Proceedings of the 59th Meeting of the Society for Machine Failure Prevention Technology, MFPT '05, pp 167-176, 2005.

FEATURE DIMENSIONALITY REDUCTION FOR PARTIAL DISCHARGE DIAGNOSIS OF AIRCRAFT WIRING

Weizhong Yan, Kai F. Goebel

GE Global Research Center
Niskayuna, NY 12309
{yan, goebelk} @ crd.ge.com

Abstract: Partial discharge (PD) analysis has been successfully used as a means to evaluate the integrity of insulation systems, especially high voltage electrical systems such as generators, transformers, and capacitors. In recent years, PD has also gained increasing attention for its application in low voltage insulation systems such as aircraft wiring systems. Besides detecting and acquiring discharge signals/pulses with sufficiently high measuring sensitivities, one of the challenges of using PD for insulation fault diagnosis is to accurately interpret the acquired signals. That is, to associate the PD signals with real physical phenomena and specifically the states of the insulation that is being monitored. Like in any diagnostic/classification system, the key to an accurate and reliable PD diagnosis is a set of high quality features/attributes that represent/capture the characteristics of PD signals. More importantly, these features must posses strong discriminant power so that the classifier designed based on those features gives desired performance. This paper is concerned with the application of linear feature transformation techniques for reducing the number of features that are extracted by collectively using different feature extraction methods. More specifically, this paper investigates the effectiveness of the two well-studied linear feature transformation methods, namely, principal component analysis (PCA) and linear discriminant analysis (LDA), in improving the classification performance of PD diagnostic systems. We apply these techniques to experimental partial discharge data from an ongoing study of aircraft wiring diagnostics. We also compare the results of this study with those from using other feature selection techniques.

Keywords: Dimensionality reduction; Feature selection; Partial discharge; PD diagnosis; Classifier; Principal component analysis; Linear discriminant analysis

1. Introduction: Partial discharges (PD) are a local, partial breakdown event that occurs for example, on the surface or inside insulation of electrical products due to possibly minute defects in insulation structure. These minute defects may be the result of the manufacturing process and/or the result of ageing and mechanical damage of the products. While normal (healthy) condition of insulation gives a baseline level of partial discharge activities, increase of partial discharge activities indicates insulation degradation or faults. Diagnosing insulation defects/faults based on partial discharge activities, generally known as “PD diagnosis” [Gulski (1995)], has played a critical role in condition-based maintenance of insulation systems, especially high voltage insulation systems, such as generators, transformers, and capacitors.

PD diagnosis, a typical classification problem, is to classify measured PD activities into the underlying insulation defects or sources that generate PDs. PD diagnosis is a complex classification problem because 1) PD is inherently a stochastic process, namely, the occurrence of PD very much depends on many factors, such as temperature, pressure, applied voltage, and the test duration [Gulski (1995)], and 2) PD signals contain noise and interference. Consequently, like in any diagnostic/classification systems, the key to an accurate and reliable PD diagnostic system is to identify a set of high quality features/attributes. These features should represent/capture the characteristics of PD signals. More importantly, these features must posses strong discriminant power so that the classifier designed based on those features has the desired performance.

Motivated by finding salient features for PD diagnosis, researchers have introduced several different feature extraction methods, including the widely used statistical analysis of phase-resolved PD patterns and more modern methods, such as, fractal analysis [Gulski (1996)] and textual analysis [Rahman et al (2000)], among others. However, so far no single feature extraction method has been proved to be effective for *all* problems. In fact, effectiveness of features from those individual feature extraction methods on classification is highly problem-dependent. Therefore, in designing PD diagnostic systems, the designer still faces a great challenge in deciding which feature extraction method(s) is more appropriate for the given problem at hand and, ultimately, in how to find a set of features that is optimal for the problem concerned in terms of classification performance. Individually evaluating those different feature extraction methods for the given problem is not only time-consuming, but also unreliable since it fails to take into consideration of feature interaction. To tackle this challenge, in our previous paper [Yan & Goebel (2005)], we proposed a novel approach that consists of two steps. First collectively utilize all different feature extraction methods, without discerning how good they are for the problem, to generate a feature pool; then apply feature selection to the feature pool to choose the optimal subset of features. The two-step approach not only gives an optimal set of features, but also alleviates the designer from the difficulties of finding appropriate feature extraction methods. In this paper, we follow the two-step approach; however, we investigate using linear feature transformation in step 2 to replace feature selection to find a small number of features that are most discriminative. Linear feature transformation has been widely used in various fields for feature extraction and feature dimensionality reduction. However, linear feature transformation, especially, linear discriminative analysis (LDA), has not been actively studied by the researchers in the field of PD diagnosis. This paper demonstrates the effectiveness of linear feature transformation in improving performance of PD diagnostic systems.

