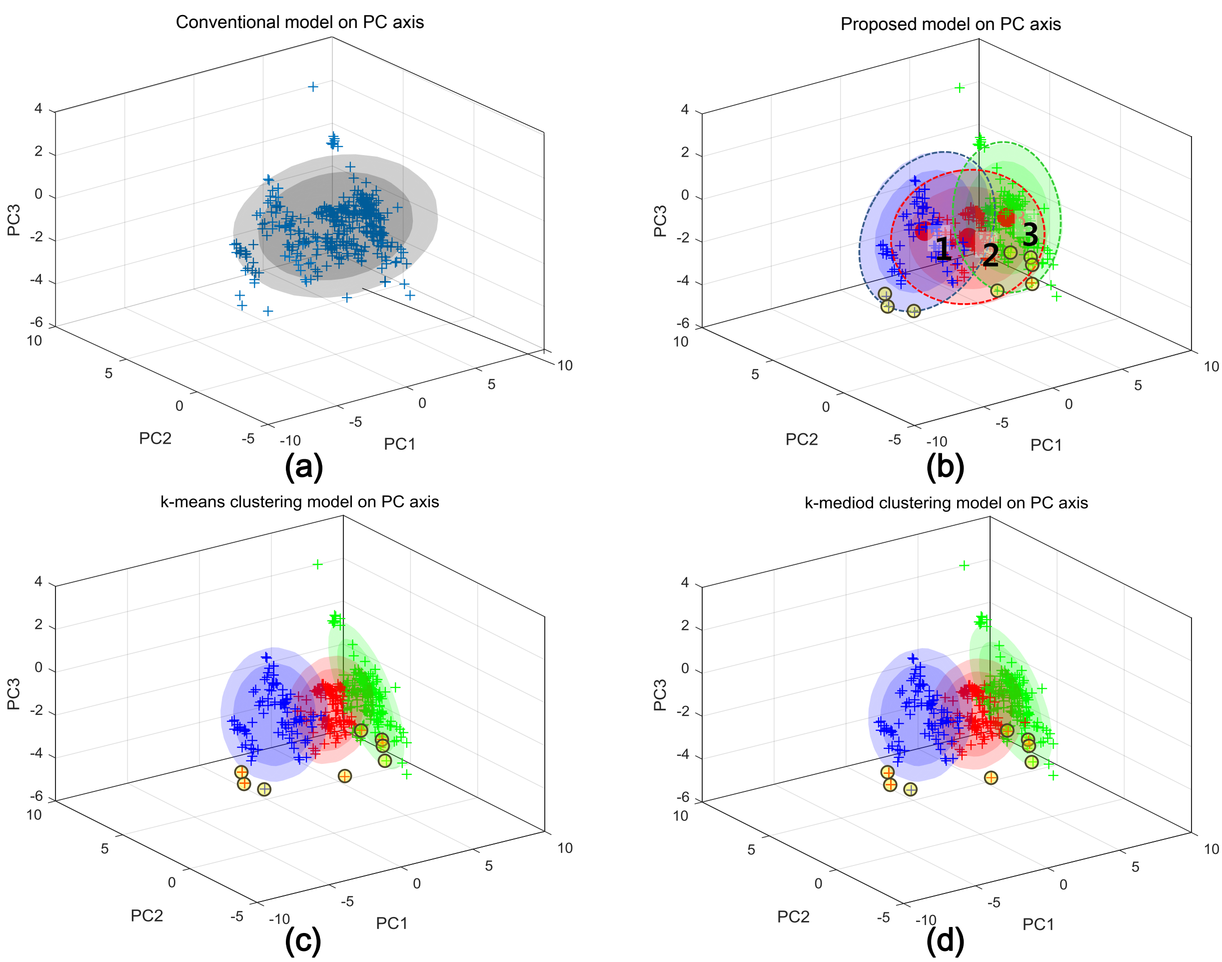


Fig. 8. A specific view of score plot of the operation data: (a) conventional PCA score (b) proposed multi-mode PCA score (c) multi-mode score using $k$-means clustering (d) multi-mode score using $k$-medoids clustering.

