Received September 10, 2017, accepted October 16, 2017, date of publication October 20, 2017,
date of current version November 14, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2764518

Nonlinear Multimode Industrial Process Fault Detection Using Modified Kernel Principal Component Analysis

XIAOGANG DENG, NA ZHONG, AND LEI WANG
College of Information and Control Engineering, China University of Petroleum, Qingdao 266580, China

Corresponding author: Xiaogang Deng (dengxiaogang@upc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61403418, Grant 21606256, and Grant 61273160, in part by the Natural Science Foundation of Shandong Province, China, under Grant ZR2014FL016, Grant ZR2016BQ14, and Grant ZR2016FQ21, and in part by the Fundamental Research Funds for the Central Universities, China, under Grant 17CX02054.

ABSTRACT Kernel principal component analysis (KPCA) has been a state-of-the-art nonlinear process monitoring method. However, KPCA assumes the single operation mode while the real industrial processes often run under multiple operation conditions. In order to monitor the nonlinear multimode processes effectively, this paper proposes a modified KPCA method assisted by the local statistical analysis, referred to as local statistics KPCA (LSKPCA). In the proposed method, two kinds of strategies, including local probability density estimation and statistics pattern analysis, are integrated to improve the traditional KPCA method. To handle the multimode characteristic of industrial processes, local probability density estimation is developed to transform the monitored variables into their probability density values, which follow the unimodal data distribution. For further extracting the statistical information among the process data, statistics pattern analysis technique is applied to capture various orders of statistics, including one-order, second-order, and high-order ones, which constitute the statistics pattern matrix of the monitored data. Furthermore, KPCA modeling is performed on the statistics pattern matrix. The simulations on one numerical example and the continuous stirred tank reactor system demonstrate that the proposed LSKPCA method has the superior fault detection performance compared with the conventional KPCA method.

INDEX TERMS Nonlinear process, multimode process, kernel principal component analysis, local probability density estimation, statistics pattern analysis.

I. INTRODUCTION

As modern industrial systems are becoming large-scale and complicated, real-time fault detection and diagnosis technologies are of vital importance to assure process safety and prevent quality degradation. The commonly used fault detection and diagnosis methods can be categorized into three groups: model-based methods, knowledge-based methods and data-based methods [1]–[7]. As for the large-scale industrial processes, there are abundant process data available in the industrial databases because of the application of distributed computer control systems. Therefore, data-based methods have shown their advantages over the other methods. The typical data-based fault detection and diagnosis methods include principal component analysis (PCA), partial least squares (PLS), independent component analysis (ICA) and canonical variate analysis (CVA), etc. [8]–[11]. Among these methods, PCA is one of the most popular data mining methods and has been applied in many different process monitoring cases [12]–[16].

PCA can extract the low-dimensional uncorrelated data feature information from the high-dimensional process data by orthogonal linear transformation. Based on the extracted data features, two monitoring statistics T^2 and SPE are usually constructed for convenient fault detection. However, traditional PCA method is in essence one linear transformation technique, which may not provide satisfactory performance for nonlinear process monitoring cases. In fact, many industrial systems, such as chemical reactors and biological processes, are often with particular nonlinear characteristics. To deal with the nonlinear system monitoring problem, many nonlinear PCA versions have been proposed, such as principal curve PCA [17], neural network based PCA [18], and kernel PCA (KPCA) [19], [20]. By utilizing the kernel function to solve nonlinear optimization effectively, KPCA has been one state-of-the-art method in nonlinear process monitoring field. Lee *et al.* [20] first proposed the KPCA based fault detection method for continuous process monitoring and then Lee *et al.* [21] developed a multiway KPCA (MKPCA) method for batch process monitoring. Considering the process dynamic property, Choi *et al.* [22] presented a dynamic KPCA method which monitors the time-lagged vectors instead of the original process vectors. Furthermore, Jia *et al.* [23] built a batch dynamic KPCA (BDKPCA) based process monitoring method by considering the auto-correlation and cross-correlation of process variables. In order to detect faults in nonlinear plant-wide processes, Jiang *et al.* [24] divided the measured variables into sub-blocks by performing mutual information-spectral clustering, then established multi-block KPCA monitoring model. Considering that KPCA only captures the global data structure but ignores the local structure information, Deng *et al.* [25] imposed local structure preserving on the KPCA optimization objective and proposed a modified KPCA algorithm, referred to as the local KPCA (LKPCA). For monitoring nonlinear processes with outliers, Zhang *et al.* [26] combined sliding median filtering technology to improve the KPCA method. Many other KPCA related studies can be seen in the literature [27]–[33].

Traditional PCA and KPCA methods assume one single normal operation mode, while real industrial processes often run under multiple operation modes due to raw material fluctuations, seasonal variations and market demand changes. In view of multimode process monitoring problems, some multimode monitoring methods have been discussed. Zhao *et al.* [34] firstly proposed a multiple PCA model based multimode process monitoring method. Natarajan and Srinivasan [35] used a k-means clustering strategy to classify the operating data and then trained the PCA model for each data cluster. Xu *et al.* [36] proposed a PCA mixture model by integrating Gaussian mixture model, in which each Gaussian component describes an individual operation mode. Further considering the non-Gaussianity of process data, Ge and Song [37] applied a two-step feature extraction technique ICA-PCA to build a statistical model for each mode and combined all monitoring results by Bayesian inference strategy. All the above methods adopt the same multiple modeling strategy, which divides the multimode normal operation data into multiple groups and builds the individual PCA model for each group. The disadvantage of these methods lies in the requirement of enough prior knowledge on mode partition. To avoid this disadvantage, another kind of multimode process monitoring method based on the local learning technology is developed. This kind of method deals with multimode data by mining the relationship between the monitored data sample and its local neighborhood data points. He and Wang [38] proposed a principal component based KNN rule, which applies PCA to reduce data dimension and then computes the KNN distance for fault detection. Deng and Tian [39] developed two local neighborhood similarity factors, including the PCA similarity factor and the distance similarity factor, to monitor multimode process changes. More recently, Ma *et al.* [40] proposed a novel local neighborhood standardization based principal component analysis (LNS-PCA) method, which preprocesses the multimode data samples in their local neighborhood domain so that the multimode property is cancelled. Wang *et al.* [41] also proposed a modified neighborhood standardization based PCA to address the multimode process monitoring problem.

To sum up, KPCA has achieved great success in nonlinear process monitoring field but it does not consider the multimode characteristic of industrial process data, while the multimode process monitoring methods mentioned above are all based on the linear PCA methods and do not consider the nonlinear process characteristics. At the same time, the present KPCA methods develop statistical models directly based on the original process variables, which only analyze the two-order statistical information, i.e., variance and covariance. However, due to process nonlinearity, various order of statistics including high-order statistics are very useful to monitor process status.

Motivated by the above analysis, this paper is to propose an improved KPCA method, named local statistics principal component analysis (LSKPCA), for nonlinear multimode process fault detection. The contribution of the proposed method includes the following two aspects: (1) By using local probability density estimation to preprocess the operating data, the multimode characteristic can be removed without any dependence of the mode prior knowledge. (2) For better process monitoring performance, statistics pattern analysis (SPA) is applied to capture the internal high-order statistical information in the local probability density variables.

The remainder of the paper is arranged as follows. The traditional KPCA algorithm is briefly overviewed in Section II. Then the proposed LSKPCA method is described in details in Section III. Section IV formulates the fault detection procedure based on LSKPCA. In Section V two case studies on one nonlinear numerical system and the benchmark continuous stirred tank reactor (CSTR) system are used to demonstrate the validity of the proposed method. Finally, conclusions are drawn in Section VI.

II. KERNEL PRINCIPAL COMPONENT ANALYSIS

KPCA is a well-known nonlinear PCA algorithm and has been widely used in many different fields [20], [29], [42], [43]. Its main idea is to firstly project the original input data to a high-dimensional feature space by a nonlinear transformation function and then perform PCA decomposition in the feature space to extract the nonlinear principal components. The algorithm details are listed as follows.

