Statistics Local Fisher Discriminant Analysis for Industrial Process Fault Classification

Xiaogang Deng[†], Xuemin Tian[†], Sheng Chen[‡,§], Chris J. Harris[‡]

[†]College of Information and Control Engineering, China University of Petroleum, Qingdao 266580,
China (E-mails: dengxg2002@gmail.com, tianxm@upc.edu.cn)

[‡]Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ,
UK (E-mails: sqc@ecs.soton.ac.uk, cjh@ecs.soton.ac.uk)

[§]King Abdulaziz University, Jeddah 21589, Saudi Arabia

Abstract—In order to effectively identify industrial process faults, an improved Fisher discriminant analysis (FDA) method, referred to as the statistics local Fisher discriminant analysis (SLFDA), is proposed for fault classification. For mining statistics information hidden in process data, statistics pattern analysis is firstly applied to transform the original measured variables into the corresponding statistics, including second-order and higher-order ones. Furthermore, considering the local structure characteristics of fault data, local FDA (LFDA) is performed which computes the discriminant vectors by modifying the optimization objective with local weighting factor. Simulation results on the benchmark Tennessee Eastman process show that the proposed SLFDA has a better fault classification performance than the FDA and LFDA methods.

I. INTRODUCTION

Industrial process fault diagnosis technology plays a key role in ensuring process safety and enhancing product quality. In the past two decades, data-based fault diagnosis approach becomes ever increasingly popular because huge amounts of process data are collected and stored by computer control systems [1]–[3]. In order to deal with complicated process characteristics, many data-based fault diagnosis methods have been developed, which includes principal component analysis (PCA), independent component analysis (ICA), canonical variate analysis (CVA), Fisher Discriminant analysis (FDA), *et al.* [4]–[8]. In process monitoring application, these methods utilize historical data to build statistical models for fault detection and fault classification.

Among the existing statistical analysis methods, FDA is a well-known pattern classification technique [9]. Because of its simplicity and effectiveness, FDA has been widely used to construct different fault classification algorithms. Chiang *et al.* developed an FDA based industrial process fault classification method [10] and then proposed a dynamic FDA method by augmenting the observed vectors [11]. To deal with the serial correlation property of process data, Jiang *et al.* [12] developed a CVA-FDA method which uses CVA to extract dynamic features for helping fault discrimination. For utilizing both labeled and unlabeled samples, Zhong *et al.* [13] presented a semi-supervised FDA model which combines the objectives of PCA and FDA. Considering process nonlinearity, Zhu

and Song [14] studied kernel FDA method by using kernel, while Li and Cui [15] proposed an improved kernel FDA which applies feature vector selection to reduce computational complexity. To improve the discriminant power of FDA, a nested loop FDA algorithm was built by Zhao and Gao [16] with inner-loop and outer-loop iterative calculations.

All the aforementioned methods are based on global data structure analysis and do not take into account data local structure characteristics, such as multi-modality. To overcome this disadvantage, Sugiyama [17] developed a local FDA (LFDA) method by combining the idea of local structure preserving. Yu [18] studied a localized FDA based process monitoring method and proved its superiority over the FDA approach. Ren *et al.* [19] proposed to monitor multimode processes by integrating the discriminant local consistency Gaussian mixture model (DLCGMM) with the LFDA. For nonlinear system monitoring, Van and Kang [20] developed a wavelet kernel LFDA method whose parameters are optimized by a particle swarm optimization algorithm. To ensure the orthogonality of discriminant vectors, Li *et al.* [21] presented a supervised orthogonal local Fisher discriminant analysis method and tested its performance on a rotating machinery system.

Although FDA and LFDA have achieved great success in the fault diagnosis field, there are some issues worthy of further studies. Both the FDA and LFDA methods are based on second-order statistics analysis which only involve the mean and variance-covariance structure analysis of the original measured variables. In other words, FDA and LFDA ignore higher-order statistics information. Higher-order statistics of process variables often have close relationship with process operation states and can be very useful for fault classification. Therefore, mining process higher-order statistics information is extremely valuable to assist the FDA based methods. Recently, statistics pattern analysis (SPA) was proposed by Wang and He [22], and the results of [22] confirm that including both second-order and higher-one statistics of process variables can significantly enhance fault detection performance. Therefore, combining SPA with FDA related methods will provide a practical and powerful means to improve fault classification performance,

which motivates our current study. To our best knowledge, no work to date applies SPA to FDA-based fault classification.