Table 2. Average $T^2$ according to the number of GMM

| number of the GMM | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Average $T^2$ vaule | 3.000 | 2.6623 | 1.9189 | 2.0538 | 2.2766 |

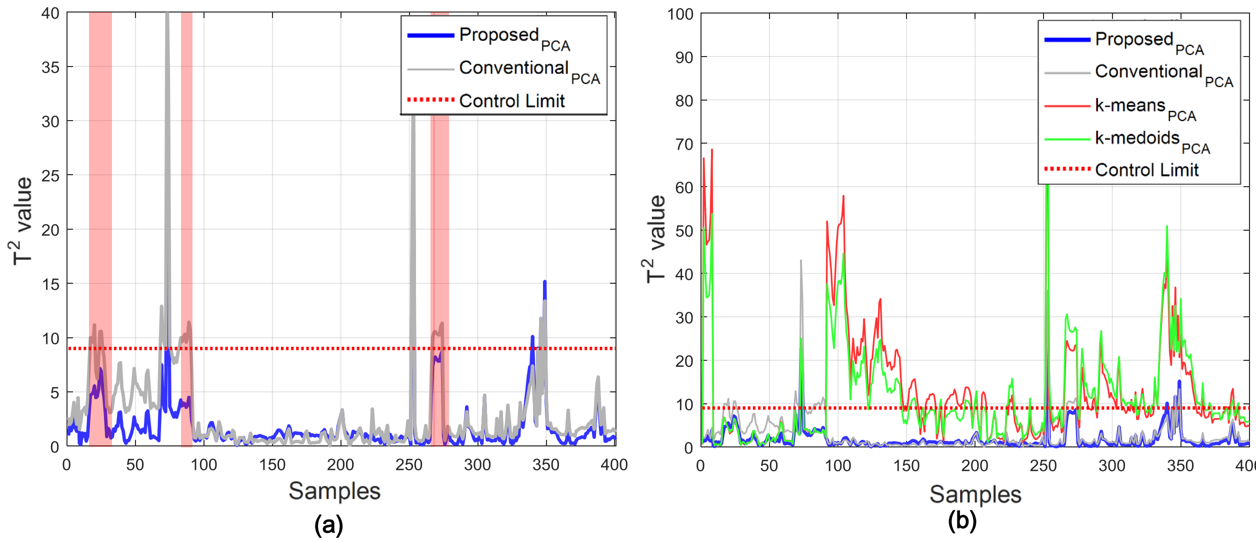


Fig. 9. $T^2$ plot : (a) the conventional PCA and the proposed multi-mode PCA (b) the proposed multi-mode PCA, conventional PCA, $k$-means PCA, and $k$-medoids PCA.

