HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Engineering Science and Technology, an International Journal

journal homepage: http://ees.elsevier.com/jestch/default.asp

JESTECH

Full length article

Comparison of dimensionality reduction techniques for the fault diagnosis of mono block centrifugal pump using vibration signals

N.R. Sakthivel [a,*], Binoy B. Nair [b], M. Elangovan [a], V. Sugumaran [c], S. Saravanmurugan [a]

[a] Department of Mechanical Engineering, Amrita Vishwa Vidyapeetham, Amrita School of Engineering, Coimbatore, Tamilnadu, India
[b] Department of Electronics and Communication Engineering, Amrita Vishwa Vidyapeetham, Amrita School of Engineering, Coimbatore, Tamilnadu, India
[c] School of Building and Mechanical Sciences, Vellore Institute of Technology, Chennai, Tamilnadu, India

ARTICLE INFO

ABSTRACT

Bearing fault, Impeller fault, seal fault and cavitation are the main causes of breakdown in a mono block centrifugal pump and hence, the detection and diagnosis of these mechanical faults in a mono block centrifugal pump is very crucial for its reliable operation. Based on a continuous acquisition of signals with a data acquisition system, it is possible to classify the faults. This is achieved by the extraction of features from the measured data and employing data mining approaches to explore the structural information hidden in the signals acquired. In the present study, statistical features derived from the vibration data are used as the features. In order to increase the robustness of the classifier and to reduce the data processing load, dimensionality reduction is necessary. In this paper dimensionality reduction is performed using traditional dimensionality reduction techniques and nonlinear dimensionality reduction techniques. The effectiveness of each dimensionality reduction technique is also verified using visual analysis. The reduced feature set is then classified using a decision tree. The results obtained are compared with those generated by classifiers such as Naïve Bayes, Bayes Net and kNN. The effort is to bring out the better dimensionality reduction technique—classifier combination.

1. Introduction

Condition monitoring is used to find faults at an early stage [10] so that diagnosis and correction of those faults can be initiated before they become more prominent and lead to loss in productivity. Unexpected breakdown of mono block centrifugal pump parts increases the down time and maintenance costs. This has motivated academic researchers and industrial experts to focus their attention on such studies using the contemporary techniques and algorithms available in this field.

Bearings fault, impeller fault, seal fault and cavitation [13] can cause serious problems such as noise, high vibration, leakage etc. and degrade the performance of the mono block centrifugal pump [12]. In order to keep the pump performing [35] at its best, different

methods of fault diagnosis have been developed and used effectively to detect the machine faults at an early stage. Vibration analysis is the one of the prime tools to detect and diagnose mono block centrifugal pump faults [11]. Vibration based pump condition monitoring and diagnosis involves data acquisition from the mono block centrifugal pump, feature extraction from the acquired data, feature selection, and interpreting the results.

Different methods and approaches are used for fault diagnosis. Rajakarunakaran et al. [21] proposed a model for the fault detection of centrifugal pumping system using feed forward network with back propagation algorithm and binary adaptive resonance network (ART1) for classification of seven categories of faults in the centrifugal pumping system. In the work reported by Sakthivel et al. [23]; the use of Support Vector Machines (SVMs) and Proximal Support Vector Machines (PSVMs) as a tool for accurately identifying and classifying pump faults was presented. SVM was found to have a slightly better classification capability than PSVM. Sakthivel et al. [24] presented the use of decision tree and rough sets to generate rules from statistical features extracted from vibration signals under good and faulty conditions of a mono block centrifugal pump. A fuzzy classifier is built using decision tree and rough

set rules, tested and the results are compared with those generated by a PCA based decision tree-fuzzy classifier. Sakthivel et al. [25] reported the use of statistical features extracted from time domain signals for classification of faults in centrifugal pumps using decision tree. For the fault classification of mono block centrifugal pump, Sakthivel et al. [26] have used artificial immune recognition system (AIRS). The fault classification efficiency of AIRS is compared with hybrid systems such as PCA-Naïve Bayes and PCA-Bayes Net. AIRS was found to outperform other hybrid systems. Wang and Chen [31] introduced a fault diagnosis method for a centrifugal pump with frequency domain symptom parameter using wavelet transforms for feature extraction, rough sets for rule generation and fuzzy neural network for classification to detect faults and distinguish fault types at early stages. Rafiee et al. [20] illustrated an artificial neural network (ANN)-based procedure for fault detection and identification in gearboxes using a new feature vector extracted from standard deviation of wavelet packet coefficients of vibration signals. The use of vibration signals requires minimum instrumentation but the use of wavelet transforms increases the computational requirements. M. Zhao et al. [42] employed TR-LDA based dimensionality reduction technique for fault diagnosis of rolling element bearings. Y. Zhang et al. [37] proposed an improved manifold learning algorithm by combining adaptive local linear embedding and recursively applying normalised cut algorithm for nonlinear dimensionality reduction, which was then proven to be effective in dealing with standard test data sets as well as on the Tennessee—Eastman process. D.Q. Zhang et al. [36] proposed an efficient dimensionality reduction algorithm called semi supervised dimensionality reduction. Zhang et al. [38–40] presented a novel feature extraction technique called group sparse canonical correlation analysis (GSCCA). M.S. Baghshah et al. [2] proposed a kernel based metric learning technique that can produce nonlinear transformation of the input features resulting in better learning performance. Zhang et al. [38–40] proposed an improvement over the Isomap dimensionality reduction technique called the Marginal Isomap (M-Isomap) which was able to provide better separation of data clusters. F.P. Nie et al. [17] developed an algorithm to find the global optimum for the orthogonal constrained trace ratio problem. Yaguo Lei et al. [33] proposed a system for fault diagnosis of rolling element bearings based on empirical mode decomposition (EMD), an improved distance evaluation technique and the combination of multiple adaptive neuro-fuzzy inference systems (ANFISs). Van Tung Tran et al. [30] discussed a combined fault diagnosis system for induction motor based on classification and regression tree (CART) algorithm and ANFIS. Yaguo Lei [34] proposed fault diagnosis of rotating machinery based on statistic analysis and ANFIS. Necla Togun and Sedat Baysec [16] presented the application of genetic programming (GP) to predict the torque and brake specific fuel consumption of a gasoline engine. In the work reported by Demetgul [5] fault diagnosis of pneumatic systems using ANN was presented.