This paper is organized as follows. Section 2 gives an introduction to linear feature transformation methods (principal component analysis and linear discriminant analysis). Section 3 describes the exemplar case that we used for the demonstration and the feature extraction methods used for extracting features from PD signals. Classification results of the example case study using PCA and LDA are presented and compared in Section 4. Section 5 concludes the paper.

2. Feature dimensionality reduction: Feature dimensionality reduction provides several important benefits in classifier design, including: 1) reducing the computational complexity of classifiers; 2) providing a remedy to "curse of dimensionality"; and 3) improving the generalization of classifiers. As a result, feature dimensionality reduction is almost always a necessary step in the design of classification systems. While there are several different methods for feature dimensionality reduction [Fodor (2002)], linear feature transformation is regarded as the most attractive one since it is "simple to compute and analytically tractable" [Duda et al (2000)]. Linear feature transformation reduces feature dimensionality by representing the original features with a smaller number of new features that are the linear combinations of the original ones. Graphically it projects the data from a high dimensional feature space to a low dimensional one.

Mathematically, linear transformation has the general form of $\mathbf{Y} = \mathbf{W}^{\mathbf{T}}\mathbf{X}$, where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_\mathbf{n}] \in \Re^p$ is a data matrix formed by n p-dimensional vectors, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_\mathbf{n}] \in \Re^d$ is the transformed data matrix containing n d-dimensional vectors in the new feature space, and $\mathbf{W} \in \Re^{p \times d}$ is the transformation matrix. For feature dimensionality reduction, $d \le p$ and preferably $d << p$ if the original feature dimension (p) is high. The transformation matrix, $\mathbf{W}$, is obtained by optimizing a specific objective/criterion function. It is the objective/criterion function used for determining the transformation matrix that distinguishes one transformation method from the others. In this paper, we are specifically interested in two most popular linear transformation methods, namely, principal component analysis (PCA) and linear discriminant analysis (LDA). From the standpoint of the mathematical form, both PCA and LDA are the same. The difference of the two methods comes from the criterion functions used for determining the transformation matrix, $\mathbf{W}$. A brief explanation of the two methods is given as follows.

PCA is a classical statistical technique and has been widely used in a range of applications. For feature dimensionality reduction, PCA attempts to find a smaller set of variables that best describe/represent the given data in terms of representation error. Following the above-defined notations, the criterion function for obtaining the transformation matrix, $\mathbf{W}$, in PCA is $J_{PCA}(\mathbf{W}) = \mathrm{E}(\|\mathbf{WY} - \mathbf{X}\|^2)$. It has proved that the transformation matrix, $\mathbf{W}$, that minimizes $J_{PCA}(\mathbf{W})$ is formed by the eigenvectors corresponding to the d largest eigenvalues of the scatter matrix, $\mathbf{S} = E[(\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^{\mathbf{T}}]$, where $\mathbf{m} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$.

LDA projects the variables to a feature space in such directions that provide optimal separation among classes. Class separability is typically quantified as a function of the between-class scatter, $\mathbf{S}_\mathbf{B}$, and the within-class scatter, $\mathbf{S}_w$, as defined by Duda et al

(2000). Therefore, LDA attempts to determine the transformation matrix, $\mathbf{W}$, by maximizing the Fisher's criterion function defined as $J_{LDA}(\mathbf{W}) = \frac{|\mathbf{W}^T \mathbf{S_B} \mathbf{W}|}{|\mathbf{W}^T \mathbf{S_W} \mathbf{W}|}$ [Duda et al (2000)].