In this paper, by utilizing both SPA for statistical information extraction and local structure analysis, we develop a new enhanced FDA method, which is referred to as the statistics LFDA (SLFDA). Specifically, in the proposed method, SPA is applied to extract different orders of statistics and then LFDA is used to build discriminant model with local structure analysis. Fault discriminant function is constructed to determine the class of fault samples. The rest of this paper is organized as follows. In Sections II and III, the principles of FDA and LFDA are briefly introduced, respectively. Section IV details our proposed SLFDA. Simulation results obtained using the Tennessee Eastman process are presented in Section V, and our conclusions are drawn in Section VI.

II. FISHER DISCRIMINANT ANALYSIS

FDA is a supervised learning technique widely used for pattern classification [9]–[11]. The training feature data set $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 \; \boldsymbol{x}_2 \cdots \boldsymbol{x}_N \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{N \times M}$, including N samples of the measured vector $\boldsymbol{x} \in \mathbb{R}^M$, consists of K classes of fault patterns $\{ \mathsf{C}_k, 1 \leq k \leq K \}$ with N_k samples for fault class C_k. The objective of FDA is to seek a set of projection vectors so that the between-class scatter is maximized and the within-class scatter is minimized simultaneously.

The within-class scatter matrix is defined as

$$\boldsymbol{S}_w = \sum_{k=1}^{K} \boldsymbol{S}_{w,k} = \sum_{k=1}^{K} \sum_{\boldsymbol{x}_i \in \mathsf{C}_k} \left(\boldsymbol{x}_i - \boldsymbol{\mu}_k \right) \left(\boldsymbol{x}_i - \boldsymbol{\mu}_k \right)^{\mathrm{T}} \quad (1)$$

where $\boldsymbol{S}_{w,k}$ is the scatter matrix for the k-th class, and $\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{\boldsymbol{x}_i \in \mathsf{C}_k} \boldsymbol{x}_i$ represents the mean vector of the k-th class. The between-class scatter matrix is given as

$$\boldsymbol{S}_b = \sum_{k=1}^{K} N_k \left(\boldsymbol{\mu}_k - \boldsymbol{\mu} \right) \left(\boldsymbol{\mu}_k - \boldsymbol{\mu} \right)^{\mathrm{T}} \quad (2)$$

where $\boldsymbol{\mu} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}_i$ is the mean vector of all fault samples.

FDA finds the discriminant vector $\boldsymbol{p}_j$ to maximize the between-class scatter as well as to minimize the within-class scatter, which is defined by the following optimization problem

$$\max_{\boldsymbol{p}_j} \frac{\boldsymbol{p}_j^{\mathrm{T}} \boldsymbol{S}_b \boldsymbol{p}_j}{\boldsymbol{p}_j^{\mathrm{T}} \boldsymbol{S}_w \boldsymbol{p}_j} \quad (3)$$

This optimization problem can be solved by the generalized eigenvalue decomposition according to

$$\boldsymbol{S}_b \boldsymbol{p}_j = \lambda_j \boldsymbol{S}_w \boldsymbol{p}_j \quad (4)$$

where λ_j is the generalized eigenvalue corresponding to the vector $\boldsymbol{p}_j$. Here the value of j ranges from 1 to $K - 1$, because the rank of $\boldsymbol{S}_b$ is less than K. For fault discrimination, we retain the eigenvectors corresponding to the first A $(A \leq K - 1)$ largest eigenvalues, which form the discriminant matrix $\boldsymbol{P}_A = \begin{bmatrix} \boldsymbol{p}_1 \; \boldsymbol{p}_2 \cdots \boldsymbol{p}_A \end{bmatrix} \in \mathbb{R}^{M \times A}$.

For a testing vector $\boldsymbol{x}$, its projection onto the discrimination vector $\boldsymbol{p}_j$ is called the j-th discriminant component, which is computed as

$$y_j = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{p}_j \quad (5)$$

By defining $\boldsymbol{\Gamma} = \frac{1}{N_k - 1} \boldsymbol{P}_A^{\mathrm{T}} \boldsymbol{S}_{w,k} \boldsymbol{P}_A$, the FDA discriminant function is constructed as [11], [12]

$$f_k(\boldsymbol{x}) = -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu}_k)^{\mathrm{T}} \boldsymbol{P}_A \boldsymbol{\Gamma}^{-1} \boldsymbol{P}_A^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{\mu}_k) - \frac{1}{2} \ln(\det(\boldsymbol{\Gamma})) \quad (6)$$

Based on the discriminant functions $f_k(\boldsymbol{x})$, $1 \leq k \leq K$, the class index $I_{C(\boldsymbol{x})}$ of the testing vector $\boldsymbol{x}$ is determined by

$$I_{C(\boldsymbol{x})} = \arg \max_{1 \leq k \leq K} \{ f_k(\boldsymbol{x}) \} \quad (7)$$

III. LOCAL FISHER DISCRIMINANT ANALYSIS

The optimization of FDA is based on the global structure analysis, which is effective for classifying the data with multivariate Gaussian distribution. However, fault data often show non-Gaussian characteristics, where local data structure becomes important for data discrimination. To overcome the shortcoming of FDA, LFDA has been proposed by integrating local structure preserving technique with FDA [17].