However, to the best of our knowledge, the comparison of traditional dimensionality reduction techniques with nonlinear dimensionality reduction techniques for fault classification of a mono block centrifugal pump has not been reported so far. This paper investigates traditional dimensionality reduction technique PCA and the following nine nonlinear dimensionality reduction techniques: (1) Kernel PCA, (2) Isomap, (3) Maximum Variance Unfolding, (4) diffusion maps, (5) Locally Linear Embedding, (6) Laplacian Eigenmaps, (7) Hessian LLE, (8) Local Tangent Space Analysis, and (9) manifold charting. These dimensionality reduction techniques are used to transform statistical features extracted from the pump vibration signals. Decision tree, naïve Bayes Bayes Net and kNN classifiers are then used to classify the faults. The main objectives of the work are:

(i) To investigate to what extent the nonlinear dimensionality reduction techniques outperform the traditional PCA on centrifugal pump data sets
(ii) To find out the best dimensionality reduction technique—classifier combination.

2. Experimental studies

2.1. Experimental setup

Fig. 1 shows the schematic diagram of the experimental test rig. The motor (2HP) drives the pump. The control valve is used to adjust the flow at the inlet and the outlet of the pump. The inlet valve is used to create pressure drop between the suction and at the eye of the impeller to simulate cavitation. An accelerometer mounted at the eye of the impeller (location shown in Fig. 1), is used to measure the vibration signals. This is due to the fact that the mechanical components under consideration in the present study (seal, impeller, bearing), are located close to the eye of the impeller.

2.2. Experimental procedure

The vibration signals are acquired from the centrifugal pump working under normal condition at a rated speed of 2880 rpm. Centrifugal pump specification is shown in Table 1. Vibration signals from the accelerometer are measured. The sampling frequency is 24 kHz. 250 sets of readings are taken for each centrifugal pump condition.

In the present study, the following faults are simulated

Bearing fault — Inner and Outer race fault
Seal fault — Broken seal
Impeller fault — Damaged impeller
Bearing and Impeller fault together
Cavitation — at the eye of the impeller

The faults are introduced one at a time and the pump performance characteristic and vibration signals are taken.

3. Feature extraction

Statistical analysis of vibration signals yields different descriptive statistical parameters. The statistical parameters taken for this study are mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum, and sum. These eleven features are extracted from vibration signals.

4. Dimensionality reduction techniques [14]

4.1. Linear dimensionality reduction techniques

4.1.1. Principle component analysis (PCA)

PCA is a linear technique for dimensionality reduction. It performs dimensionality reduction by embedding the data into a linear subspace of lower dimensionality. PCA is one of the most popular (unsupervised) techniques. Therefore, PCA is included in this comparative study.

PCA constructs a low-dimensional representation of the data that describes as much of the variance in the data as possible. This is done by finding a linear basis of reduced dimensionality for the data, in which the amount of variance in the data is maximal. The basic working of a PCA is presented below.
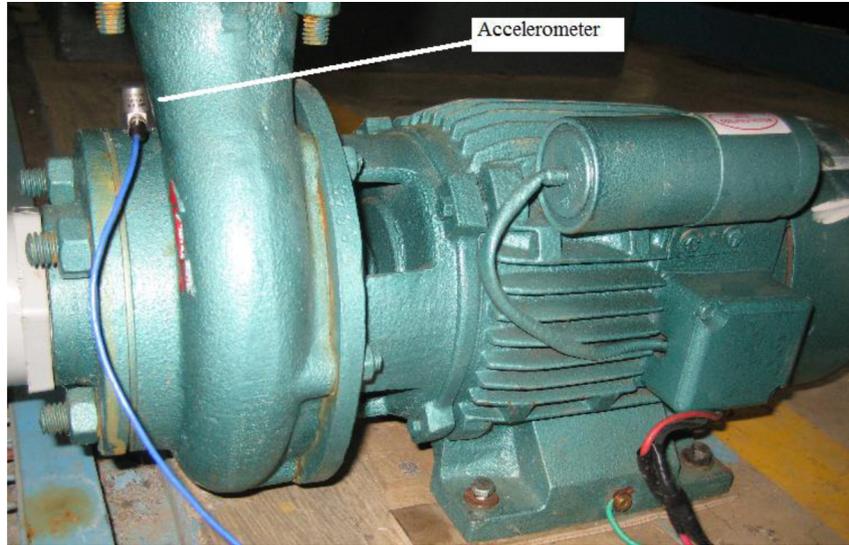
Let $x_1, x_2,..., x_n$ be $N \times 1$ vectors

Fig. 1. Experimental test rig.

Step 1 Mean value $\bar{x}$ is calculated using the equation:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

Step 2 The mean value is subtracted from each feature:

$$\Phi_i = x_i - \bar{x} \tag{2}$$

Step 3 Matrix $A = [\Phi_1, \Phi_2, ..., \Phi_N]$ is generated with and covariance matrix C is computed as follows:

$$C = \frac{1}{M}\sum_{i=1}^{N} \Phi_i \Phi_i^T = AA^T \tag{3}$$

The covariance matrix characterizes the distribution of the data.

Step 4 Eigenvalues are computed as:

$$C = \lambda_1 > \lambda_2 > ...\lambda_N \tag{4}$$

Step 5 Eigenvectors are computed as:

$$C = u_1, u_2, ...u_N \tag{5}$$

Since C is symmetric, $u_1, u_2,...,u_N$ form a basis, $(x_i - \bar{x})$, can be written as a linear combination of the eigenvectors:

$$x_i - \bar{x} = b_1 u_1 + b_2 u_2 + ... + b_N u_N = \sum_{i=1}^{N} l \tag{6}$$

where $b_1, b_2,..., b_N$ are scalars.

Table 1
Centrifugal pump specification.