Given a training data matrix $\boldsymbol{X} \in \mathbb{R}^{n\times m}$, where n is the sample number, m is the variable number and $\mathbb{R}$ represents a set of real numbers, a nonlinear function $\Phi(\cdot)$ is assumed to map the original data $\boldsymbol{X}$ onto the high-dimensional feature space $\mathcal{F}$, where the mapped data is denoted as $\Phi(\boldsymbol{X}) \in \mathbb{R}^n \times \mathcal{F}$.

Suppose that $\Phi(\boldsymbol{X})$ has been mean-centered, the linear PCA decomposition is executed as

$$\Phi(\boldsymbol{X}) = \sum_{i=1}^{k} \boldsymbol{t}_i \boldsymbol{p}_i^T + \boldsymbol{E}, \quad (1)$$

where $\boldsymbol{t}_i \in \mathbb{R}^n$ is the score vector, also called principal component (PC) vector, $\boldsymbol{p}_i \in \mathcal{F}$ is the loading vector representing the PC direction, k is the number of the retained PCs in KPCA model, and $\boldsymbol{E}$ is the residual matrix.

The goal of PCA decomposition is to seek the PC direction $\boldsymbol{p}_i$, which can be calculated by solving the eigenvalue decomposition as

$$\frac{1}{n-1}\Phi^T(\boldsymbol{X})\Phi(\boldsymbol{X})\boldsymbol{p}_i = \lambda_i \boldsymbol{p}_i, \quad (2)$$

where λ_i is the i-th eigenvalue corresponding to the eigenvector $\boldsymbol{p}_i$. It is known that the eigenvector $\boldsymbol{p}_i$ can be expressed by the linear expansion of the training data matrix $\Phi(\boldsymbol{X})$ as [19] [20]

$$\boldsymbol{p}_i = \Phi^T(\boldsymbol{X})\boldsymbol{v}_i = \sum_{j=1}^{n} v_{ij}\Phi(\boldsymbol{x}_j), \quad (3)$$

where $\boldsymbol{v}_i \in \mathbb{R}^n$ is the coefficient vector with its j-th element as v_{ij}, and $\boldsymbol{x}_j$ is the j-th sample of the training data matrix $\boldsymbol{X}$.

Combining Eqs. (2) and (3) will lead to the following equation as [19]

$$\frac{1}{n-1}\Phi(\boldsymbol{X})\Phi^T(\boldsymbol{X})\Phi(\boldsymbol{X})\Phi^T(\boldsymbol{X})\boldsymbol{v}_i = \lambda_i\Phi(\boldsymbol{X})\Phi^T(\boldsymbol{X})\boldsymbol{v}_i, \quad (4)$$

In order to avoid the explicit use of the nonlinear mapping function $\Phi(\cdot)$, the well-known kernel trick is introduced [19]. By defining a kernel matrix $\boldsymbol{K} = \Phi(\boldsymbol{X})\Phi^T(\boldsymbol{X})$, Eq. (4) is reformulated as

$$\frac{1}{n-1}\boldsymbol{K}\boldsymbol{K}\boldsymbol{v}_i = \lambda_i \boldsymbol{K}\boldsymbol{v}_i, \quad (5)$$

whose solutions are obtained by solving the eigenvalue decomposition as

$$\frac{1}{n-1}\boldsymbol{K}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i, \quad (6)$$

where the (i,j)-th element of the kernel matrix $\boldsymbol{K}$ is computed by $\boldsymbol{K}_{ij} = ker(\boldsymbol{x}_i, \boldsymbol{x}_j) = \Phi^T(\boldsymbol{x}_i)\Phi(\boldsymbol{x}_j)$. $ker(\cdot,\cdot)$ represents the kernel function which transforms the inner product of $\Phi(\boldsymbol{x}_i)$ and $\Phi(\boldsymbol{x}_j)$ into the kernel computation of $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. Some typical kernel functions are the Gaussian kernel function, the sigmoid kernel function and the polynomial kernel function [19], [20]. In this paper, we adopt the most commonly used Gaussian kernel function $ker(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp(-||\boldsymbol{x}_i - \boldsymbol{x}_j||^2/c)$, where c is the kernel width parameter.

By solving Eq. (6), we can obtain a series of eigenvectors $\boldsymbol{v}_i$ ordered by their eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. For a testing data vector $\boldsymbol{x}_t \in \mathbb{R}^m$, its i-th nonlinear principal component t_i is computed by [20]

$$t_i = \Phi^T(\boldsymbol{x}_t)\boldsymbol{p}_i = \boldsymbol{K}_t^T \boldsymbol{v}_i, \quad (7)$$

where $\boldsymbol{K}_t = \Phi(\boldsymbol{X})\Phi(\boldsymbol{x}_t) \in (R)^n$ is the kernel vector of the testing vector $\boldsymbol{x}_t$.

Based on the kernel principal components (KPCs) from Eq. (7), two monitoring statistics T^2 and SPE are constructed to detect process faults. The T^2 statistic is built to monitor the changes of the PC subspace, while the SPE statistic is developed to monitor the changes of the residual subspace. Their definitions are given by [20], [21]

$$T^2 = \bar{\boldsymbol{t}}_k^T \Lambda_k^{-1} \bar{\boldsymbol{t}}_k, \quad (8)$$

$$SPE = \bar{\boldsymbol{t}}^T \bar{\boldsymbol{t}} - \bar{\boldsymbol{t}}_k^T \bar{\boldsymbol{t}}_k, \quad (9)$$

where $\bar{\boldsymbol{t}} = [t_1,\ t_2,\ \cdots,\ t_n]$ while $\bar{\boldsymbol{t}}_k = [t_1,\ t_2,\ \cdots,\ t_k]$, Λ_k is a $k \times k$ diagonal matrix with its diagonal elements are the eigenvalues $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_k$. The retained KPCs number k is determined by the average eigenvalue approach [20], which selects the retained KPCs with the larger eigenvalues than the average eigenvalue. The confidence limits of the T^2 and SPE statistics can be computed by two ways. One way is based on the pre-assumed data distribution [20], while another way is to apply the non-parametric density estimation technologies such as the kernel density estimation (KDE) [44]. In this paper, we adopt the KDE method to obtain the confidence limits.

III. MODIFIED KPCA METHOD USING LOCAL STATISTICS INFORMATION

Traditional KPCA based process monitoring method assumes that the process data obey the unimodal distribution. However, industrial processes often include several operating modes due to the changes of raw materials, production environments or market demands. If KPCA is directly applied to the multimode process, it may not provide the best monitoring performance. In order to deal with the multimode nonlinear process monitoring problem, this paper is to propose an improved KPCA method, referred to as local statistics KPCA (LSKPCA). In this proposed method, local probability density estimation is firstly applied to convert the original multimode data into the unimodal probability density variables. Then for better nonlinear process monitoring, statistics pattern analysis is used to compute the low-order and high-order statistics of the monitored probability density variables. Lastly in the statistics space, KPCA statistical model is developed for process monitoring. The proposed LSKPCA method has the two-fold advantages including canceling the single mode assumption and providing deeper statistical information mining.

A. MOTIVATION ANALYSIS

To illustrate why the proposed method is developed, a simple numerical example with two monitored variables x_1, x_2 is designed as follows.

$$x_1 = s_1 + e_1,$$

$$x_2 = s_1^3 + s_2 + e_2, \quad (10)$$

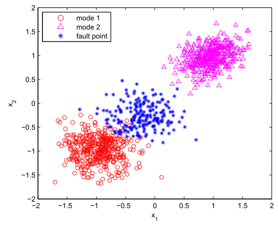

FIGURE 1. The scatter plot of the normal and fault data of the numerical system.