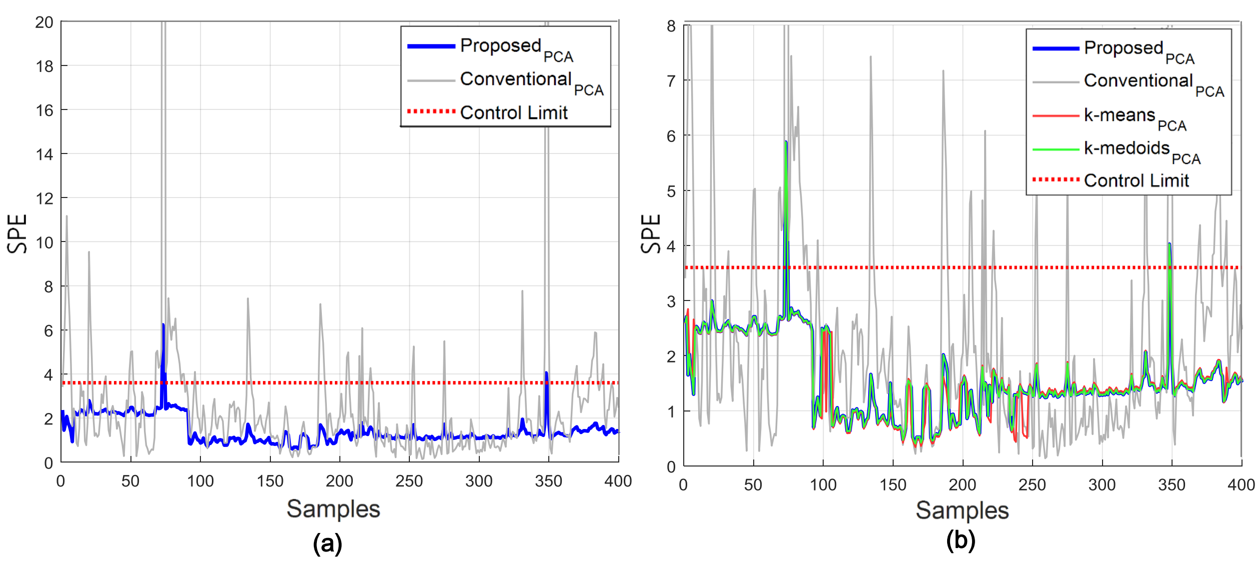


Fig. 10. *SPE* plot : (a) the conventional PCA and the proposed multi-mode PCA (b) the proposed multi-mode PCA, conventional PCA, $k$-means PCA, and $k$-medoids PCA.

ysis based on Gaussian mixture model. When the fault detection performance is evaluated with real sewage heat pump data, the Gaussian mixture integrated multi-mode model PCA results more accurate detection capability than other clustering algorithms such as $k$-means and $k$-medoids as well as conventional PCA based model. Since the faults of the heat pump system downtime may cause energy wastage, system unreliability, and shorter equipment life. Therefore, it is important to develop a fault detection algorithm of the sewage source heat pump equipment for safety and cost-effectiveness. Considering the results of the proposed methodology, it is expected that the proposed Gaussian integrated multi-mode model PCA model will play an important role in monitoring the heat pump system.

## REFERENCES


[1] N. C. Baek, U. C. Shin, and J. H. Yoon, "A study on the design and analysis of a heat pump heating system using wastewater as a heat source," *Solar Energy*, vol. 78, no.3, pp. 427-440, 2005.

[2] N. Funamizu, M. Iida, and Y. Sakakura, "Reuse of heat energy in wastewater implementation examples in Japan," *Water Sci Technol.*, vol. 43, no. 10, pp. 277-286, 2001.

[3] R. Wu and D. Sun, "Research on the soft-dirt characteristic of heat-exchanging pipe in using urban original wastewater as cooling and heating source," *J Fluid Mach.*, vol. 33, no. 12, pp. 58-61, 2005.

[4] R. Wu, Y. Xu, and D. Sun, "Technology economic analysis of heat pump temperature rising in sewage disposal process in cold climate area," *Act Energiae Solaris Sinica.*, vol. 29, no. 3, pp. 267-271, 2008.

[5] M. Abuasbehm, and H. Madani, "Fault detection and diagnosis for brine to water heat pump systems," *Proc. of 12th IEA Heat Pump Conference*, 2017.

[6] D. Yu, J. Yu, F. Sun, Y. Deng, Q. Wu, and G. Cong, "Research on the PCA-based intelligent fault detection methodology for sewage source heat pump system," *Procedia Engineering*, vol. 205, pp. 1064-1071, 2017.

[7] V. Uraikul, C. W. Chan, and P. Tontiwachwuthikul, "Artificial intelligence for monitoring and supervisory control of process systems," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 2, pp. 115-131, 2007.

[8] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, "A review of process fault detection and diagnosis part III: process history based methods," *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 327-346, 2003.

[9] S. W. Choi, J. H. Park, and I. B. Lee, "Process monitoring using Gaussian mixture model via principal component analysis and discriminant analysis," *Computers & Chemical Engineering*, vol. 28, no. 8, pp. 1377-1387, 2004.