Speed: 2880 rpm	Pump size: 50 × 50 mm
Current: 11.5 A	Discharge: 392 L per second
Head: 20 m	Power: 2 HP

Step 6 For dimensionality reduction, it keeps only the terms corresponding to the K largest eigenvalues:

$$x_i - \bar{x} = \sum_{i=1}^{N} b_i u_i \text{ where } K \ll N \tag{7}$$

The representation of $x \cdot x$ into the basis $u_1, u_2,..., u_K$ is thus

$$\begin{matrix} b_1 \\ b_2 \\ ... \\ b_K \end{matrix} \tag{8}$$

4.2. Nonlinear dimensionality reduction techniques

4.2.1. Diffusion maps

Diffusion maps are based on defining a Markov random walk on the graph of the data. By performing the random walk for a number of time steps, a measure for the proximity of the data points is obtained. Using this measure, the diffusion distance is defined. The basic working of diffusion map is presented below :

Step 1 A graph of the data is constructed first in the diffusion maps frame work. Using the Gaussian Kernel function, the weights of the edges in the graph are computed, leading to a matrix W with entries

$$W_{ij} = e^{-\frac{x_i - x_j^2}{2\sigma^2}} \tag{9}$$

where σ indicates the variance of the Gaussian.

Step 2 normalization of the matrix W is performed in such a way that its rows add up to 1. In this way, a matrix $P^{(1)}$ is formed with entries

$$p_{ij}^{(1)} = \frac{w_{ij}}{\sum_k w_{ik}} \tag{10}$$

Since diffusion maps originate from dynamical systems theory, the resulting matrix $P^{(1)}$ is considered a Markov matrix that defines

the forward transition probability matrix of a dynamical process. Hence, the matrix $P^{(1)}$ represents the probability of a transition from one data point to another data point in a single time step. The forward probability matrix for t time steps $P^{(t)}$ is thus given by $(P^{(1)})^t$.

Step 3 Diffusion distance is defined using random walk forward probabilities p_{ij}^t

$$D^{(t)}(x_i, x_j) = \sqrt{\sum_k \frac{\left(p_{ik}^{(t)} - p_{jk}^t\right)}{\Psi(x_k)^{(0)}}} \qquad (11)$$

In the above equation, $\Psi(x_k)^{(0)}$ is a term that attributes more weight to parts of the graph with high density. It is defined by $\Psi(x_k)^{(0)} = m_i / \sum_j m_j$, where m_i is the degree of node x_i defined by $m_i = \sum_j p_{ij}$

Step 4 In the low-dimensional representation of the data Y, diffusion maps attempt to retain the diffusion distances. Using spectral theory on the random walk, the low-dimensional representation Y that retains the diffusion distances $D^{(t)}(x_i, x_j)$ is formed by the d nontrivial principal eigenvectors of the eigenproblem

$$P^{(t)} v = \lambda v \qquad (12)$$

Because the graph is fully connected, the largest eigenvalue is trivial (viz. $\lambda_1 = 1$), and its eigenvector v_1 is thus discarded. The low-dimensional representation Y is given by the next d principal eigenvectors. In the low-dimensional representation, the eigenvectors are normalized by their corresponding eigenvalues. Hence, the low-dimensional data representation is given by

$$Y = \{\lambda_2 v_2, \lambda_3 v_3, \ldots, \lambda_{d+1} v_{d+1}\} \qquad (13)$$

4.2.2. Hessian LLE

Hessian LLE (HLLE) [7] is a variant of LLE that minimizes the 'curviness' of the high-dimensional manifold when embedding it into a low-dimensional space, under the constraint that the low-dimensional data representation is locally isometric. This is done by an eigenanalysis of a matrix H that describes the curviness of the manifold around the data points. The curviness of the manifold is measured by means of the local Hessian at every data point. The local Hessian is represented in the local tangent space at the data point, in order to obtain a representation of the local Hessian that is invariant to differences in the positions of the data points. It can be shown that the coordinates of the low-dimensional representation can be found by performing an eigenanalysis of an estimator H of the manifold Hessian.

Step 1 Identifying the k nearest neighbours for each data point x_i using Euclidean distance. Assume local linearity of the manifold in the neighbourhood.
Step 2 Applying PCA on its k nearest neighbours, the local tangent space at point x_i can be found.
Step 3 An estimator for the Hessian of the manifold at point x_i in local tangent space coordinates is computed. In order to do this, a matrix Z_i is formed that contains (in the columns) all cross products of M up to the d_{th} order.
Step 4 The matrix Z_i is orthonormalized by applying Gram-Schmidt ortho-normalization [1].

The estimation of the tangent Hessian H_i is now given by the transpose of the last $d(d + 1)/2$ columns of the matrix Z_i. Using the

Hessian estimators in local tangent coordinates, a matrix H is constructed

$$H_{lm} = \sum_i \sum_j \left((H_i)_{jl} X (H_i)_{jm}\right) \qquad (14)$$

The matrix H represents information on the curviness of the high-dimensional data manifold. An eigenanalysis of H is performed in order to find the low-dimensional data representation that minimizes the curviness of the manifold. The eigenvectors corresponding to the d smallest nonzero eigenvalues of H are selected and form the matrix Y, which contains the low-dimensional representation of the data.

4.2.3. Isomap

Isomap [28] is a technique that attempts to preserve pairwise geodesic (or curvilinear) distances between data points. Geodesic distance is the distance between two points measured over the manifold.

Step 1 In Isomap, the geodesic distances between the data points x_i ($i = 1, 2, \ldots, n$) are computed by constructing a neighbourhood graph G, in which every data point x_i is connected with its k nearest neighbours x_{ij} ($j = 1, 2, \ldots, k$) in the data set X.
Step 2 The shortest path between two points in the graph forms an estimate of the geodesic distance between these two points, and can easily be computed using Dijkstra's or Floyd's shortest-path algorithm [6,8].
Step 3 The geodesic distances between all data points in X are computed, thereby forming a pairwise geodesic distance matrix. The geodesic distances between all data points in X are computed, thereby forming a pairwise geodesic distance matrix.
Step 4 The low-dimensional representations y_i of the data points x_i in the low-dimensional space Y are computed by applying classical scaling.

4.2.4. Kernel PCA

Kernel PCA (KPCA) is the reformulation of traditional linear PCA in a high-dimensional space that is constructed using a kernel function [27]. Kernel PCA computes the principal eigenvectors of the kernel matrix, rather than those of the covariance matrix. The reformulation of PCA in kernel space is straightforward, since a kernel matrix is similar to the inner product of the data points in the high-dimensional space that is constructed using the kernel function. The application of PCA in the kernel space provides Kernel PCA the property of constructing nonlinear mappings.