Note that for FDA, the within-class scatter matrix (1) can be rewritten as [17], [18]

$$\boldsymbol{S}_w = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} w_{i,j}^{(w)} \left(\boldsymbol{x}_i - \boldsymbol{x}_j \right) \left(\boldsymbol{x}_i - \boldsymbol{x}_j \right)^{\mathrm{T}} \quad (8)$$

with the within-class weighting coefficient $w_{i,j}^{(w)}$ defined by

$$w_{i,j}^{(w)} = \begin{cases} \frac{1}{N_k}, & \text{if } \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathsf{C}_k \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

while the between-class scatter matrix (2) can be reformulated as [17], [18]

$$\boldsymbol{S}_b = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} w_{i,j}^{(b)} \left(\boldsymbol{x}_i - \boldsymbol{x}_j \right) \left(\boldsymbol{x}_i - \boldsymbol{x}_j \right)^{\mathrm{T}} \quad (10)$$

with the between-class weighting coefficient $w_{i,j}^{(b)}$ given by

$$w_{i,j}^{(b)} = \begin{cases} \frac{1}{N} - \frac{1}{N_k}, & \text{if } \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathsf{C}_k \\ \frac{1}{N}, & \text{otherwise} \end{cases} \quad (11)$$

The formulations (8) and (10) indicate that the distances of data pairs between the same classes have important influence on within-class and between-class scatters. According to (9) and (11), the weighting coefficients are the same for the data points in the same class.

However, the distribution of fault data is complex because of its dynamic and non-Gaussian properties. Even in the same fault class, multimodality may be observed among data points. For example, some data points may represent the starting stage of a fault, while other data points may reflect the ending stage of the fault. Therefore, to impose the same importance on all data points as FDA does is unreasonable. LFDA introduces local structure preserving to modify the weighting coefficients.

In the framework of LFDA, the weighting coefficients given in (9) and (11) are modified respectively as

$$\overline{w}_{i,j}^{(w)} = \begin{cases} r_{i,j}\frac{1}{N_k}, & \text{if } \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathsf{C}_k \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

$$\overline{w}_{i,j}^{(b)} = \begin{cases} r_{i,j}\left(\frac{1}{N} - \frac{1}{N_k}\right), & \text{if } \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathsf{C}_k \\ \frac{1}{N}, & \text{otherwise} \end{cases} \tag{13}$$

where $r_{i,j}$ is a local weighting factor used to measure the closeness degree between the data pair $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. If the two data points are close, the factor should be large to emphasize the local structure preserving. Otherwise, for the two points that are far apart, the factors should be small to avoid destroying the data structure. According to [17], [18], the local weighting factor $r_{i,j}$ can be computed as

$$r_{i,j} = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|}{\sigma}\right) \tag{14}$$

where σ is the local scaling parameter represented by the distance between the sample and its nearest neighbors. By substituting (9) and (11) for (12) and (13), the FDA is transformed into the LFDA, which is capable of extracting discriminant features while imposing local structure preserving.

IV. Statistics Local Fisher Discriminant Analysis

The FDA and LFDA methods are directly based on the original measured process variables. More specifically, they extract the second-order statistics from the original measured variables for fault classification, and they do not carry out deep mining for higher-order statistical information. In fact, different orders of the process variables' statistics, including second-order and higher-order statistics, are all closely related to fault operation mode, which are very important for fault classification. Using only the second-order statistics may be unable to discover certain complicated faults. In the SPA framework [22]–[24], the process behavior is described not directly by the process variables but by their different statistics.

Our proposed SLFDA combines SPA with LFDA, whose schematic is depicted in Fig. 1. Two steps are involved in the SLFDA based fault classification. In the first step, SPA is performed to acquire the statistics of all the measured variables. Then in the second step, LFDA is applied to build a discriminant model. We now detail our SLFDA modeling.