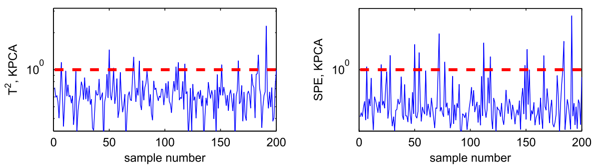

FIGURE 2. The KPCA monitoring charts on the fault data of the numerical system.

where $e_i (i = 1, 2)$ is the zero-mean Gaussian noise with a standard deviation of 0.01, while $s_i (i = 1, 2)$ is the data source variable involving two different operation modes as

$$\begin{aligned} mode\ 1: \ & s_1 \sim N(0, 0.08^2), s_2 \sim N(0, 0.15^2), \\ mode\ 2: \ & s_1 \sim N(0.5, 0.07^2), s_2 \sim N(1, 0.13^2). \end{aligned} \tag{11}$$

By Eq. (10), one two-dimensional normal dataset containing two modes are generated. Each mode involves 400 samples and a total of 800 samples comprise the training dataset. When the system is running under mode 1, another dataset with 200 samples are simulated as the fault dataset according to the following expressions

$$\begin{aligned} x_1^f &= s_1 + f_1 + e_1, \\ x_2^f &= s_1^3 + s_2 + f_2 + e_2, \end{aligned} \tag{12}$$

where the f_1 is a bias of 0.1 while the f_2 is a bias of 0.4.

When the KPCA model is built, the normal and fault data are scaled by the mean and variance of normal data. The scaled normal and fault samples are plotted in Fig. 1. Obviously, the normal data cover two separated areas while the fault data samples lie between these areas. This may lead to the difficulty of detecting the fault samples because the whole normal area encompasses the fault data. The KPCA based monitoring results are shown in Fig. 2, where the monitoring statistics are plotted with solid line while the 95% confidence limits, used as detection thresholds, are plotted with dashed line. In this paper, all monitoring charts are normalized by the corresponding detection thresholds. From the KPCA monitoring charts, the two monitoring statistics are always below the confidence limits. This means that the tested fault is totally miss-alarmed. So the conventional KPCA method performs rather poorly, which can not separate the fault samples from the normal data because of its unimodal data assumption. This motivates our study in the following sections.

B. LOCAL PROBABILITY DENSITY ESTIMATION

In order to improve KPCA method for multimode process monitoring, this paper is to propose a local KPCA method. Different from the traditional KPCA model which monitors the original process variables, the proposed local KPCA method monitors the local probability density values of the original variables. In the proposed method, a novel data processing strategy, called local probability density estimation, is adopted to transform the multimode data to the unimodal data, and then KPCA statistical model is developed based the unimodal data. The idea of the local probability density estimation is formulated as follows.

According to probability theory and statistics [45], probability density function (PDF) describes the probability distribution of a random variable, which is able to exhibit the dynamic behavior of data samples. In some cases, the random variables are supposed to follow some known PDFs, such as Gaussian distribution function, uniform distribution function. However, in many cases, the PDF of a random variable is unknown and the nonparametric density estimation methods are applied to estimate the data distribution. The commonly used nonparametric probability density estimation tool is the kernel density estimator (KDE), also known as the Parzen window method [45], which applies the kernel function to approximate the probability density curve. This method has been widely used in image processing, signal analysis and other fields [46]–[48].

For a continuous variable x, its n random samples are denoted by $x_1, x_2, x_3, \cdots, x_n$. Then the probability density function of variable x is estimated by the KDE method as

$$p(x) = \frac{1}{n} \sum_{j=1}^{n} exp(-\frac{||x - x_j||^2}{\rho}), \tag{13}$$

where ρ is the window width parameter.

When KDE is applied, all the training data samples are used to compute the overall probability density values. However, this paper is concerned about the changes of the local probability density. To achieve this goal, we propose a local KDE (LKDE) method. Different from KDE, LKDE computes the probability density values of each data sample only based on its local neighborhood samples. For one given sample x, the local probability density $lp(x)$ is estimated by

$$lp(x) = \frac{1}{L} \sum_{x_j \in NN(x)} exp(-\frac{||x - x_j||^2}{\rho(x_j)}), \tag{14}$$

where L is the designated number of the local neighborhood samples, $NN(x)$ represents the dataset consisting of the L nearest neighbors of the sample x, expressed by

$$NN(x) = \{x^1, x^2, \cdots, x^L\}, \tag{15}$$

where x^j is the j-th nearest neighbor of x, and $\rho(x_j)$ is the kernel width parameter formulated by

$$\rho(x_j) = d(x_j, x_j^L), \tag{16}$$

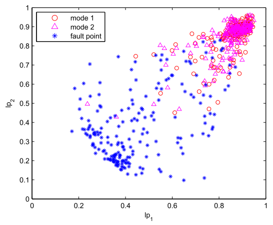

FIGURE 3. The scatter plot of the local probability density data of the numerical system.

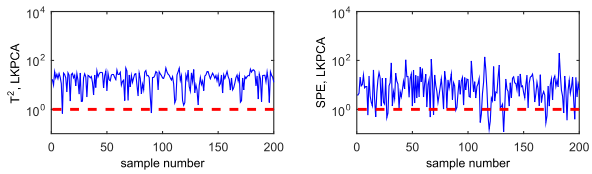

FIGURE 4. The LKPCA monitoring charts on the fault data of the numerical system.

where $d(x_j, x_j^L)$ represents the Euclidean distance between x_j and its L-th neigbor x_j^L.

Given a multimode training dataset $\boldsymbol{X} \in \mathbb{R}^{n\times m}$, the local neighbors of each vector are found by comparing its distances with the other vectors in the training dataset. Then LKDE can be used to map the original multimode data matrix $\boldsymbol{X}$ into the local probability density data matrix $\boldsymbol{LP} \in \mathbb{R}^{n\times m}$. To exhibit the ability of LKDE, the motivating example in Section III-A is further analyzed. The local density variables $lp_i(i = 1, 2)$, corresponding to the variables $x_i(i = 1, 2)$, are plotted in Fig. 3. It can be seen that the normal data under the two modes are grouped together while the fault data samples are separated basically. This is because LKDE focuses on the local data distribution in view of its neigbors so that the multimode characteristic of the training data is cancelled.

With the local probability density variables as the monitored variables, the modified KPCA method is called local KPCA (LKPCA). The LKPCA monitoring charts are shown in Fig. 4, where it is observed that for most of the fault samples, their monitoring statistics exceed the confidence limits. Comparing Figs. 3 and 4, it is clear that LKPCA gives a better monitoring result than the KPCA method. Therefore, the LKPCA method addresses the multimode data distribution problem successfully by transforming the original process variables into the corresponding local density variables. However, it should be pointed out that the separation between normal data and fault data is not very perfect. By Fig. 3, some fault samples still fall into the normal data distribution area. The similar results are found in Fig. 4, where the monitoring statistics are very close to the confidence limits so that fault is not isolated very clearly.

C. STATISTICS PATTERN ANALYSIS

The statistical models in KPCA and LKPCA only involve the low-order statistics, i.e. variance and covariance, of the monitored variables, but omit the utilization of high-order statistics. In some cases, the fault will not affect the local density variables clearly but leads to the obvious change of the local density variables' statistics, especially the high-order statistics. In order to make better use of the statistical information of the process data, this paper integrates statistics pattern analysis (SPA) with the LKPCA, which results in a novel nonlinear multimode process monitoring method, called local statistics KPCA (LSKPCA).

According to the SPA theory proposed by He and Wang [49], [50], various orders of variable statistics, including first order, second order and high-order statistics, can provide rich information for process monitoring and fault detection. Some researchers have demonstrated the effectiveness of SPA in process monitoring. He and Wang [49] firstly applied SPA in semiconductor process monitoring. Zhang *et al.* [51], Deng and Tian [52] discussed the SPA based fault variable identification and fault pattern recognition, respectively. These studies do not involve the nonlinear multimode process monitoring problem. This paper is to integrate SPA with LKPCA for monitoring nonlinear multimode processes. The details of SPA procedure in LSKPCA are given as follows.

Based on the local probability density dataset $\boldsymbol{LP} \in \mathbb{R}^{n\times m}$, a data window $\boldsymbol{LP}_q \in \mathbb{R}^{w\times m}$ at the time instant q is picked out, expressed by

$$\begin{aligned}\boldsymbol{LP}_q &= [\boldsymbol{lp}_1\ \boldsymbol{lp}_2\ \cdots\ \boldsymbol{lp}_m] \\ &= \begin{bmatrix} lp_1(q-w+1) & lp_2(q-w+1) & \cdots & lp_m(q-w+1) \\ lp_1(q-w+2) & lp_2(q-w+2) & \cdots & lp_m(q-w+2) \\ \vdots & \vdots & \vdots & \vdots \\ lp_1(q) & lp_2(q) & \cdots & lp_m(q) \end{bmatrix},\end{aligned} \tag{17}$$

where $lp_i(j)$ denotes the j-th sample of the i-th variable, and w is the window width.