It has been proved that finding $\mathbf{W}$ that maximizes $J_{LDA}(\mathbf{W})$ is equivalent to solving a generalized eigenvalue problem $\mathbf{S_B W} = \lambda \mathbf{S_W W}$, which yields a maximum of $c-1$ (c is the number of classes) eigenvectors if $\mathbf{S}_w$ is nonsingular. Those eigenvectors corresponding to the d largest eigenvalues form the columns of the transformation matrix, $\mathbf{W}$.

From the above-defined objective functions, we can see that LDA finds new features by maximizing discriminating power between classes, while PCA attempts to best represent the entirety of the data without paying any attention to the underlying class structure. As the result, it is intuitive and generally accepted that for classification problems, LDA should perform better that PCA does. Recently, however, some empirical studies, e.g., Martinez and Kak (2001), have shown that when the training data set is small, PCA can outperform LDA. Since LDA requires class information for all data examples, it is a supervised method. On the other hand, PCA is an unsupervised method.

3. A case study: The case studied in this paper is the design of a PD diagnostic system for aircraft wiring. It is a part of our multiple-generation initiative project funded by ONR. Presented here as an example case study is the initial design of an on-ground, de-energized prototype system.

Experiment test for generating PD pulses: PD measurements for aircraft wiring are generated and recorded through laboratory tests. Damage to wiring insulation generally takes two forms: material degradation due to aging or thermal/electrical environment, and chafing that may occur during maintenance and mechanical abrasion during operation-induced vibration. The lab tests conducted in this paper focus on the later, i.e., wire chafing. For classifier design purpose, two wiring conditions are tested. One is for normal condition wires and another for wires with artificial chafing. Two different ways are considered in producing artificial defects. The first type represents defects occurring in twisted pair wires. For that a small piece of the upper insulation layer was removed from both wires in the twisted pair. This type of defect is the most reproducible one since the thickness of the remaining insulator layer was not changed. Samples for this type were prepared from two aircraft grade wires, types M22759/90-22-95 and M22759/81-22-52. The second type of defect simulated a chafed wire touching the shielding of the cable bundle or a metal part of the aircraft. The wire was chafed in a short length and a piece of tinned copper wire was twisted around the chafed area. The tinned copper wire was also pressed into the chafed section in order to touch the remained insulation layer. Such samples were prepared from three aircraft grade wires, types M22759/90-22-95, M22759/81-22-52 and M81044/6-22-9.

A total of nine wire samples are tested with maximum of 10 repetitions for each sample. After data cleansing (including the removal of noise and incomplete data sets), 596 PD sequences are used for designing the PD diagnostic system. Out of the 596 PD sequences, 225 are for normal wires and 371 for chafed ones.

Feature extraction: We use the following five different methods for extracting features from the PD measurements/signals.

A) Features from statistical analysis of phase-resolved PD patterns. Phase-resolved PD pattern analysis is the most commonly used method for feature extraction in PD diagnosis. Given a sequence of PD pulses and the recorded voltage phase angles at corresponding pulse peaks, a 3D PD pattern is generated, where the number of pulses (pulse count) is plotted as a function of magnitude and phase of the PD pulses. A typical 3D PD pattern is shown in Figure 1. 3D PD patterns are a good representation/summary of all PD pulses recorded within a specified time window and should show different characteristics for different PD activities, thus different PD sources. For the convenience of statistical analysis, the 3D patterns are decomposed into two 2D distributions by projecting it into the two axes - phase and magnitude. Statistical analysis is performed separately for those two distributions. Also, statistical analysis is performed separately for phase angles from 0° to 180° ("positive" PDs), for phase angles from 180° to 360° ("negative" PDs), and on the difference between positive and negative PDs. For each of the distributions, two types of statistics, names amplitude statistics and shape statistics, are calculated. The statistical descriptors are mean, standard deviation, skewness and kurtosis. In addition, overall maximum magnitudes of positive and negative PDs and correlation between positive and negative PD patterns are also calculated as features.