For the data matrix $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \cdots \boldsymbol{x}_N \end{bmatrix}^{\mathrm{T}}$, a data window with width L is captured at time instant t ($L \le t \le N$), which
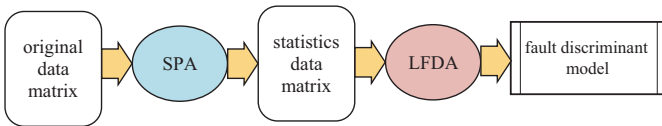
Fig. 1: The Schematic of SLFDA modeling.

is denoted as $\boldsymbol{X}_t = [\boldsymbol{x}_{t-L+1} \ \boldsymbol{x}_{t-L+2} \cdots \ \boldsymbol{x}_t]^{\mathrm{T}} \in \mathbb{R}^{L \times M}$. This data window can be represented explicitly as

$$\boldsymbol{X}_t = \begin{bmatrix} x_{t-L+1}(1) & x_{t-L+1}(2) & \cdots & x_{t-L+1}(M) \\ x_{t-L+2}(1) & x_{t-L+2}(2) & \cdots & x_{t-L+2}(M) \\ \vdots & \vdots & \ddots & \vdots \\ x_t(1) & x_t(2) & \cdots & x_t(M) \end{bmatrix} \tag{15}$$

where $x_i(j)$ denotes the i-th sample of the j-th variable.

Given the data window (15), three types of the process variables' statistics are constructed, which includes first-order, second-order and higher-order statistics. The j-th process variable's first-order statistic or mean, $\mu_t(j)$, is defined as

$$\mu_t(j) = \frac{1}{L}\sum_{i=t-L+1}^{t} x_i(j) \tag{16}$$

The second order statistics of the j-th process variable are the variance $v_t(j)$ and the one-lag autocorrelation $c_t(j)$, which are given respectively by

$$v_t(j) = \frac{1}{L-1}\sum_{i=t-L+1}^{t} \big(x_i(j) - \mu_t(j)\big)^2 \tag{17}$$

$$c_t(j) = \frac{1}{L-1}\frac{\sum\limits_{i=t-L+1}^{t-1}\big(x_i(j)-\mu_t(j)\big)\big(x_{i+1}(j)-\mu_t(j)\big)}{v_t(j)} \tag{18}$$

The higher-order statistics considered are skewness and kurtosis, which quantify the nonlinearity and non-Gaussianity of process variables [22]. In particular, skewness measures the asymmetry of the process variable distribution, while the kurtosis measures the peakiness of the distribution. The skewness of the j-th process variable is defined as

$$\gamma_t(j) = \frac{\frac{1}{L}\sum\limits_{i=t-L+1}^{t}(x_i(j) - \mu_t(j))^3}{\left(\frac{1}{L}\sum\limits_{i=t-L+1}^{t}(x_i(j) - \mu_t(j))^2\right)^{3/2}} \tag{19}$$

while the kurtosis of the j-th process variable is given by

$$\kappa_t(j) = \frac{\frac{1}{L}\sum\limits_{i=t-L+1}^{t}(x_i(j) - \mu_t(j))^4}{\left(\frac{1}{L}\sum\limits_{i=t-L+1}^{t}(x_i(j) - \mu_t(j))^2\right)^2} - 3 \tag{20}$$

By computing the various statistics, (16) to (20), for $L \le t \le N$, the original data matrix $\boldsymbol{X}$ is transformed into the new statistics matrix $\boldsymbol{X}_S$ given by

$$\boldsymbol{X}_S = \begin{bmatrix} \boldsymbol{\mu}_L & \boldsymbol{\nu}_L & \boldsymbol{c}_L & \boldsymbol{\gamma}_L & \boldsymbol{\kappa}_L \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\mu}_t & \boldsymbol{\nu}_t & \boldsymbol{c}_t & \boldsymbol{\gamma}_t & \boldsymbol{\kappa}_t \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\mu}_N & \boldsymbol{\nu}_N & \boldsymbol{c}_N & \boldsymbol{\gamma}_N & \boldsymbol{\kappa}_N \end{bmatrix} \tag{21}$$

where $\boldsymbol{\mu}_t = [\mu_t(1)\cdots\mu_t(M)]$, $\boldsymbol{\nu}_t = [\nu_t(1)\cdots\nu_t(M)]$, $\boldsymbol{c}_t = [c_t(1)\cdots c_t(M)]$, $\boldsymbol{\gamma}_t = [\gamma_t(1)\cdots\gamma_t(M)]$ and $\boldsymbol{\kappa}_t =$

$[\kappa_t(1) \cdots \kappa_t(M)]$. As the statistics matrix $\boldsymbol{X}_S$ has a higher dimension than the original data matrix, PCA is executed to reduce data dimension before discriminant analysis. PCA is capable of producing the uncorrelated low-dimensional principal component matrix $\boldsymbol{T}_S$ and the decomposition is expressed as

$$\boldsymbol{X}_S = \boldsymbol{T}_S \boldsymbol{Q}^{\mathrm{T}} + \boldsymbol{E} \tag{22}$$

where $\boldsymbol{Q}$ is the loading matrix and $\boldsymbol{E}$ is the residual matrix. In the PCA decomposition procedure, the principal component matrix $\boldsymbol{T}_S$ is required to retain 99.9% information of the statistics matrix $\boldsymbol{X}_S$. Further explanations for the PCA algorithm can be found in the related references [2], [10], [11].