4.2.5. Laplacian Eigenmaps

In Laplacian Eigenmaps, the local properties are based on the pairwise distances between near neighbours. Laplacian Eigenmaps compute a low-dimensional representation of the data in which the distances between a data point and its k-nearest neighbours are minimized. This is done in a weighted manner, i.e., the distance in the low-dimensional data representation between a data point and its first nearest neighbour contributes more to the cost function than the distance between the data point and its second nearest neighbour. Using spectral graph theory, the minimization of the cost function is defined as an eigenproblem.

4.2.6. Local linear embedding (LLE)

LLE [22] is a technique similar to Isomap (and MVU) in that it constructs a graph representation of the data points. In contrast to Isomap, it attempts to preserve solely local properties of the data. As a result, LLE is less sensitive to short-circuiting than Isomap,

because only a small number of local properties are affected if short-circuiting occurs. Furthermore, the preservation of local properties allows for successful embedding of non-convex manifolds. In LLE, the local properties of the data manifold are constructed by writing the high-dimensional data points as a linear combination of their nearest neighbours. In the low-dimensional representation of the data, LLE attempts to retain the reconstruction weights in the linear combinations as good as possible.

4.2.7. Local tangent space analysis (LTSA)

LTSA describes local properties of the high-dimensional data using the local tangent space of each data point [41]. It is based on the observation that, if local linearity of the manifold is assumed, there exists a linear mapping from a high-dimensional data point to its local tangent space, and that there exists a linear mapping from the corresponding low-dimensional data point to the same local tangent space. LTSA attempts to align these linear mappings in such a way, that they construct the local tangent space of the manifold from the low-dimensional representation.

4.2.8. Manifold charting

Manifold charting constructs a low-dimensional data representation by aligning a MoFA or a MoPPCA model [4]. Manifold charting minimizes a convex cost function that measures the amount of disagreement between the linear models on the global coordinates of the data points. The minimization of this cost function can be performed by solving a generalized eigenproblem.

4.2.9. Maximum variance unfolding (MVU)

MVU is a technique that attempts to resolve this problem by learning the kernel matrix. MVU learns the kernel matrix by defining a neighbourhood graph on the data (as in Isomap) and retaining pairwise distances in the resulting graph [32]. MVU is different from Isomap in that it explicitly attempts to 'unfold' the data manifold. It is done by maximizing the Euclidean distances between the data points, under the constraint that the distances in the neighbourhood graph are left unchanged (i.e., under the constraint that the local geometry of the data manifold is not distorted). The resulting optimization problem can be solved using semi definite programming.

5. Classifiers

5.1. Decision tree algorithm

A decision tree [19] represents the information in the signal, presented to it as features, in the form of a tree. Decision trees are built recursively, following a top-down approach. A tree induced using the C5.0 (or ID3 or C4.5) algorithm consists of a number of branches, one root, a number of nodes and a number of leaves. One branch is a chain of nodes from root to a leaf; and each node
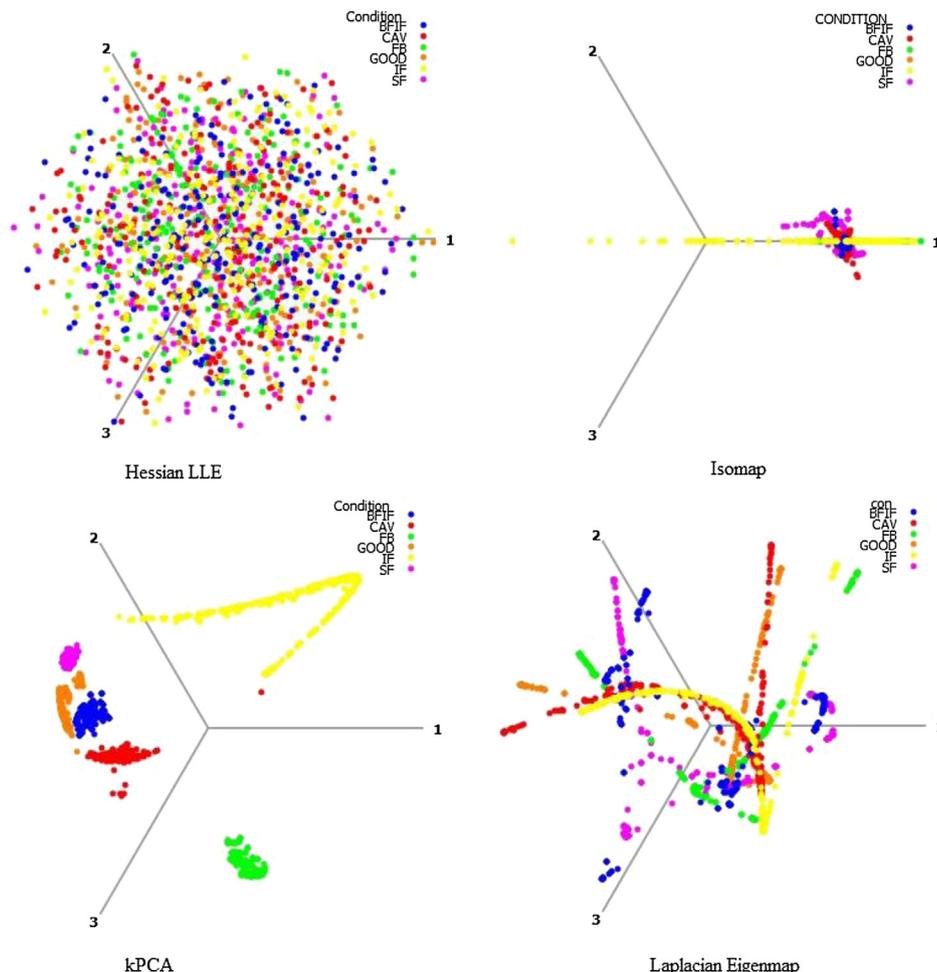

Fig. 2. Three dimensional representation of centrifugal pump fault data set for Hessian LLE, Isomap, kPCA and Laplacian Eigenmap based dimensionality reduction techniques.

involves one attribute. Classification is done through the decision tree with its leaves representing the different conditions of the mono block centrifugal pump.

5.1.1. The information gain and entropy reduction is calculated in the following way

Information gain measures how well a given attribute separates the training examples according to their target classification. The measure is used to select among the candidate features at each step while growing the tree. Information gain is the expected reduction in entropy caused by partitioning the samples according to this feature.