For the data window $\boldsymbol{LP}_q$, SPA is applied to construct the statistics pattern(SP) vector as the monitored vector. A SP vector $\boldsymbol{ls}_q$ is composed of various statistics of $\boldsymbol{LP}_q$, shown as

$$\boldsymbol{ls}_q = [\boldsymbol{\mu}|\Upsilon|\Xi], \tag{18}$$

where $\boldsymbol{\mu}$ represents the first-order statistics consisting of variable means $\mu_i(1 \leq i \leq m)$, defined by

$$\mu_i(q) = \frac{1}{w}\sum_{j=0}^{w-1} lp_i(q-j). \tag{19}$$

Υ is the second-order statistic set made up of variance (v_i), correlation $(r_{i,j})$, autocorrelation (r_i^d) and cross correlation $(r_{i,j}^d)$, which are expressed by

$$v_i = \frac{1}{w}\sum_{k=0}^{w-1}[lp_i(q-k) - \mu_i]^2, \tag{20}$$

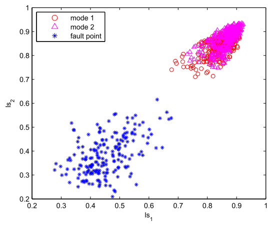

FIGURE 5. The scatter plot of the local statistics data of the numerical system.

$$r_{i,j} = \frac{1}{w\sqrt{v_i v_j}} \sum_{k=0}^{w-1} [lp_i(q-k) - \mu_i][lp_j(q-k) - \mu_j], \quad (21)$$

$$r_i^d = \frac{1}{(w-d)v_i} \sum_{k=d}^{w-1} [lp_i(q-k) - \mu_i][lp_j(q-k+d) - \mu_j], \quad (22)$$

$$r_{i,j}^d = \frac{1}{(w-d)\sqrt{v_i v_j}} \sum_{k=d}^{w-1} [lp_i(q-k) - \mu_i] \times [lp_j(q+d-k) - \mu_j], \quad (23)$$

where d is the time lag number. As the autocorrelation and cross correlation utilize the process lag information, they are able to provide the measurement of the process dynamic relationship. Ξ denotes the high-order statistics, consisting of skewness (γ_i) and kurtosis (κ_i) as

$$\gamma_i = \frac{\frac{1}{w}\sum_{k=0}^{w-1}[lp_i(q-k) - \mu_i]^3}{\left\{\frac{1}{w}\sum_{k=0}^{w-1}[lp_i(q-k) - \mu_i]^2\right\}^{3/2}}, \quad (24)$$

$$\kappa_i = \frac{\frac{1}{w}\sum_{k=0}^{w-1}[lp_i(q-k) - \mu_i]^4}{\left\{\frac{1}{w}\sum_{k=0}^{w-1}[lp_i(q-k) - \mu_i]^2\right\}^{2}} - 3. \quad (25)$$

It should be noted that not all the statistics are used in process monitoring and only these significant statistics are necessary to construct the SP vector. By referring to the present study [50], the correlation statistic $r_{i,j}$ is selected only if $|r_{i,j}| > 0.3$ for more than 70% of the training SPs. For autocorrelation and cross correlation coefficients, the selection requirement is $|r_i^d| > 0.5$ or $|r_{i,j}^d| > 0.5$ for more than 90% of the training SP vectors.

By combining Eqs. (17) to (25), the data window $\boldsymbol{LP}_q$ is transformed to a SP vector $\boldsymbol{ls}_q$. Furthermore, the moving window technique is applied on the local probability density dataset $\boldsymbol{LP}$, which is transformed to a statistic matrix $\boldsymbol{LS}$. Furtherly, KPCA modeling on the matrix $\boldsymbol{LS}$ leads to a LSKPCA statistical model.

To display the performance of statistical information mining, the motivating example is further analyzed. The first two local statistics are plotted in Fig. 5. Compared to Fig.3, the fault samples in Fig.5 are significantly separated from the normal samples. The introduction of statistics pattern analysis will benefit the KPCA based fault detection. This can be verified by the monitoring results in Fig. 6. By this figure, it is obvious that both the T^2 and SPE statistics exceed the confidence limits evidently and there is no missing alarming samples. These analyses prove that the LSKPCA has the capability of providing much better fault detection results than the traditional KPCA method.

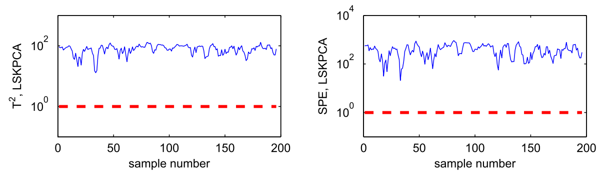

FIGURE 6. The LSKPCA monitoring charts on the fault data of the numerical system.

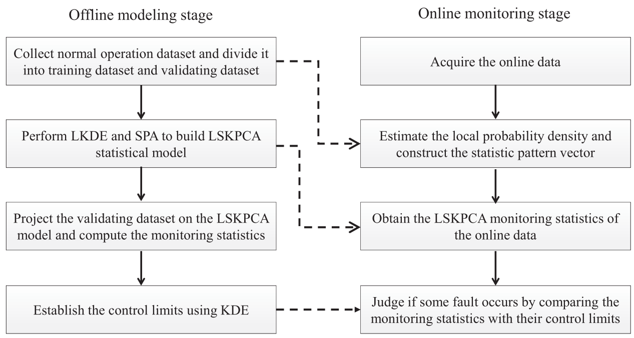

FIGURE 7. The flow chart of LSKPCA-based fault detection.

IV. FAULT DETECTION PROCEDURE BASED ON LSKPCA

The LSKPCA based process monitoring procedure includes two stages: offline modeling stage and online detection stage. In the offline modeling stage, the normal dataset is collected and LSKPCA statistical model is developed, while in the online monitoring stage, new data vector is acquired and the corresponding LSKPCA monitoring statistics are computed for fault detection. The whole fault detection procedure, shown in Fig. 7, is formulated as follows.

Offline Modeling Stage:

1) Collect the normal operation data and divide them into two datasets: training dataset and validating dataset. Normalize the two datasets with the mean and variance of the training dataset.
2) Perform the LKDE on the training dataset to obtain the local probability density variables. Apply the moving window technique on the local probability density dataset and build the local statistic matrix by computing the statistic pattern for each data window.
3) Build LSKPCA statistical model by applying KPCA to the local statistic matrix.
4) For the validating dataset, construct its local statistic matrix and project it on the LSKPCA model.
5) Compute the monitoring statistics corresponding to the validating dataset, and establish the confidence limits using the kernel density estimation.

Online Monitoring Stage:

1) Gather a new online data vector and normalize it with the mean and variance of the normal training dataset.

2) Estimate its local probability density by referring to the normal training data, and then compute the statistic pattern vector.
3) Project the statistic pattern vector onto the LSKPCA model, and compute the corresponding monitoring statistics.
4) A fault is alarmed if any statistic exceeds its confidence limit.

V. CASE STUDY

In this section, two cases including one numerical nonlinear system and the simulated continuous stirred tank reactor (CSTR) are applied for method evaluation. Three methods including the traditional KPCA, local density estimation based KPCA (LKPCA), and the proposed LSKPCA are used for fault detection. For the monitoring statistics of all the methods, 95% confidence limits are adopted as fault detection threshold. In the following charts, the confidence limit is plotted with dashed line while the monitoring statistic is plotted with solid curve. If the solid line is above the dashed line, it means a fault is detected.

A. A NUMERICAL NONLINEAR SYSTEM

A nonlinear numerical example, designed by Dong and McAvoy [17], is used to test the validity of the proposed method. Its mathematical model is given as

$$\begin{cases} x_1 = s + e_1 \\ x_2 = s^2 - 3s + e_2 \\ x_3 = -s^3 + 3s^2 + e_3, \end{cases} \tag{26}$$

where $x_i(i = 1, 2, 3)$ is the monitored variable, $e_i(i = 1, 2, 3) \sim N(0, 0.01)$ denotes the dependent noises and s represents the system source variable following the uniform distribution. Two normal operating modes are designed by setting $s \in [0.01, 1]$ and $s \in [3.5, 4.3]$, respectively. A total of 1000 normal samples, including 500 samples for each mode, are simulated to constitute the normal training dataset. Another 1000 normal samples are simulated as the normal validating dataset. Two types of faults with 1000 samples for each fault, shown as following, are applied to test the fault detection algorithms.