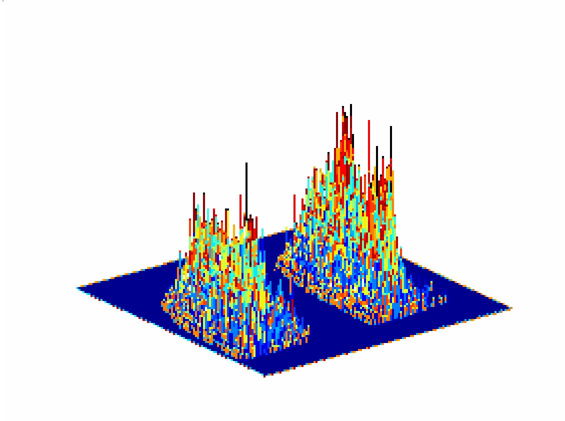

Figure 1: An example of 3D PD patterns

B) Features from PD height distribution analysis. Heights (peaks) of a sequence of PD pulses can be represented in a histogram that shows number of pulses as a function of their magnitude. According to Cacciari et al (2002), PD pulse height distribution tends to fit well with the two-parameter Weibull function defined as: $F(q) = 1 - \exp(-(q/\alpha)^\beta)$, where q is the pulse height, α and β are the scale and shape parameters of the Weibull function. They have found that the scale and shape (especially shape) parameters differ with different PD sources, thus can be used as features for PD identification or classification.

C) Features from "classification map". PD pulses are different in wave shape depending on the location and nature of the underlying defect that generates PD. One way to capture the different wave shapes is to use so-called "equivalent time-length", T^2, and "equivalent bandwidth", W^2 [Contin et al (2000)].

In the $T^2 - W^2$ plane (also called the "classification map" by [Contin et al (2000)]), each PD pulse is presented as a point and each sequence of PD pulses, which are similar in shape, fall into a well-defined area (cluster) in the $T^2 - W^2$ plane. The location and shape of the clusters in the $T^2 - W^2$ plane differ corresponding to different PD sources [Contin et al (2000 & 2002)]. In this paper, characteristics of the clusters are extracted through statistical analysis and used as features for classification purpose. The features extracted include overall mean, means and standard deviations in both T^2 and W^2 directions, respectively, 1st through 4th orders of moments of distributions in both T^2 and W^2 directions, respectively, direction of the 1st eigenvector of the cluster, and ratio of the first two eigenvalues of the cluster.

D) Features from spectrum analysis. Frequency spectrum of a PD pulse indicates frequency components of the PD pulse. Thus the shape or distribution of frequency spectra should be correlated with different PD sources. In this paper, the first 3 frequencies corresponding to the highest three magnitudes, the three highest magnitudes themselves, the difference between the three frequencies, and the difference between the three magnitudes are used as features.

E) Features from raw PD signals. They are the maximum and minimum peaks of PD pulses, mean and standard deviation of peaks of PD pulses, inception voltage, and PD rate.

The five feature extraction methods yield 86 features in total. Including the temperature and humidity measurements, we have 88 features for each of the 596 data samples, which we can use for feature transformation study.

4. Classification results: PD diagnosis is a classification problem where the extracted features from PD measurements are the inputs and the sources of PD or condition status of the wire monitored are the class targets. There are a great number of classifiers available, ranging from traditional statistical methods to more modern methods (such as neural network classifiers and support vector machines (SVMs)). To study the effectiveness of feature transformation on improving classification performance of PD diagnostic systems, in this paper, we use support vector machines as the classifier for diagnosing aircraft wiring faults. SVMs are a recently developed learning system originated from the statistical learning theory [Vapnik (1995)]. One distinction between SVMs and many other learning systems is that its decision surface is an optimal hyperplane in a high dimensional feature space. The optimal hyperplane is mathematically found by solving a properly formed convex quadratic problem with optimization theory, which is well studied in the field of mathematical programming and can be solved in a relatively straightforward way. For the classification problem we are concerned in this paper, two of these unique properties, namely, better handling of sparse data and good average performance over a wide spectrum of different classification problems make SVM classifiers a better choice here.

To evaluate classification performance, the data set is randomly split into two disjoint subsets: one for training and another for evaluation. Also, to improve the robustness of evaluation, the SVM classifier is trained and evaluated 10 times and each time the data is randomly split into two disjoint subsets. The mean, the standard deviation, the minimum, and the maximum of accuracies of the 10 evaluations are reported here. Additionally, 1-sigma random noise is added to the original feature values in order to account for the facts that only limited PD examples are obtained in well-defined lab environment and that PD measurements in real aircraft wires will be much more noisy.