Information gain (S, A) of a feature A relative to a collection of examples S, is defined as

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{v \in \text{Value}(A)} \frac{|S_V|}{|S|} \text{Entropy}(S_V) \qquad (15)$$

where Values(A) is the set of all possible values for attribute A, and S_v is the subset of S for which feature A has value v (ie., $S_v = \{s \in S | A(s) = v\}$).

The first term in the equation for Gain is the entropy (measure of disorder in the data) of the original collection S and the second term is the expected value of the entropy after S is partitioned using feature A. The expected entropy described by the second term is simply the sum of the entropies of each subset S_v, weighed by the fraction of samples $|S_v|/|S|$ that belong to S_v. Gain(S, A) is therefore the expected reduction in entropy caused by knowing the value of feature A. Entropy is a measure of homogeneity of the set of examples and it is given by

$$\text{Entropy}(S) = \sum_{i=1}^{c} -P_i \log_2 P_i \qquad (16)$$

where c is the number of classes, p_i is the proportion of S belonging to class 'i'.

5.2. Naïve Bayes algorithm [29]

The Naïve Bayes classifier is a highly practical Bayesian learning method. The following description is based on the discussion in Ref. [15]. The Naïve Bayes classifier applies to learning tasks where each instance x is described by a conjunction of attribute values and the target function $f(x)$ can take on any value from some finite set C. A set of training examples of the target function is provided, and a new instance is presented, described by the tuple of attribute values $(a_1, a_2,..., a_n)$. The learner is asked to predict the target value, or classification, for this new instance.

The Bayesian approach to classifying the new instance is to assign the most probable target value, c_{MAP}, given the attribute values $(a_1, a_2,..., a_n)$ that describe the instance.
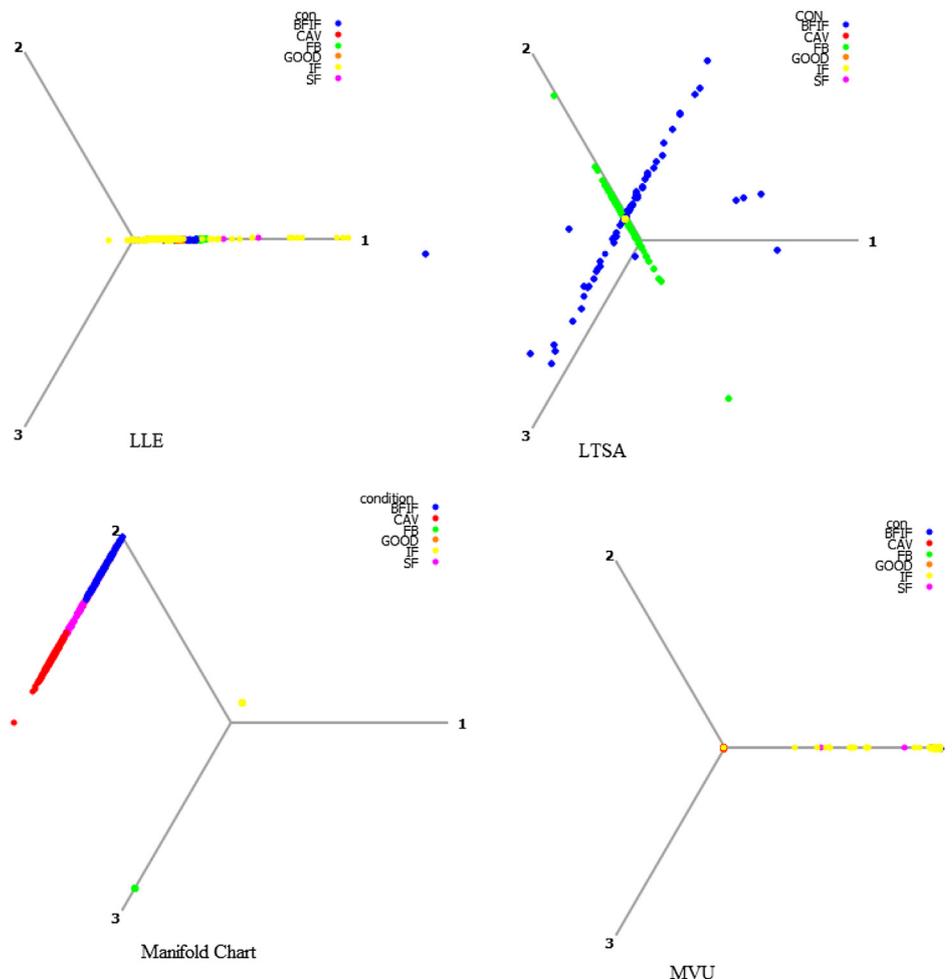

Fig. 3. Three dimensional representation of centrifugal pump fault data set for LLE, LTSA, manifold chart and MVU based dimensionality reduction techniques.

$$c_{\text{MAP}} = \underset{c_j \in C}{\text{argmax}} \left(P(c_j | a_1, a_2, ..., a_n) \right) \qquad (17)$$

Using Bayes theorem,

$$c_{\text{MAP}} = \underset{c_j \in C}{\text{argmax}} \left(\frac{P(a_1, a_2, ..., a_n | c_j) P(c_j)}{P(a_1, a_2, ..., a_n)} \right)$$

$$= \underset{c_j \in C}{\text{argmax}} \left(P(a_1, a_2, a_3, ..., a_n | c_j) P(c_j) \right) \qquad (18)$$

The Naïve Bayes classifier makes the further simplifying assumption that the attribute values are conditionally independent given the target value. Therefore,

$$c_{\text{NB}} = \underset{c_j \in C}{\text{argmax}} \left(P(c_j) \prod_i P(a_i | c_j) \right) \qquad (19)$$

where c_{NB} denotes the target value output by the Naïve Bayes classifier.

The conditional probabilities $P(a_i | c_j)$ need to be estimated from the training set. The prior probabilities $P(c_j)$ also need to be fixed in some fashion (typically by simply counting the frequencies from the training set). The probabilities for differing hypotheses (classes) can also be computed by normalizing the values received for each hypothesis (class).