- Fault D1: The system is runing under mode 1, and s is imposed with a step bias of -0.2 from the 401th sample.
- Fault D2: The system is running under mode 2, and x_1 has a ramp change with the slope of 0.005 from the 401th sample.

KPCA and LSKPCA are applied to monitor this nonlinear system. Both methods apply the Gaussian kernel width parameter, which is set as $c = 100m$, where m is the dimension of input space. When LSKPCA is used, the local neighbor number L is set to 50 and the moving window width w in SPA is chosen as 20. In order to maintain the independence of SPs, the interval between the initial samples of the moving windows is set to 10 during the offline training stage. But in online detection stage, this interval is set as 1 to reduce the detection delay.

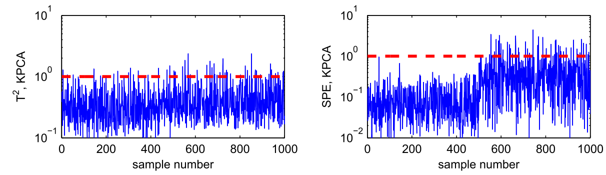

FIGURE 8. KPCA monitoring charts on the normal validating dataset.

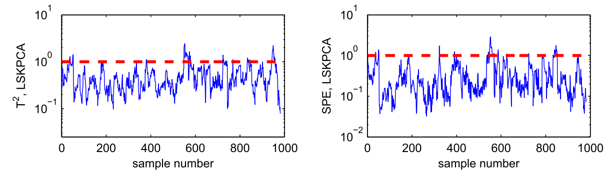

FIGURE 9. LSKPCA monitoring charts on the normal validating dataset.

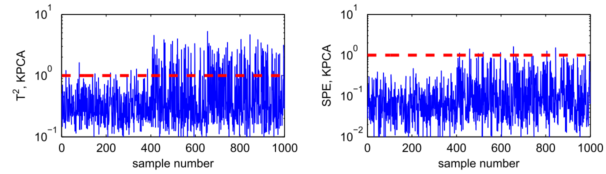

FIGURE 10. KPCA monitoring charts for the numerical system in the case of fault D1.

The monitoring results on the normal validating dataset are shown in Figs. 8 and 9. It is seen that most of the normal operating samples are under the confidence limits. The false alarming rates of both methods are all 5% since the 95% confidence limits are applied. However, it should be pointed out that the data distribution of KPCA *SPE* statistic is unbalanced because two operation modes are utilized in KPCA modeling. It is obvious from KPCA *SPE* chart that the confidence limit is too large for mode 1 while too small for mode 2, which will surely cause high false alarming rate or miss alarming rate. However, after using local statistics, LSKPCA method can remove the multimode data characteristic and balance the monitoring results of two modes.

Fault D1 is firstly tested and the monitoring results of the two methods are shown in Figs. 10 and 11. According to Fig. 10, the KPCA T^2 and *SPE* statistics fail to detect the fault and the corresponding fault detection rates are 22.2% and 1.67%, respectively. Compared to KPCA, LSKPCA gives the alarming signals more clearly. The fault detection rates of the LSKPCA T^2 and *SPE* statistics reach up to 97.2% and 96.7%, respectively. Obviously, LSKPCA method provides the better monitoring performance in the case of fault D1.

Fault D2 is further illustrated, which involves the drift change of process variable. The monitoring results of two methods are plotted in Figs. 12 and 13, respectively. By Fig. 12, the KPCA T^2 statistic is unable to respond to the fault in time, whose fault detection rate is only 16.2%. The KPCA *SPE* statistic alarms the fault with a low fault

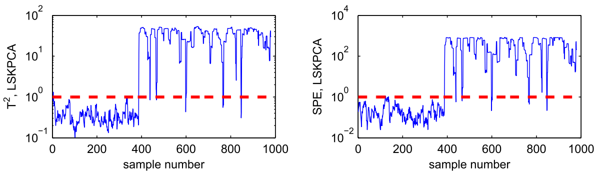

FIGURE 11. **LSKPCA monitoring charts for the numerical system in the case of fault D1.**

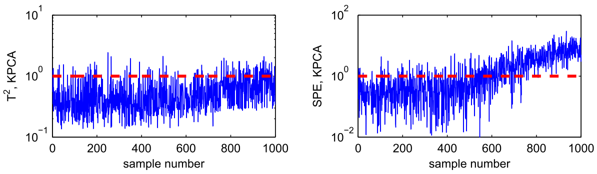

FIGURE 12. **KPCA monitoring charts for the numerical system in the case of fault D2.**

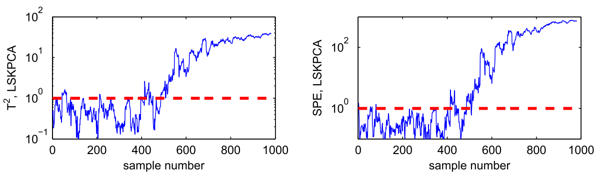

FIGURE 13. **LSKPCA monitoring charts for the numerical system in the case of fault D2.**

TABLE 1. **Fault detection rates (%) of numerical system faults D1 and D2 obtained by KPCA and LSKPCA.**

Testing Fault	KPCA		LSKPCA	
	T^2	SPE	T^2	SPE
D1	22.2	1.67	97.2	96.7
D2	16.2	67.7	87.5	85.8
Average	19.2	34.7	92.4	91.3

detection rate of 67.7%. When LSKPCA is applied to detect fault D2, its T^2 statistic gives a fault detection rate of 87.5%, superior to the KPCA T^2 statistic, while its *SPE* statistic has a fault detection rate of 85.8%, also higher than the KPCA *SPE* statistic. The monitoring results on fault D2 demonstrate the effectiveness of LSKPCA again.

The fault detection rates of KPCA and LSKPCA on the two tested faults are summarized in the Table 1. it is clear that LSKPCA achieves the obvious performance improvement in the process monitoring of the numerical system.

B. CSTR SYSTEM

The continuous stirred tank reactor (CSTR) simulation system is a typical industrial process which is commonly used to evaluate control algorithms and fault diagnosis methods [25], [53]–[55]. As shown in Fig. 14, material A is fed into the reactor with a steady flow rate. Under strong agitation, A is completely mixed with the liquid in the reactor meanwhile chemical exothermic reaction happens with a product B. By regulating the flow rate of outlet stream, the liquid level in reactor remains stable. There is a jacket filled with cooling water outside the reactor. The continuous coolant takes away heat to maintain a constant temperature.

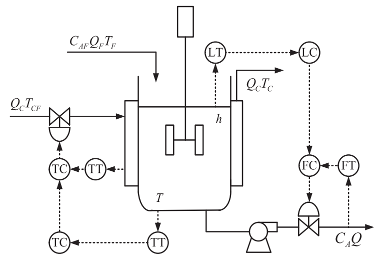

FIGURE 14. **Flowchart of the CSTR system.**

TABLE 2. **Different operating modes of the CSTR process.**

Mode	$Q_F(L \cdot min^{-1})$	$C_{AF}(mol \cdot L^{-1} \cdot min^{-1})$
1	100	1
2	110	1
3	100	0.9

TABLE 3. **Five fault patterns of the CSTR system.**

Fault	Description
F1	Feed concentration ramps up.
F2	Catalyst deactivation occurs.
F3	Heat exchanger fouling occurs.
F4	Reactor temperature sensor has a bias.
F5	Coolant temperature sensor has a bias.