After the completion of the first step of SPA, we obtain the statistic training matrix $\boldsymbol{T}_S$. In the second step of LFDA, the LFDA is implemented on this statistic training matrix and the constructed discriminant model is called the SLFDA discriminant model.

We can now summarize our SLFDA based fault classification, which involves two stages: offline training and online testing. In the offline training, the SLFDA is performed on the training fault datasets to build the discriminant function for each class of fault. During the online testing, new fault data is collected and projected onto the discriminant model. The class of the new fault data is determined by comparing the test discriminant functions of all the known fault classes.

V. A SIMULATION CASE STUDY

The proposed SLFDA based fault classification approach is evaluated using the well-known benchmark Tennessee Eastman (TE) process. The TE process whose flowchart is depicted in Fig. 2 simulates a real chemical process, which involves five units: reactor, condenser, compressor, separator and stripper. The TE process simulator was firstly built by Downs and Vogel [25] and has been widely used for validating various process monitoring and fault diagnosis techniques [4], [6], [10], [12], [26]. With the TE simulator, operation datasets for normal operation mode and 21 classes of fault modes have been generated. The data of each mode contains 52 process variables, including 11 manipulated variables and 41 measurement variables. Details of this TE process, including
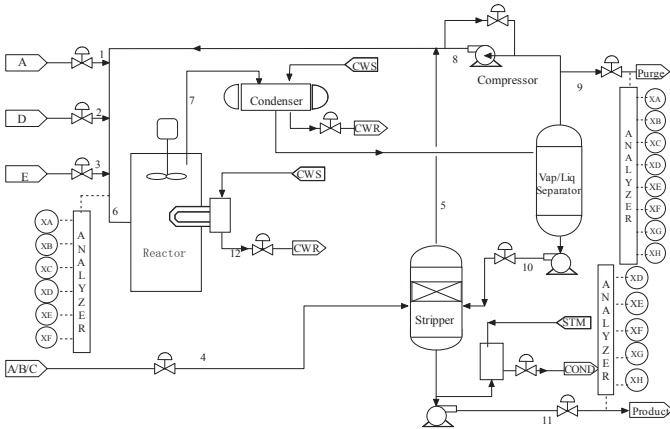

Fig. 2: Flow chart of the Tennessee Eastman process.

the descriptions of process variables and faults, can be found in [11], [25]. In this simulation study, we collect 480 samples for each fault as training data and 400 different samples for each fault as testing data. We compare the performance of the proposed SLFDA with those of the FDA and LFDA.

First, the classification of faults 3, 4 and 11 is illustrated. The training data includes 1440 samples with 480 data points from each of the three faults, while the testing dataset contains 1200 samples with the first 400 testing samples coming from fault 3, the second 400 testing samples from fault 4 and the last 400 testing samples from fault 11. The FDA, LFDA and SLFDA are applied to the training data to build their respective discriminant models. For the SLFDA, the time window L for calculating statistics is set to 30. The classification results on the testing data obtained by the three methods are now shown. For a clear visualization of the test results, the first two discriminant components y_1 and y_2 obtained by each method are plotted in Fig. 3, where each ellipse represents 99% confidence region for the corresponding fault data. The results of the FDA shown in Fig. 3 (a) exhibits serious overlaps among the three faults. Thus, the FDA is unable to classify these three faults. From the LFDA based results shown in Fig. 3 (b), it can be seen that fault 3 and fault 4 are separated clearly but both overlap with fault 11. Therefore, although some improvement over the FDA is achieved, the LFDA is unable to classify all these three faults satisfactory. By contrast, the results of the proposed SLFDA depicted in Fig. 3 (c) indicate that all the three faults are clearly separated, and the SLFDA provides the best fault classification results.

The y_1-y_2 plots given in Fig. 3 only reflect the position difference among different fault classes. However, the classification accuracy is also influenced by the covariance structures of fault data. Therefore, we further analyze the discriminant function outputs of the three methods plotted in Fig. 4. As three classes of faults 3, 4 and 11 are used to build discriminant models, there are three discriminant functions for each test sample, which are correspondingly denoted as $f_1(\boldsymbol{x})$, $f_2(\boldsymbol{x})$ and $f_3(\boldsymbol{x})$. The maximum discriminant function output among the three discriminant functions indicates the estimated fault class of the test sample, which is illustrated in Fig. 5.