5.3. Bayes Net algorithm

Bayesian network [9] consists of a set of variables, $V = \{A_1, A_2, ..., A_N\}$ and a set of directed edge, E, between variables, which form a directed acyclic graph (DAG) $G = (V, E)$ where a joint distribution of variables is represented by the product of conditional distributions of each variable given its parents [3]. Each node, $A_i \in V$ represents a random variable and a directed edge from A_i to A_j, $(A_i, A_j) \in E$, represents the conditional dependency between A_i and A_j. In a Bayesian networks, each variable is independent on its non-descendants, given a value of its parents in G. This independence encoded in G reduces the number of parameters which is required to characterize a joint distribution, so that posterior distribution can be efficiently inferred.

In a Bayesian network over $V = \{A_1, A_2, ..., A_n\}$, the joint distribution $P(V)$ is the product of all conditional distributions specified in the Bayesian network such as

$$P(A_1, A_2, ..., A_N) = \prod_{i=1}^{N} P(A_i / Pa_i) \qquad (20)$$

where, $P(A_i / Pa_i)$ is the conditional distribution of A_i, given Pa_i which denotes the parent set of A_i. A conditional distribution for each variable has a parametric form that can be learnt by the maximum likelihood estimation.

6. Results and discussion

The study of fault classification performance of mono block centrifugal pump using traditional dimensionality reduction technique and nonlinear dimensionality reduction techniques is discussed in the following phases:

  (i) Visual analysis of the reduced feature set.
 (ii) Comparison of dimensionality reduction techniques with decision tree, Bayes Net Naïve Bayes and kNN classifiers.
(iii) Comparison of the above results with a stand-alone decision tree classifier.

As can be seen from Figs. 2–4, different dimensionality reduction techniques offer differing degrees of separation of different fault categories. From observation of the above Figs. 2–4 it is clear that kPCA and PCA provide the maximum separation between different fault categories hence combining these dimensionality reduction techniques with a classifier is very likely to yield a high degree of classification accuracy. On the other hand, Hessian LLE based dimensionality reduction offers the least visual separation between fault classes and hence is likely to result in very poor classification accuracy when combined with a classifier. These assertions on the likely impact of different dimensionality reduction techniques on the classification accuracy are verified with the help of decision tree, naïve Bayes, Bayes Net and kNN classifiers in the subsequent sections. Tables 1–4 show the classification accuracy obtained using decision tree, Bayes Net, Naïve Bayes and kNN classifiers respectively, with various nonlinear dimensionality reduction techniques and PCA. It must be noted that normalisation is performed on all the features before the process of dimensionality reduction is carried out. From Tables 1–4 among the nonlinear dimensionality reduction techniques and PCA, the PCA outperforms when using decision tree, Bayes Net, Naïve Bayes and kNN classifiers. In this work, the lowest accuracy is obtained by the Hessian LLE transformed features with all the four classifiers. PCA-decision tree classifier outperforms the other three classifiers, namely, PCA-naïve Bayes, PCA-Bayes Net and PCA-kNN. Fig. 5 shows the plot of various dimensionality reduction techniques and their percentage classification accuracy with decision tree, naïve Bayes, Bayes Net and kNN classifiers. The overall classification accuracy was found to be 99.45% for the PCA-decision tree classifier, which is slightly higher than the overall classification accuracy of PCA-kNN (99.43%), PCA-Naïve Bayes (99.3%) and PCA-Bayes Net (99.18%) (Table 5).

The above results are also compared with the results obtained during the same experimental setup using a stand-alone C4.5 decision tree algorithm, as reported by the same author, [25]. Table 6 shows the results of % classification accuracy for the same data and for the same statistical features that were extracted. It was observed that reducing the dimensionality of the feature set to 3 was able to generate best classification accuracy for all the dimensionality reduction techniques.

In PCA decision tree classifier, the application of PCA transforms the original features of the data set into a reduced number of uncorrelated features which are the principle components. This results in a small loss of information contained
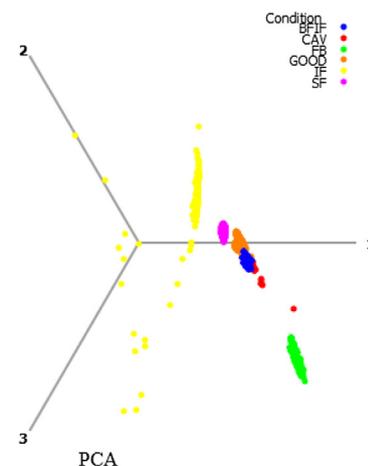

Fig. 4. Three dimensional representation of centrifugal pump fault data set for PCA based dimensionality reduction technique.

Table 2
Tabulation of classification accuracy of mono block centrifugal pump for decision tree classifier with different nonlinear dimensionality reduction techniques and PCA.
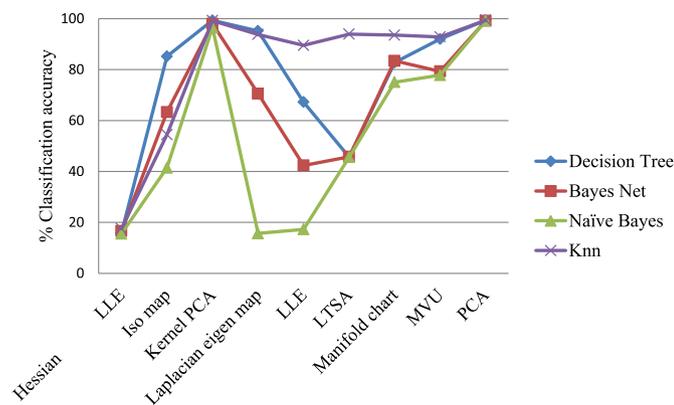
Dimensionality reduction technique	Hessian LLE	Isomap	Kernel PCA	Laplacian Eigenmap	LLE	LTSA	Manifold chart	MVU	PCA
Classifier: decision tree									
GOOD	70	66.2	98.8	94.3	32.6	100	72.4	99.3	97.6
CAV	0	79.01	99.6	92.8	41.17	0	98	83.6	99.6
FB	14.4	92.7	100	97.2	92.4	100	100	99.5	99.6
BFIF	3.2	85.7	99.2	97.2	96.55	73.6	49.2	91.4	100
IF	10.4	93.8	98.4	94	55.36	0	100	100	100
SF	8	93.9	100	95.43	85.21	0	77.2	78.7	100
% Classification accuracy	15.9	85.21	99.33	95.18	67.22	45.6	82.8	92	99.45

Table 3
Tabulation of classification accuracy of Mono block centrifugal pump for Bayes Net classifier with different nonlinear dimensionality reduction techniques and PCA.