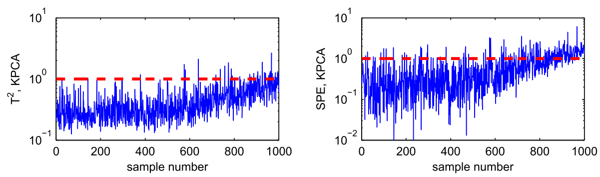

FIGURE 15. **KPCA monitoring charts for the CSTR sytem in the case of Mode 1-fault 1.**

The CSTR process mechanics can be seen in the related literature [53], [54]. In the CSTR simulation process, three operation modes are listed in shown in Table 2, where Q_F, C_{AF} represent the feed flow and the feed concentration, respectively. For each mode, ten process variables are recorded as the monitored variables. Normal operation is firstly simulated to generate 2000 samples for each mode. These samples are divided into the training and validating datasets, which include 1000 samples respectively. For method testing, five types of fault, listed in Table 3, are introduced to generate the fault data [53], [55]. These five faults represent some common process changes. Specifically, fault F1 is the operation condition change, fault F2 and F3 are the process parameter changes, while fault F4 and F5 are the sensor bias cases. Each fault dataset consists of 1000 samples, where fault is imposed from the 401th sample.

Three methods of KPCA, LKPCA and LSKPCA are applied to detect the CSTR system faults. The statistical model parameters are determined by the rules same to the numerical example. The monitoring results on fault F1 under mode 1 are illustrated in Figs. 15 to 17. In Fig. 15, KPCA

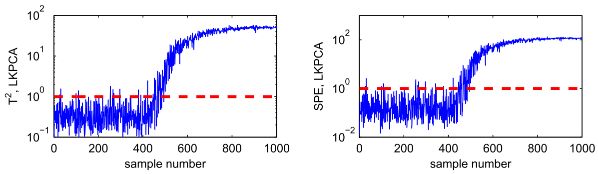

FIGURE 16. LKPCA monitoring charts for the CSTR sytem in the case of Mode 1-fault 1.

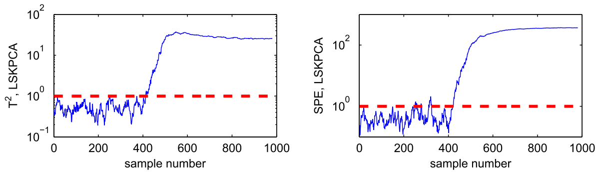

FIGURE 17. LSKPCA monitoring charts for the CSTR sytem in the case of Mode 1-fault 1.

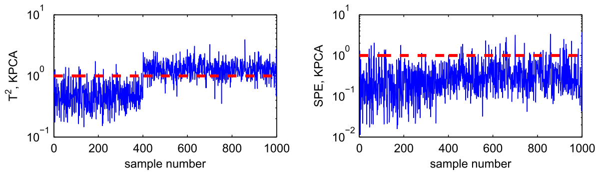

FIGURE 18. KPCA monitoring charts for the CSTR sytem in the case of Mode 2-fault 5.

performs so poorly that its T^2 statistic can not detect the abnormality effectively and its *SPE* statistic gives alarming signal until the 902th sample. In contrast to the KPCA results, the LKPCA T^2 and *SPE* charts give the alarming signals at the 471th sample simultaneously. LSKPCA detects this fault successfully at the 413th sample for the T^2 chart and at the 423th sample for the *SPE* chart. The results indicate that LSKPCA gives a better fault detection performance with the help of local probability density estimation and SPA technique.

The fault detection charts for the fault F5 under mode 2 are shown in Figs. 18 to 20. It is clear that LSKPCA is the most sensitive to the fault among these three methods, whose montoring statistics exceed the confidence limits clearly. By contrast, KPCA and LKPCA do not as well as LSKPCA. The KPCA T^2 statistic is fluctuant around the confidence limits while the KPCA *SPE* statistic fails to detect the abnormalities from beginning to the end. LKPCA performs better than KPCA. When LKPCA is applied, the monitoring statistics of most samples go beyond the confidence limits clearly but some faulty samples are still under the confidence limits. The results can be further analyzed in terms of fault detection rate. Traditional KPCA provides the fault detection rates of 80.3% and 8.33% for the T^2 and *SPE* statistics, respectively. In contrast to KPCA, LKPCA performs better whose fault detection rates are 98% and 98.3%, for the T^2 and *SPE* statistics respectively. For this fault, LSKPCA gives the highest detection rates of 100% for both the T^2 and *SPE* statistics. Generally, LSKPCA has the best monitoring performance on this fault.

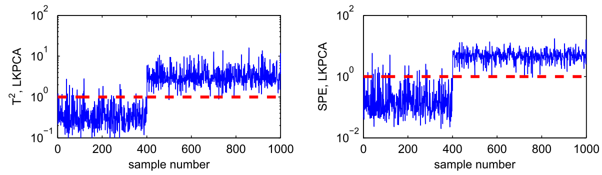

FIGURE 19. LKPCA monitoring charts for the CSTR sytem in the case of Mode 2-fault 5.

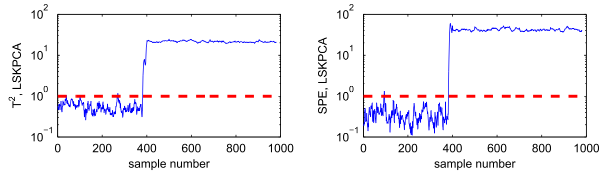

FIGURE 20. LSKPCA monitoring charts for the CSTR system in the case of Mode 2-fault 5.

TABLE 4. Fault detection rates (%) of the CSTR system faults obtained by KPCA, SKPCA, LKPCA and LSKPCA.

Mode	Fault	KPCA		SKPCA		LKPCA		LSKPCA	
No.	No.	T^2	*SPE*	T^2	*SPE*	T^2	*SPE*	T^2	*SPE*
1	F1	7.5	30.8	9.5	37.2	89.7	89.8	94.8	93.2
1	F2	66.3	72.7	87.7	75.7	79.7	79.5	90.3	87.0
1	F3	75.5	51.0	29.7	89.2	88.7	88.8	91.3	91.3
1	F4	99.7	67.5	100	100	100	100	100	100
1	F5	70.8	5.5	4.8	100	99.7	99.7	99.8	99.8
2	F1	57.3	48.2	2.3	64.5	89.2	89.7	94.7	94.2
2	F2	81.8	73.0	86.7	75.8	81.2	82.3	88.2	88.2
2	F3	85.3	62.8	40.3	90.5	91.0	92.3	93.3	94.0
2	F4	100	67.8	100	100	100	100	100	100
2	F5	80.3	8.33	7.0	100	99.3	99.7	100	100
3	F1	3.0	4.67	2.5	6.0	89.5	89.5	93.3	92.5
3	F2	71.2	69.8	84.2	69.7	73.3	75.2	83.0	83.2
3	F3	70.7	30.2	2.7	84.5	83.0	83.3	88.7	88.0
3	F4	99.8	33.2	100	100	100	100	100	100
3	F5	86.0	8.0	5.7	100	99.8	99.8	99.8	99.8
Average		70.3	42.2	44.2	79.5	90.8	91.3	94.5	94.1

Furthermore, we apply four methods of KPCA, LKPCA, SKPCA and LSKPCA to monitor all faults of the CSTR system. Here SKPCA is the combination of statistics pattern analysis (SPA) and KPCA. Fault detection rates (FDRs) of four methods for the five faults under the three modes are summarized in Table 4. Among these four methods, KPCA is the basic method. As it does not consider the multimode characteristic of process data and statistical information mining, it gives the unsatisfactory monitoring performance with the average FDRs of 70.3%, 42.2% for T^2 and *SPE*, respectively. SKPCA applies SPA to improve the process monitoring performance. However, for multimode process, it can not give clear improvements. The average FDRs of SKPCA T^2 and *SPE* are 44.2% and 79.5%, respectively. It should be pointed out that SKPCA *SPE* performs better than the one of KPCA, but SKPCA T^2 does worse than the KPCA T^2. By contrast, LKPCA achieves obvious improvement, which has the higher average FDRs of 90.8%, 91.3% for T^2 and *SPE*, respectively. This is because the multimode distribution is one main data property and LKPCA is capable of dealing

TABLE 5. Average false alarm rates (%) of the CSTR system normal samples obtained by KPCA, EKPCA, LKPCA and LSKPCA.