According to the FDA chart of Fig. 4 (a), the three discriminant function curves are intertwined for the first 400 samples and the last 400 samples, which leads to large numbers of misclassified samples for faults 3 and 11, as can be clearly seen from Fig. 5 (a). For the LFDA method shown in Fig. 4 (b), the situation is better than the case of the FDA but the discriminant function outputs for the last 400 samples are still seriously intertwined, while for the first 400 samples, $f_1(\boldsymbol{x})$ and $f_3(\boldsymbol{x})$ are intertwined. This results in many misclassified samples as indicated in Fig. 5 (b). When the SLFDA is applied, it provides excellent discrimination results as shown in Fig. 4 (c), which can clearly and correctly identify the fault classes of testing samples, as confirmed in Fig. 5 (c).

The fault misclassification rates obtained by the three methods for classifying faults 3, 4 and 11 are listed in Table I. For the FDA method, fault 4 is recognized reasonably well with
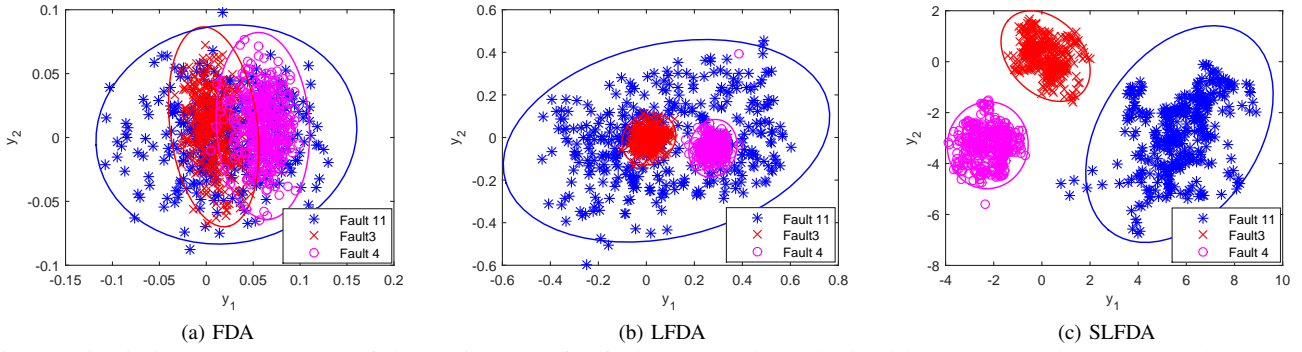
Fig. 3: Discriminant components of the testing data for faults 3, 4 and 11 obtained by (a) FDA, (b) LFDA, and (c) SLFDA.
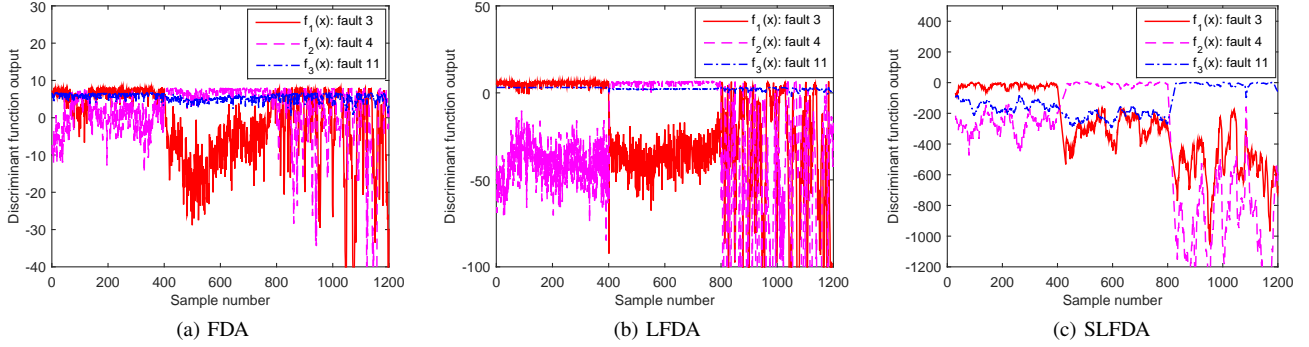