The PCA results: First, we apply PCA to the 88-feature data set described in the previous section. Figure 2 shows the variance explained by each of the principal components (PCs). For illustration convenience, only the first 30 PCs with largest variance explained are shown in the figure. The vertical bars show the percentage of the total variance explained by each individual principal component, while the line with dots shows the cumulative percentage of the total variance explained by all principal components up to the specific one.

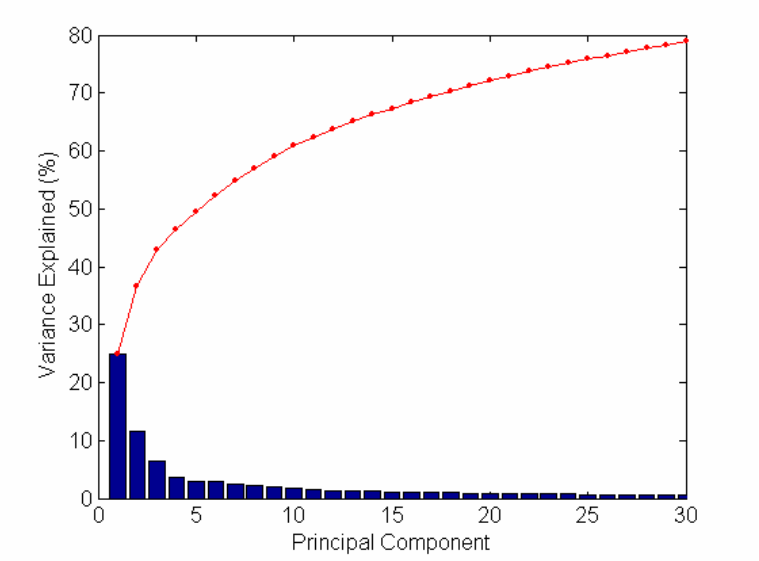

Figure 2: Variance explained by the first 30 principal components

As we can see from the figure, the first 10 PCs explain approximately 60% of the total variance of the data whereas the first 30 PCs explain about 80% of the total variance. Classification accuracies for the designs using the first 3, 10, and 30 PCs as classifier inputs are shown in Table I, respectively. For comparison, also included in Table I are the classification accuracies for the design with all 88 features extracted from the PD signals as its inputs. From Table I, we can see that using only the first 3 PCs as the classifier inputs already provides a reasonable performance – better than that from using all 88 features. Using the first 10 PCs leads to a classification performance improvement. However, further increasing the number of PCs from 10 to 30 does not gain anything in classification performance.

Table I: Classification accuracies summary

		Classification accuracy			
Methods		**Mean**	**Std. Dev.**	**Min.**	**Max.**
All 88 features		0.9566	0.0206	0.9224	0.9811
PCA	**3**	0.9644	0.0055	0.9560	0.9706
	10	0.9839	0.0044	0.9748	0.9874
	30	0.9813	0.0036	0.9748	0.9874
LDA		0.9906	0.0048	0.9832	0.9958

The LDA results: The PD diagnostic system concerned in this paper is a 2-class classification problem, with the 2 classes being either normal or chafed condition of aircraft wires. As discussed in Section 2, the maximal number of new features in LDA space is equal to the number of classes minus one. Thus applying LDA to the 88 features extracted in Section 3 results in a single feature in the LDA space. The distributions of the single LDA feature over the 2 wiring conditions (normal and chafed) are show in Figure 3. It clearly shows a good separation between the 2 conditions/classes, which means that the LDA feature carries high discriminant power, thus expects to give a good classification performance.

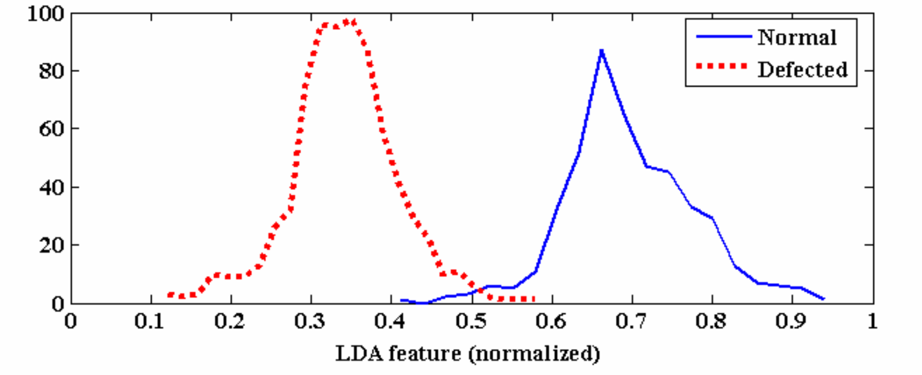

Figure 3: Feature value distribution of the LDA feature

The classification accuracies for the design using the LDA feature are also shown in Table I (last row). Using the single LDA feature achieves the highest classification accuracies.