Dimensionality reduction technique	Hessian LLE	Isomap	Kernel PCA	Laplacian Eigenmap	LLE	LTSA	Manifold chart	MVU	PCA
Classifier: Bayes Net									
GOOD	100	24.8	95.2	72.9	21.7	23.5	74	100	97.6
CAV	0	75	99.6	78	32.4	36.4	99	90.3	99.2
FB	0	55.6	100	72.8	86.5	95.6	98	100	99.5
BFIF	0	71.4	98	75.6	29.3	13.5	52	38.8	99.2
IF	0	92	100	58.8	61.8	17.8	99.7	100	99.6
SF	0	61.2	95.2	64.8	22.48	88	78	46.7	100
% Classification accuracy	16.6	63.3333	98	70.55	42.36	45.8	83.45	79.3	99.18

Table 4
Tabulation of classification efficiency of mono block centrifugal pump for naïve Bayes classifier with different nonlinear dimensionality reduction techniques and PCA.

Dimensionality reduction technique	Hessian LLE	Isomap	Kernel PCA	Laplacian Eigenmap	LLE	LTSA	Manifold chart	MVU	PCA
Classifier: Naïve Bayes									
GOOD	30.4	0.6	91.6	6.6	5.4	100	43.2	100	97.6
CAV	9.2	99.55	98.8	31.2	56.3	0	98	91.6	99.6
FB	16.8	37.9	100	8	1.6	100	100	100	100
BFIF	10.4	0	98	15.6	23.27	73.6	48	23.3	98.8
IF	10.8	82.7	100	27.6	14.6	0	100	100	100
SF	15.6	27.8	90.8	5	2.4	0	61.2	52.07	100
% Classification accuracy	15.53	41.425	96.53	15.67	17.26	45.6	75.06	77.82	99.3

Fig. 5. Percentage classification accuracy for decision tree, naïve Bayes, Bayes Net and kNN classifiers with different nonlinear dimensionality reduction techniques and PCA.

in the original data set. This is the reason attributed to the slightly lower accuracy (while significantly reducing the computational expense due to feature reduction) obtained in the present study.

7. Conclusions and future directions

Condition monitoring of a mono block centrifugal is carried out using vibration signals and the statistical features are extracted. These features are transformed using dimensionality reduction techniques and then, they are classified using decision tree, Bayes Net and naïve Bayes classifiers. On observation of the results, it may be concluded that among the various dimensionality reduction techniques, the traditional dimensionality technique PCA gives higher classification efficiency, for all the classifiers considered. Among the classifiers, PCA-decision tree combination is found to be more effective than all other dimensionality reduction technique–

Table 5
Tabulation of classification efficiency of mono block centrifugal pump for kNN (1 NN) classifier with different nonlinear dimensionality reduction techniques and PCA.

Dimensionality reduction technique	Hessian LLE	Isomap	Kernel PCA	Laplacian Eigenmap	LLE	LTSA	Manifold chart	MVU	PCA
Classifier: kNN									
GOOD	19.6	40.4	100	82.8	79.6	88.4	70.4	83.6	99.3
CAV	12.8	76.8	99.6	96.8	86.8	98.4	97.2	90.8	98.2
FB	20	48.8	100	100	90.4	100	100	92.4	100
BFIF	15.2	17.6	100	99.2	98.4	92.4	98	99.2	99.1
IF	16.4	86.8	98	97.6	84.8	99.2	100	88.8	100
SF	21.2	56.4	98	86.4	72.4	85.2	95.6	76.4	100
% Classification accuracy	17.53	54.47	99.23	93.8	89.5	93.93	93.53	92.86	99.43

Table 6
Comparison of efficiency for various classifiers [25].

Classifier	% Classification accuracy
Decision tree	100
PCA-kNN	99.43
PCA-decision tree	99.45
PCA-Naïve Bayes	99.30
PCA-Bayes Net	99.18

classifier combinations. From the results obtained, we may conclude that nonlinear techniques for dimensionality reduction are not capable of outperforming traditional linear techniques such as PCA. Evaluation of recently developed dimensionality reduction techniques such as sparse distance preserving embedding, sparse proximity preserving embedding [38–40], Sparse local discriminant projections [43], dynamic transition embedding (DTE) [44] and sparsity preserving projections [18] etc. for fault detection of centrifugal pumps could also be carried out. A possible application of the system proposed in the present study is a completely automated on-line pump condition monitoring system which can automatically inform the operator of any faults, point out the faulty part and hence, the maintenance may be carried out only when there is a malfunction. This will help in significantly reducing the maintenance overheads.

References

[1] G. Afken, Gram-Schmidt Orthogonalization, Academic Press, Orlando, FL, USA, 1985.

[2] M.S. Baghshah, S.B. Shouraki, Semi-supervised metric learning using pairwise constraints, in: Proceedings of the International Joint Conferences on Artificial Intelligence, 2009, pp. 1217—1222.

[3] Bo-Chiuan Chen, Yuh-Yih Wu, Feng-Chi Hsieh, Go-Long Tsai, Crank Angle Estimation with Kalman Filter and Stroke Identification for Electronic Fuel Injection Control of a Scooter Engine, SAE Document Number: 2005-01-0076, 2005.

[4] M. Brand, Charting a manifold, in: Advances in Neural Information Processing Systems, vol. 15, The MIT Press, Cambridge, MA, USA, 2002, pp. 985—992.

[5] M. Demetgul, I.N. Tansel, S. Taskin, Fault diagnosis of pneumatic systems with artificial neural network algorithms, Expert Syst. Appl. 36 (2009) 10512—10519.

[6] E.W. Dijkstra, A note on two problems in connexion with graphs, Numer. Math. 1 (1959) 269—271.

[7] D.L. Donoho, C. Grimes, Hessian eigenmaps: new locally linear embedding techniques for high-dimensional data, Proc. Natl. Acad. Sci. 102 (21) (2005) 7426—7743.

[8] R.W. Floyd, Algorithm 97: shortest path, Commun. ACM 5 (6) (1962) 345.

[9] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifier, Mach. Learn. 29 (1997) 131—163.

[10] D.H. Hellmann, Early fault detection, World Pumps (2002) 54—57.