Fig. 4: Discriminant functions of the testing data for faults 3, 4 and 11 obtained by (a) FDA, (b) LFDA, and (c) SLFDA.
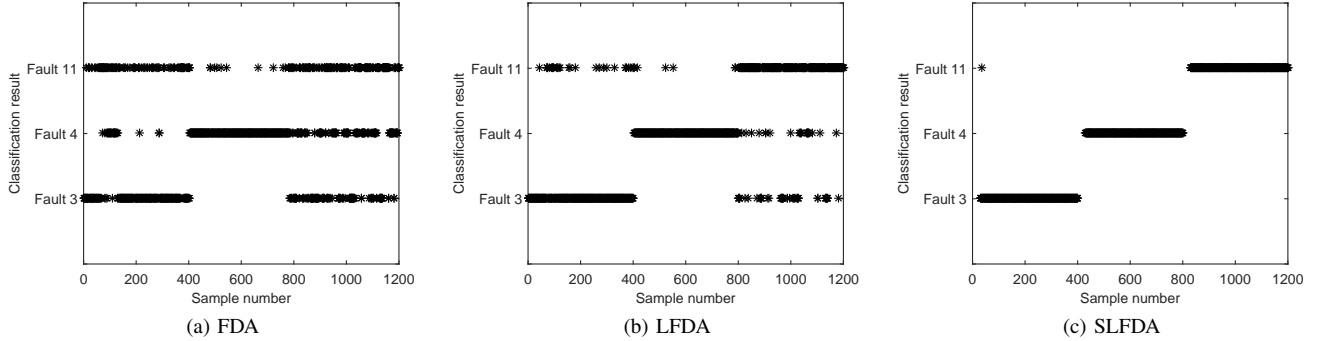

Fig. 5: Classification results of the testing data for fault 3, 4 and 11 obtained by (a) FDA, (b) LFDA, and (c) SLFDA

a misclassification rate of 7.25% but its fault misclassification rates for faults 3 and 11 are very high, reaching 33.75% and 61.5%, respectively. For the LFDA method, most of the testing samples from faults 3 and 4 are identified reasonably well and its misclassification rates for faults 3 and 4 are 7% and 2.25%, respectively. But its misclassification rate for fault 11 is very high, reaching 17.5%. By contrast, our SLFDA correctly recognizes all the test samples of faults 4 and 11, while its

TABLE I: Fault misclassification rates (%) of testing samples for faults 3, 4 and 11 obtained by the FDA, LFDA and SLFDA.

Method	Misclassification rate of testing samples			
	Samples No. 1-400	Samples No. 401-800	Samples No. 801-1200	All samples
FDA	33.75	7.25	61.50	34.17
LFDA	7.0	2.25	17.5	8.92
SLFDA	0.27	0	0	0.09

misclassification rate for fault 3 is only 0.27%. Clearly, the proposed SLFDA offers the best fault discrimination capability for classifying faults 3, 4 and 11.

Next, all the 21 faults are used to investigate the classification performance of the three methods. The average misclassification rates of the FDA, LFDA and SLFDA over the test data of all the 21 faults are plotted in Fig. 6 as the functions of the number of discriminant components selected. As expected, the LFDA has a lower fault misclassification rate than the FDA, and the SLFDA achieves the lowest misclassification rate among these three methods.

VI. CONCLUSIONS

This paper has developed a SLFDA method, which integrates the SPA and the LFDA to provide better fault classification performance. In particular, the SPA has been applied
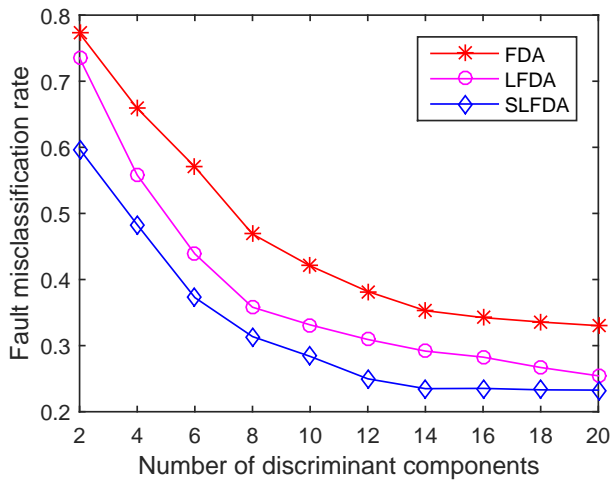
Fig. 6: Fault misclassification rates of the three methods as functions of the number of discriminant components used, averaged over the test data of all the 21 faults.

to mine intrinsic statistical information from the process variables, including higher-order statistics, while the LFDA has been utilized to handle local structure preserving for enhancing classification. The testing results on the benchmark Tennessee Eastman process have shown that the proposed SLFDA provides significantly better fault classification capability than the FDA and LFDA methods. There are two key issues deserving further research. One is how to choose statistics optimally for fault discrimination, and the other is to investigate the influence of statistic window width on classification results. The study on these two problems can help realizing the full potential of the SLFDA.

References

[1] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annual Reviews in Control*, vol. 36, no. 2, pp. 220–234, Dec. 2012.

[2] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Industrial & Engineering Chemistry Research*, vol. 52, no. 10, pp. 3543-3562, Feb. 2013.

[3] S. Yin, X. Li, H. Gao, and O. Kaynak, "Data-based techniques focused on modern industry: an overview," *IEEE Trans. Industrial Electronics*, vol. 62, no. 1, pp. 657–667, Jan. 2015.