[10] X. Xu, L. Xie, and S. Wang, "Multimode process monitoring with PCA mixture model," *Computers & Electrical Engineering*, vol. 40, no. 7, pp. 2101-2112, 2014.

[11] G. Wang, J. Liu, Y. Zhang, and Y. Li, "A novel multimode data processing method and its application in industrial process monitoring," *Journal of Chemometrics*, vol. 29, no. 2, pp.126-138, 2015.

[12] D. Ha, U. Ahmed, H. Pyun, C. J. Lee, K. H. Baek, and C.H. Han, "Multi-mode operation of principal component analysis with k-nearest neighbor algorithm to monitor compressors for liquefied natural gas mixed refrigerant processes," *Computers & Chemical Engineering*, vol. 106, no. 2, pp. 96-105, 2017.

[13] W. Deng, R. Yao, H. Zhao, X. Yang, and G. Li, "A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm," *Soft Computing*, vol. 23, no. 7, pp. 2445-2462, 2017.

[14] W. Deng, S. Zhang, H. Zhao, and X. Yang, "A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing," *IEEE Access*, vol. 6, no. 1, pp. 35042-35056, 2018.

[15] W. Deng, H. Zhao, X.Yang, J.Xiong, M. Sun, and B. Li, "Study on an improved adaptive PSO algorithm for solving multi-objective gate assignment," *Applied Soft Computing*, vol. 59, pp. 288-302, 2017.

[16] W. Deng H. Zhao, L. Zou, G. Li, X. Yang, and D. Wu, "A novel collaborative optimization algorithm in solving complex optimization problems," *Soft Computing*, vol. 21 no. 15, pp. 4387-4398, 2017.

[17] H. Zhao, R. Yao, L. Xu, Y. Yuan, G. Li, and W. Deng, "Study on a novel fault damage degree identification method using high-order differential mathematical morphology gradient spectrum entropy," *Entropy*, vol. 20, no. 9, Article no. 682, 2018.

[18] H. Zhao, M. Sun, W. Deng, and X. Yang, "A new feature extraction method based on EEMD and multi-scale fuzzy entropy for motor bearing," *Entropy*, vol. 19, no. 1, Article no. 14, 2017.

[19] W. Deng, J. J. Xu, and H. M. Zhao, "An improved ant colony optimization algorithm based on hybrid strategies for scheduling problem," *IEEE ACCESS*, vol. 7, pp. 20281-20292, 2019.

[20] A. Beghi, R. Brignoli, L. Cecchinato, G. Menegazzo, M. Rampazzo, and F. Simmini, "Data-driven fault detection and diagnosis for HVAC water chillers," *Control Engineering Practice*, vol. 53, pp. 79-91, 2016.

[21] Y. P. Hu, H. X. Chen, G. N. Li, H .R. Li, R. J. Xu, and J. Li, "A statistical training data cleaning strategy for the PCA-based chiller sensor fault detection, diagnosis and data reconstruction method," *Energy and Buildings*, vol. 112, pp. 270-278, 2016.

[22] N. Cotrufo, and R. Zmeureanu, "PCA-based method of soft fault detection and identification for the ongoing commissioning of chillers," *Energy and Buildings*, vol. 130, pp. 443-452, 2016.

[23] G. N. Li, Y. P. Hu, H. X. Chen, L. M. Shen, H. R. Li, M. Hu, J. Y. Liu, and K. Sun, "An improved fault detection method for incipient centrifugal chiller faults using the PCA-R-SVDD algorithm," *Energy and Buildings*, vol. 116, pp. 104-113, 2016.

**Young-Jun Yoo** received the B.S. degree from POSTECH in 2009, and the Ph.D. degree Electronic and Electrical Engineering from POSTECH in 2014. His research interests include disturbance observer, time delay systems, robot manipulator control and fault detection of the mechanical system.