[11] I. Hamernick, Ludeca, Vibration analysis for condition monitoring, Pumps Syst. (2006) 62—63.

[12] K. Hartigan, Inside monitoring to extend pump life, World Pumps 502 (2008) 28—30.

[13] J. Jensen, K. Dayton, Detecting cavitation in centrifugal pumps, Orbit (2000) 26—30.

[14] Laurens van der Maaten, Eric Postma, Jaap van den Herik, Dimensionality Reduction: a Comparative Review, 2009.

[15] T. Mitchell, Machine Learning, first ed., McGraw-Hill Science/Engineering/Math, New York, 1997.

[16] Necla Togun, Sedat Baysec, Genetic programming approach to predict torque and brake specific fuel consumption of a gasoline engine, Appl. Energy 87 (2010) 3401—3408.

[17] F.P. Nie, S.M. Xiang, Y.Q. Jia, C.S. Zhang, Semi-supervised orthogonal discriminant analysis via label propagation, Pattern Recognit. 42 (11) (2009) 2615—2627.

[18] L.S. Qiao, S.C. Chen, X.Y. Tan, Sparsity preserving projections with applications to face recognition, Pattern Recognit. 43 (1) (2010) 331—341.

[19] J.R. Quinlan, Improved use of continuous attributes in C4, J. Artif. Intell. Res. (1996) 77—90.

[20] J. Rafiee, F. Arvani, A. Harifi, M.H. Sadeghi, Intelligent condition monitoring of a gearbox using artificial neural network, Mech. Syst. Signal Process. 21 (2007) 1746—1754.

[21] S. Rajakarunakaran, P. Venkumar, D. Devaraj, K. Surya Prakasa Rao, Artificial neural network approach for fault detection in rotary system, Appl. Soft Comput. 8 (2008) 740—778.

[22] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323—2326.

[23] N.R. Sakthivel, V. Sugumaran, B.B. Nair, Application of support vector machine (SVM) and proximal support vector machine (PSVM) for fault classification of mono block centrifugal pump, Int. J. Data Anal. Tech. Strat. 2 (2010a) 38—61.

[24] N.R. Sakthivel, V. Sugumaran, B.B. Nair, Comparison of decision tree-fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump, Mech. Syst. Signal Process. 24 (2010b) 1887—1906.

[25] N.R. Sakthivel, V. Sugumaran, S. Babudevasenapati, Vibration based fault diagnosis of mono block centrifugal pump using decision tree, Expert Syst. Appl. 37 (2010c) 4040—4049.

[26] N.R. Sakthivel, V. Sugumaran, B.B. Nair, R.S. Rai, Decision support system using artificial immune recognition system for fault classification of centrifugal pump, Int. J. Data Anal. Tech. Strat. 3 (1) (2011) 66—84.

[27] B. Schölkopf, A.J. Smola, K.R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (5) (1998) 1299—1319.

[28] J.B. Tenenbaum, V. De Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319—2323.

[29] J. Vaidya, M. Kantarcıoglu, C. Clifton, Privacy-preserving Naïve Bayes classification, VLDB J. 17 (2008) 879—898.

[30] Van Tung Tran, Bo-Suk Yang, Myung-Suck Oh, Andy Chit Chiow Tan, Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference, Expert Syst. Appl. 36 (2009) 1840—1849.

[31] H.Q. Wang, P. Chen,) Fault diagnosis of centrifugal pump using symptom parameters in frequency domain, CGIR E-J. 9 (2007) 1—14.

[32] K.Q. Weinberger, F. Sha, L.K. Saul, Learning a kernel matrix for nonlinear dimensionality reduction, in: Proceedings of the 21st International Conference on Machine Learning, 2004.

[33] Yaguo Lei, Zhengjia He, Yanyang Zi, Qiao Hu, Fault diagnosis of rotating machinery based on multiple ANFIS combination with gas, Mech. Syst. Signal Process. 21 (2007) 2280—2294.

[34] Yaguo Lei, Zhengjia He, Yanyang Zi, A new approach to intelligent fault diagnosis of rotating machinery, Expert Syst. Appl. 35 (2008) 1593—1600.

[35] M. Yates, Pump performance monitoring complements condition monitoring, World Pumps 262 (2002) 36—38.

[36] D.Q. Zhang, Z.H. Zhou, S.C. Chen, Semi-supervised dimensionality reduction, in: Proceedings of the 7th SIAM International Conference on Data Mining (SDM), Minneapolis, MN, 2007, pp. 629—634.

[37] Y. Zhang, J. Zhuang, S. Wang, Fusion of manifold learning and spectral clustering algorithm with applications to fault diagnosis, in: Second International Conference on Machine Learning and Computing, ICMLC, 2010, pp. 155—160.

[38] Z. Zhang, M.B. Zhao, T.W.S. Chow, Binary- and multi-class group sparse canonical correlation analysis for feature extraction and classification, IEEE Trans. Knowl. Data Eng. 25 (10) (2013a) 2192—2205.

[39] Z. Zhang, T.W.S. Chow, M.B. Zhao, M-Isomap: orthogonal constrained marginal isomap for nonlinear dimensionality reduction, IEEE Trans. Syst. Man. Cybern. Part B Cybern. 43 (1) (2013b) 180—192.

[40] Z. Zhang, S. Yan, M. Zhao, Pairwise sparsity preserving embedding for unsupervised subspace learning and classification, IEEE Trans. Image Process. 22 (12) (2013c) 4640—4651.

[41] Z. Zhang, H. Zha, Principal manifolds and nonlinear dimensionality reduction via local tangent space alignment, SIAM J. Sci. Comput. 26 (1) (2004) 313—338.

[42] M. Zhao, X. Jin, Z. Zhang, B. Li, Fault diagnosis of rolling element bearings via discriminative subspace learning: visualization and classification, Expert Syst. Appl. 41 (2014) 3391—3401.

[43] Z. Lai, Z. Jin, J. Yang, W.K. Wong, Sparse local discriminant projections for face feature extraction, in: Proceedings of International Conference on Pattern Recognition, Istanbul, Turkey, 2010.

[44] Z. Lai, Z. Jin, M. Sun, J. Yang, Dynamic transition embedding for image feature extraction and recognition, Neural Comput. Appl. 21 (2011) 1905—1915.