	KPCA		SKPCA		LKPCA		LSKPCA	
	T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE
FAR	4.32	4.48	1.91	6.04	4.58	4.51	3.60	3.81

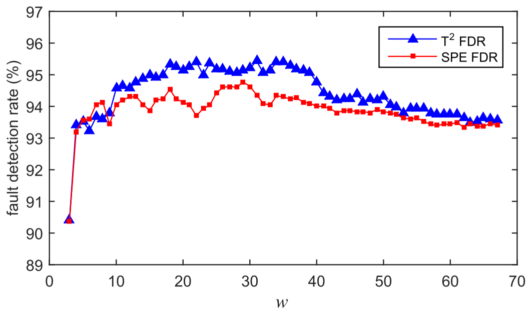

FIGURE 21. Average fault detection rates of 5 faults using different *w* values.

with the multimodal data characteristic. When LKPCA is integrated with SPA, the LSKPCA method further increases the average FDRs to 94.5% and 94.1% for T^2 and SPE, respectively. To sum up, local density estimation and statistics pattern analysis are helpful to improve the fault detection performance of the nonlinear multimode processes.

In process monitoring field, the false alarm rate (FAR) is also an important performance index, which indicates the monitoring performance on normal samples. In this paper, the first 400 samples of each testing fault dataset are under the normal condition. We compute the the average false alarm rates on all the normal samples and list the results in Table 5. From this table, most of the false alarm rates are below 5% except that the FAR of SKPCA SPE statistic is 6.04%, a little higher than 5%. It is also found that LSKPCA FARs are a little smaller than the other two methods. This means better process monitoring performance on the normal data. Combining the analyses about FDR and FAR, it can be concluded that the proposed method LSKPCA are more excellent than traditional methods in detecting the CSTR system faults.

In LSKPCA algorithm, the moving window width w is a crucial parameter that has a significant effect on the monitoring performance. As various statistics are computed within the moving window, the statistical randomness will be high if w is too small. With the increase of w value, the faulty property will be weaken especially when the window includes both normal and fault samples. As a result, the fault detection delay will certainly be increased. In order to analyze the influence of w, 5 faults under mode 1 are monitored under different w, whose average fault detection rates are presented in Fig. 21. It can be observed from Fig. 21 that when $w < 10$ the fault detection rates of two statistics are increasing with a distinct trend. After the rapid rise, the fault detection rates keep stable around the maximum value. When $w > 40$, the two curves both go down the slope due to the monitoring delay. This is consistent to the above analysis. Therefore, the window width parameter should be selected between 10 and 40. In this paper w=20 is used. It should be noted that the influences of parameters may be different for various processes. To obtain an optimal w value, some experiments should be designed to obtain the reasonable value.

VI. CONCLUSION

This paper proposes a novel nonlinear multimode process fault detection method based on LSKPCA. The proposed method integrates two-fold data mining strategies. On one hand, local probability density estimation is applied to transform the multimode data into the unimodal data, which provides the good data foundation for the further statistical modeling. On the other hand, SPA is adopted to mine the intrinsic statistical information, which is helpful to improve the fault detection performance. A numerical nonlinear example and the CSTR simulation system are used to validate the proposed method and the application results show that the proposed method outperforms the traditional KPCA method. However, there are some related problems deserving further studies. One problem involves the transition stage monitoring, which is not considered in this work. Compared with the multiple steady stages, the transition stages are much more complex because of their obvious dynamic changes. It is of great value to design some effective transition stage monitoring methods. Another problem is about fault source variable identification. In this work, we only determine if some faults occur but do not know which variable is responsible for the fault. In the future, we will carry out the further study to identify the fault source variable for process recovery.

REFERENCES

[1] K. Tidriri, N. Chatti, S. Verron, and T. Tiplica, "Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges," *Annu. Rev. Control*, vol. 42, pp. 63–81, Sep. 2016.

[2] M. Kano and Y. Nakagawa, "Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry," *Comput. Chem. Eng.*, vol. 32, nos. 1–2, pp. 12–24, Jan./Feb. 2008.

[3] B. Cai, H. Liu, and M. Xie, "A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks," *Mech. Syst. Signal Process.*, vol. 80, pp. 31–44, Dec. 2016.

[4] B. Cai, Y. Zhao, H. Liu, and M. Xie, "A data-driven fault diagnosis methodology in three-phase inverters for PMSM drive systems," *IEEE Trans. Power Electron.*, vol. 32, no. 7, pp. 5590–5600, Jul. 2017.

[5] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Ind. Eng. Chem. Res.*, vol. 52, no. 10, pp. 3543–3562, Mar. 2013.

[6] B. Cai, L. Huang, and M. Xie, "Bayesian networks in fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2227–2240, Oct. 2017.

[7] B. Cai *et al.*, "Multi-source information fusion based fault diagnosis of ground-source heat pump using Bayesian network," *Appl. Energy*, vol. 114, no. 2, pp. 1–9, Feb. 2014.

[8] Y. Liu, Y. Pan, Q. Wang, and D. Huang, "Statistical process monitoring with integration of data projection and one-class classification," *Chemometrics Intell. Lab. Syst.*, vol. 149, pp. 1–11, Dec. 2015.

[9] L. Cai, X. Tian, and S. Chen, "Monitoring nonlinear and non-Gaussian processes using Gaussian mixture model-based weighted kernel independent component analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 1, pp. 122–135, Jan. 2017.

[10] K. Peng, Q. Li, K. Zhang, and D. Dong, "Quality-related process monitoring for dynamic non-Gaussian batch process with multi-phase using a new data-driven method," *Neurocomputing*, vol. 214, pp. 317–328, Nov. 2016.