From Table I, we can see that both PCA and LDA not only dramatically reduce the number of features needed by the classifier, but also improve the classification performance for the PD diagnosis problem. Between the two linear transformation methods (PCA and LDA), LDA seems to be more effective not only in reducing feature dimensionality, but also in improving classification performance. This observation, i.e., LDA is better that PCA in classification, is accordance with the intuition. The observation from Table I that both PCA and LDA use a few features while showing better performance than the design using all 88 features indicates that there exist a great number of redundant features among the original 88 features, which is expected since the 5 different feature extraction methods used are not necessarily independent.

5. Conclusions: In designing PD diagnostic systems and any other diagnostic systems alike, finding an optimal feature set with a smaller number of features is important. This paper is concerned with using linear feature transformation to reduce feature dimensionality. More specifically, we investigate the effectiveness of linear feature transformation on improving the classification performance of PD diagnostic systems. By using an example of designing a PD diagnostic system for aircraft wire fault diagnosis, this paper demonstrates that both PCA and LDA can be effective not only in reducing feature dimensionality, but also in improving the classification performance, in the design of PD diagnostic systems for aircraft wiring. Therefore, linear feature transformation (PCA and LDA), especially when working together with our proposed 2-step feature extraction approach, can be potentially useful in design of PD diagnostic systems. This paper also demonstrates that LDA outperforms PCA for PD diagnosis because the underlying objective for LDA is specifically to maximize class separability.

ACKNOLEGEMENT: The authors gratefully acknowledge the guidance and support of Nicole Evers, Janos Sarkozi and Karim Younsi at GE Global Research Center and of Sean Field, Charles Fulbright, and Greg Nepini at NavAir. This work was funded by ONR under contract #: N00014-02-C-0402.

REFERENCES:

1. Gulski, E. (1995), “Digital analysis of partial discharges”, IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 2, No. 5, pp822-37.
2. Gulski, E., Burger, H.P., Zielonka, A., and Brooks, R. (1996), “Classification of defects in HV components by fractal analysis of PD measurements”, IEEE Annual Report Conference on Electrical Insulation and Dielectric Phenomena, San Francisco, October 20-23, 1996
3. Abdul Rahman, M.K., Arora, R., and Srivastava, S.C. (2000), “Partial discharge classification using principal component transformation”, IEE Proceedings of Scientific Measurement Technologies, Vol. 147, No. 1, pp7-13
4. Yan, W. & Goebel, K. (2005), “Feature selection for partial discharge diagnosis”, Submitted to SPIE Smart Structures 2005, March 6 –10, San Diego, CA
5. Fodor, I.K. (2002), “A survey of dimension reduction techniques”, Lawrence Livermore National Laboratory (LLNL) technical report, June 2002, UCRL-ID-148494
6. Duda, R. O., Hart, P. E., and Stork, D. G. (2000), *Pattern Classification*, John Wiley & Sons, Inc., New York, NY

7. Contin, A., Cavallini, A., Montanari, G.C., Pasini, G., and Puletti, F. (2000), "Artificial intelligence methodology for separation and classification of partial discharge signals", Proceedings of IEEE Conference on Electrical Insulation and Dielectric Phenomena", Victoria, Canada, pp522-26.
8. Contin, A., Cavallini, A., Montanari, G.C., Pasini, G., and Puletti, F. (2002), "Digital detection and fuzzy classification of partial discharge signals", IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 9, No. 3, pp335-48
9. Vapnik, V. (1995), *The nature of statistical learning theory*, Springer-Verlag, New York, 1995
10. Martinez, A.M. and Kak, A.C. (2001), "PCA versus LDA", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 2, pp228-233.