[4] 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 and Intelligent Laboratory Systems*, vol. 127, pp. 195–209, Aug. 2013.

[5] Y. Zhang, W. Du, Y. Fan, and L. Zhang, "Process fault detection using directional kernel partial least squares," *Industrial & Engineering Chemistry Research*, vol. 54, no. 9, pp. 2509–2518, Feb. 2015.

[6] X. Tian, X. Zhang, X. Deng, and S. Chen, "Multiway kernel independent component analysis based on feature samples for batch process monitoring," *Neurocomputing*, vol. 72, nos. 7-9, pp. 1584–1596, Mar. 2009.

[7] P.-E. P. Odiowei and Y. Cao, "Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations," *IEEE Trans. Industrial Informatics*, vol. 6, no. 1, pp. 36–45, Feb. 2010.

[8] S. J. Zhao, J. Zhang, and Y. M. Xu, "Performance monitoring of processes with multiple operating modes through multiple PLS models," *J. Process Control*, vol. 16, pp. 763–772, Aug. 2006.

[9] H. Huang, H. Feng, and C. Peng, "Complete local Fisher discriminant analysis with Laplacian score ranking for face recognition," *Neurocomputing*, vol. 89, no. 15, pp. 64–77, Jul. 2012.

[10] L. H. Chiang, E. L. Russell, and R. D. Braatz, "Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 50, no. 2, pp. 243–252, Mar. 2000.

[11] L. H. Chiang, E. L. Russell, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*. Springer, 2001.

[12] B. Jiang, X. Zhu, D. Huang, J. A. Paulson, and R. D. Braatz, "A combined canonical variate analysis and Fisher discriminant analysis (CVA-FDA) approach for fault diagnosis," *Computers & Chemical Engineering*, vol. 77, pp. 1–9, Jun. 2015.

[13] S. Zhong, Q. Wen, and Z. Ge, "Semi-supervised Fisher discriminant analysis model for fault classification in industrial processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 138, pp. 203–211, Nov. 2014.

[14] Z. Zhu and Z.-H. Song, "A novel fault diagnosis system using pattern classification on kernel FDA subspace," *Expert Systems With Applications*, vol. 38, no. 6, pp. 6895–6905, Jun. 2011.

[15] J. Li and P. Cui, "Improved kernel fisher discriminant analysis for fault diagnosis," *Expert Systems With Applications*, vol. 36, no. 2, pp. 1423–1432, Mar. 2009.

[16] C. Zhao and F. Gao, "A nested-loop Fisher discriminant analysis algorithm," *Chemometrics and Intelligent Laboratory Systems*, vol. 146, pp. 396–406, Aug. 2015.

[17] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," *J. Machine Learning Research*, vol. 8, pp. 1027–1061, 2007.

[18] J. Yu, "Localized Fisher discriminant analysis based complex chemical process Monitoring," *AIChE J.*, vol. 57, no. 7, pp. 1817–1828, Jul. 2011.

[19] S. Ren, Z. Song, M. Yang, and J. Ren, "A novel multimode process monitoring method integrating LCGMM with modified LFDA," *Chinese J. Chemical Engineering*, vol. 23, no. 12, pp. 1970–1980, Dec. 2015.

[20] M. Van and H.-J. Kang, "Wavelet kernel local Fisher discriminant analysis with particle swarm optimization algorithm for bearing defect classification," *IEEE Trans. Instrumentation and Measurement*, vol. 64, no. 12, pp. 3588–3600, Dec. 2015.

[21] F. Li, J. Wang, M. K. Chyu, and B. Tang, "Weak fault diagnosis of rotating machinery based on feature reduction with supervised orthogonal local Fisher discriminant analysis," *Neurocomputing*, vol. 168, pp. 505–519, Nov. 2015.

[22] 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.

[23] J. Wang and Q. P. He, "Multivariate statistical process monitoring based on statistics pattern analysis," *Industrial & Engineering Chemistry Research*, vol. 49, no. 17, pp. 7858–7869, 2010.

[24] X. Deng and X. Tian, "Nonlinear process fault pattern recognition using statistics kernel PCA similarity factor," *Neurocomputing*, vol. 121, pp. 298–308, Dec. 2013.

[25] J. Downs and E. F. Vogel, "A plant-wide industrial problem process," *Computers & Chemical Engineering*, vol. 17, no. 3, pp. 245–255, Mar. 1993.

[26] C. Lau, K. Ghosh, M. Hussain, and C. R. Che Hassan, "Fault diagnosis of Tennessee Eastman process with multi-scale PCA and ANFIS," *Chemometrics and Intelligent Laboratory Systems*, vol. 120, pp. 1–14, Jan. 2013.