[11] L. Shang, J. Liu, and Y. Zhang, "Recursive fault detection and identification for time-varying processes," *Ind. Eng. Chem. Res.*, vol. 55, no. 46, pp. 12149–12160, Nov. 2016.
[12] A. Alawi, J. Zhang, and J. Morris, "Multiscale multiblock batch monitoring: Sensor and process drift and degradation," *Organic Process Res. Develop.*, vol. 19, no. 1, pp. 145–157, Jan. 2015.
[13] Z. Zhou, C. Yang, C. Wen, and J. Zhang, "Analysis of principal component analysis-based reconstruction method for fault diagnosis," *Ind. Eng. Chem. Res.*, vol. 55, no. 27, pp. 7402–7410, Jun. 2016.
[14] Y. Satoyama, K. Fujiwara, and M. Kano, "Variable elimination-based contribution for accurate fault identification," *IFAC-PapersOnline*, vol. 49, no. 7, pp. 383–388, Jul. 2016.
[15] J. Zhu, Z. Ge, and Z. Song, "HMM-driven robust probabilistic principal component analyzer for dynamic process fault classification," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3814–3821, Jun. 2015.
[16] Q. Jiang and X. Yan, "Plant-wide process monitoring based on mutual information–multiblock principal component analysis," *ISA Trans.*, vol. 53, no. 5, pp. 1516–1527, Sep. 2014.
[17] D. Dong and T. J. McAvoy, "Nonlinear principal component analysis—Based on principal curves and neural networks," *Comput. Chem. Eng.*, vol. 20, no. 1, pp. 65–78, Jan. 1996.
[18] F. Jia, E. B. Martin, and A. J. Morris, "Non-linear principal components analysis for process fault detection," *Comput. Chem. Eng.*, vol. 22, pp. S851–S854, Mar. 1998.
[19] B. Schölkopf, A. Smola, and K. R. Möller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.
[20] J.-M. Lee, C. Yoo, S. W. Choi, P. A. Vanrolleghem, and I.-B. Lee, "Nonlinear process monitoring using kernel principal component analysis," *Chem. Eng. Sci.*, vol. 59, no. 1, pp. 223–234, Jan. 2004.
[21] J.-M. Lee, C. Yoo, and I.-B. Lee, "Fault detection of batch processes using multiway kernel principal component analysis," *Comput. Chem. Eng.*, vol. 28, no. 9, pp. 1837–1847, Aug. 2004.
[22] S. W. Choi and I.-B. Lee, "Nonlinear dynamic process monitoring based on dynamic kernel PCA," *Chem. Eng. Sci.*, vol. 59, no. 24, pp. 5897–5908, Dec. 2004.
[23] M. Jia, F. Chu, F. Wang, and W. Wang, "On-line batch process monitoring using batch dynamic kernel principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 101, no. 2, pp. 110–122, Apr. 2010.
[24] Q. Jiang and X. Yan, "Nonlinear plant-wide process monitoring using MI-spectral clustering and Bayesian inference-based multiblock KPCA," *J. Process Control*, vol. 32, pp. 38–50, Aug. 2015.
[25] X. Deng, X. Tian, and S. Chen, "Modified kernel principal component analysis based on local structure analysis and its application to nonlinear process fault diagnosis," *Chemometrics Intell. Lab. Syst.*, vol. 127, pp. 195–209, Aug. 2013.
[26] Y. Zhang, S. Li, and Z. Hu, "Improved multi-scale kernel principal component analysis and its application for fault detection," *Chem. Eng. Res. Des.*, vol. 90, no. 9, pp. 1271–1280, Sep. 2012.
[27] Z. Ge, C. Yang, and Z. Song, "Improved kernel PCA-based monitoring approach for nonlinear processes," *Chem. Eng. Sci.*, vol. 64, no. 9, pp. 2245–2255, May 2009.
[28] Y. Zhang, S. Li, and Y. Teng, "Dynamic processes monitoring using recursive kernel principal component analysis," *Chem. Eng. Sci.*, vol. 72, pp. 78–86, Apr. 2012.
[29] Q. Jiang and X. Yan, "Weighted kernel principal component analysis based on probability density estimation and moving window and its application in nonlinear chemical process monitoring," *Chemometrics Intell. Lab. Syst.*, vol. 127, pp. 121–131, Aug. 2013.
[30] S. W. Choi, C. Lee, J.-M. Lee, J. H. Park, and I. B. Lee, "Fault detection and identification of nonlinear processes based on kernel PCA," *Chemometrics Intell. Lab. Syst.*, vol. 75, no. 1, pp. 55–67, Jan. 2005.
[31] Q. Jiang, J. Li, and X. Yan, "Performance-driven optimal design of distributed monitoring for large-scale nonlinear processes," *Chemometrics Intell. Lab. Syst.*, vol. 155, pp. 151–159, Jul. 2016.
[32] C. F. Alcala and S. J. Qin, "Reconstruction-based contribution for process monitoring with kernel principal component analysis," *Ind. Eng. Chem. Res.*, vol. 49, no. 17, pp. 7849–7857, Sep. 2010.
[33] C.-Y. Cheng, C.-C. Hsu, and M.-C. Chen, "Adaptive kernel principal component analysis (KPCA) for monitoring small disturbances of nonlinear processes," *Ind. Eng. Chem. Res.*, vol. 49, no. 5, pp. 2254–2262, Mar. 2010.
[34] S. J. Zhao, J. Zhang, and Y. Xu, "Monitoring of processes with multiple operating modes through multiple principle component analysis models," *Ind. Eng. Chem. Res.*, vol. 43, no. 22, pp. 7025–7035, Sep. 2004.
[35] S. Natarajan and R. Srinivasan, "Multi-model based process condition monitoring of offshore oil and gas production process," *Chem. Eng. Res. Des.*, vol. 88, nos. 5–6, pp. 572–591, May/Jun. 2009.
[36] X. Xu, L. Xie, and S. Wang, "Multimode process monitoring with PCA mixture model," *Comput. Elect. Eng.*, vol. 40, no. 7, pp. 2101–2112, Oct. 2014.
[37] Z. Ge and Z. Song, "Multimode process monitoring based on Bayesian method," *J. Chemometrics*, vol. 23, no. 12, pp. 636–650, Nov./Dec. 2009.
[38] Q. P. He and J. Wang, "Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 345–354, Nov. 2007.
[39] X. Deng and X. Tian, "Multimode process fault detection using local neighborhood similarity analysis," *Chin. J. Chem. Eng.*, vol. 22, nos. 11–12, pp. 1260–1267, Nov. 2014.
[40] H. Ma, Y. Hu, and H. Shi, "A novel local neighborhood standardization strategy and its application in fault detection of multimode processes," *Chemometrics Intell. Lab. Syst.*, vol. 118, pp. 287–300, Aug. 2012.
[41] G. Wang, J. Liu, Y. Zhang, and Y. Li, "A novel multi-mode data processing method and its application in industrial process monitoring," *J. Chemometrics*, vol. 29, no. 2, pp. 126–138, Feb. 2015.
[42] Y. Choi, S. Ozawa, and M. Lee, "Incremental two-dimensional kernel principal component analysis," *Neurocomputing*, vol. 134, pp. 280–288, Jun. 2014.
[43] M. Ding, Z. Tian, and H. Xu, "Adaptive kernel principal component analysis," *Signal Process.*, vol. 90, no. 5, pp. 1542–1553, May 2010.
[44] R. T. Samuel and Y. Cao, "Nonlinear process fault detection and identification using kernel PCA and kernel density estimation," *Syst. Sci. Control Eng.*, vol. 4, no. 1, pp. 165–174, Oct. 2016.
[45] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: Chapman & Hall, 1986.
[46] P. Ji, N. Zhao, S. Hao, and J. Jiang, "Automatic image annotation by semi-supervised manifold kernel density estimation," *Inf. Sci.*, vol. 281, pp. 648–660, Oct. 2014.
[47] R. Gonzalez, B. Huang, and E. Lau, "Process monitoring using kernel density estimation and Bayesian networking with an industrial case study," *ISA Trans.*, vol. 58, pp. 330–347, Sep. 2015.
[48] R. Liu and Y. Lu, "Infrared target tracking in multiple feature pseudo-color image with kernel density estimation," *Infr. Phys. Technol.*, vol. 55, pp. 505–512, Nov. 2012.
[49] Q. P. He and J. Wang, "Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes," *AIChE J.*, vol. 57, no. 1, pp. 107–121, Jan. 2011.
[50] J. Wang and Q. P. He, "Multivariate statistical process monitoring based on statistics pattern analysis," *Ind. Eng. Chem. Res.*, vol. 49, no. 17, pp. 7858–7869, Sep. 2010.
[51] H. Zhang, X. Tian, X. Deng, and L. Cai, "A local and global statistics pattern analysis method and its application to process fault identification," *Chin. J. Chem. Eng.*, vol. 23, no. 11, pp. 1782–1792, Nov. 2015.
[52] X. Deng and X. Tian, "Nonlinear process fault pattern recognition using statistics kernel PCA similarity factor," *Neurocomputing*, vol. 121, pp. 298–308, Dec. 2013.
[53] M. C. Johannesmeyer, A. Singhal, and D. E. Seborg, "Pattern matching in historical data," *AIChE J.*, vol. 48, no. 9, pp. 2022–2038, Sep. 2002.
[54] X. Deng and X. Tian, "Sparse kernel locality preserving projection and its application in nonlinear process fault detection," *Chin. J. Chem. Eng.*, vol. 21, no. 2, pp. 163–170, Feb. 2013.
[55] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Fault discriminant enhanced kernel principal component analysis incorporating prior fault information for monitoring nonlinear processes," *Chemometrics Intell. Lab.*, vol. 162, pp. 21–34, Mar. 2017.

XIAOGANG DENG received the B.Eng. and Ph.D. degrees from the China University of Petroleum, Dongying, China, in 2002 and 2008, respectively. He is currently an Associate Professor with the College of Information and Control Engineering, China University of Petroleum. From 2015 to 2016, he was a Visiting Scholar with the Department of Electronics and Computer Sciences, University of Southampton, Southampton, U.K. His research interests include industrial process modeling and simulation, data-driven fault detection and diagnosis, and control performance monitoring.

NA ZHONG received the bachelor's and M.Eng. degrees from the College of Information and Control Engineering, China University of Petroleum, in 2014 and 2017, respectively. Her research interests include statistical signal processing, machine learning theory, and industrial process monitoring.

LEI WANG received the B.Eng. degree from the College of Information and Control Engineering, China University of Petroleum, in 2015. He is currently pursuing the master's degree with the College of Information and Control Engineering, China University of Petroleum, Qingdao. His research interests include statistical signal processing, machine learning theory, and industrial process monitoring.