International Journal of Computational Intelligence and Applications
Vol. 15, No. 4 (2016) 1650019 (26 pages)

DOI: 10.1142/S146902681650019X

Hybrid Hierarchical Clustering — Piecewise Aggregate Approximation, with Applications

Yu Zhang*, Michael Gallimore†, Chris Bingham‡ and Jun Chen§

School of Engineering, University of Lincoln
Lincoln LN6 7TS, UK
**yzhang@lincoln.ac.uk*
†mgallimore@lincoln.ac.uk
‡cbingham@lincoln.ac.uk
§juchen@lincoln.ac.uk

YONG XU

Institute of Acoustics, Chinese Academy of Sciences
Beijing 100190, P. R. China
xuyong@mail.ioa.ac.cn

Received 11 August 2015
Revised 16 June 2016
Accepted 2 November 2016
Published

Piecewise Aggregate Approximation (PAA) provides a powerful yet computationally efficient tool for dimensionality reduction and Feature Extraction (FE) on large datasets compared to previously reported and well-used FE techniques, such as Principal Component Analysis (PCA). Nevertheless, performance can degrade as a result of either regional information insufficiency or over-segmentation, and because of this, additional relatively complex modifications have subsequently been reported, for instance, Adaptive Piecewise Constant Approximation (APCA). To recover some of the simplicity of the original PAA, whilst addressing the known problems, a distance-based Hierarchical Clustering (HC) technique is now proposed to adjust PAA segment frame sizes to focus segment density on information rich data regions. The efficacy of the resulting hybrid HC-PAA methodology is demonstrated using two application case studies on non-time-series data viz. fault detection on industrial gas turbines and ultrasonic biometric face identification. Pattern recognition results show that the extracted features from the hybrid HC-PAA provide additional benefits with regard to both cluster separation and classification performance, compared to traditional PAA and the APCA alternative. The method is therefore demonstrated to provide a robust readily implemented algorithm for rapid FE and identification for datasets.

Keywords: Piecewise aggregate approximation; hierarchical clustering; rundown vibration signature; high resolution range profile.

1. Introduction

Among well-known signal processing techniques dimensionality reduction is recognized as an important and is often used as a pre-processing step for more advanced analytical and numerical processing.[1] Such techniques are often subdivided into two main classes, viz. Feature Extraction (FE) and Feature Selection (FS).[2] FS is a process by which important feature subsets are separated from vast amounts of data, through wrappers, filters or embedded methods based on some correlation or mutual information criteria.[3] Whilst FS selects important features or filter out redundant features from the original data sets, FE is more transformative, identifying a subset of new features by keeping as much important information as possible, normally by distance measures and similarity searches within the original data-series.[4] Because of their selective nature, such techniques are regularly considered as underpinning methods in wider application fields of fault/anomaly detection, pattern recognition and classification systems.[5,6]

Many linear FE techniques have been reported and successfully applied. The most established being Principal Component Analysis (PCA),[7] which accomplishes FE by searching for a subset of orthogonal linear combinations of the original data with the greatest variances, i.e., the principal components.[8] PCA is considered a second-order method based on minimizing mean-square error and is useful for identifying and keeping dominant features contained in the original data, at the expense of often not providing a meaningful physical interpretation.[9]

Independent Component Analysis (ICA) has been developed recently for FE, and has been successfully applied for blind source separation.[10] It is a higher-order method that searches linear projections to maximize particular independence criteria that are not necessarily orthogonal but as statistically independent as possible.[11] It is claimed that ICA is able to extract more meaningful features than PCA from non-Gaussian data.[9]

Projection Pursuit (PP) techniques are an alternative set of linear FE methods that incorporate higher-order information.[12] PP seeks to identify projections that optimize a defined projection index that represents, in an explicit or implicit form, useful information contained within data series.[13] It is useful particularly for data sets that are non-Gaussian, but is much more computationally intensive than PCA.

More complex nonlinear FE methods are often based on extensions of the existing linear FE techniques, and include nonlinear PCA and nonlinear ICA, etc.,[14] each of which has been demonstrated to provide useful properties for a number of application sectors.[15]

These techniques generally require much greater computation effort than linear algorithms, and this is often considered as the limiting factor for use with large datasets.[9] Indeed, it is the computational load and implementation complexity that often precludes the use of even linear PCA, ICA and PP techniques in many application fields, and has resulted in the use of more fundamental methods for FE.

In terms of simplicity, low computational cost and ease of implementation, an alternative therefore, termed Piecewise Aggregate Approximation (PAA)[16,17] is a dimensionality reduction technique which was originally designed for large time-series datasets, and has been widely adopted for use in medical, financial, engineering, and speech/image processing systems due to its low computation overhead.[18–20] In Ref. 21, the authors compare PAA to other dimensionality reduction techniques, including Singular Value Decomposition (SVD), Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT), and have demonstrated its superiority both theoretically and empirically, in terms of providing much faster computational times and being suitable for arbitrary-length queries. For instance, the computational time overhead for PCA/SVD is $O(NM^2)$, where N is the number of samples, and M is the dimension, which shows that if dimensions are sufficiently high, then the computational time can be very costly. Alternatively, for PAA the computational overhead is related to $O(nM)$, where n is the number of the equal-sized frames, and therefore provides substantial computational benefit compared to alternative techniques.

Nevertheless, whilst the traditional practice of using equally distributed segments (for PAA) facilitates rapid implementation, it can lead to insufficient fidelity in some regions of interest, whilst providing over-segmentation in regions considered less information rich, thereby often reducing relative performance compared to the alternatives. Several methods have therefore been proposed to modify the segment frame sizes to enhance the quality performance, such as Adaptive Piecewise Constant Approximation (APCA) based on Haar DWTs[22] and other more generic optimization methods,[23] albeit at the expense of degrading the classical benefits of PAA due to the required additional computational overhead.

Here then, the hybrid use of PAA and Hierarchical Clustering (HC)[24] is proposed. HC has been extensively used in data analysis and signal processing due to its simplicity and visual interpretation of the hierarchy structure,[25–27] and is considered here as a means of optimizing PAA segment frame sizes according to sequence-sample similarity. HC is used to define optimal PAA frame size according to hierarchical distance measures, but at the same time not to compromise the original PAA's simplicity for further implementations. The main advantage of hybrid HC-PAA is the simplicity of both algorithms and therefore the ease of implementations.

To provide a seed for further discussion, the proposed methodology is depicted pictorially in Fig. 1. Specifically, HC is used to cluster the data series according to similarity, and PAA is applied to the clustered data series for data dimensionality reduction and FE. To validate the performance of the extracted features, both

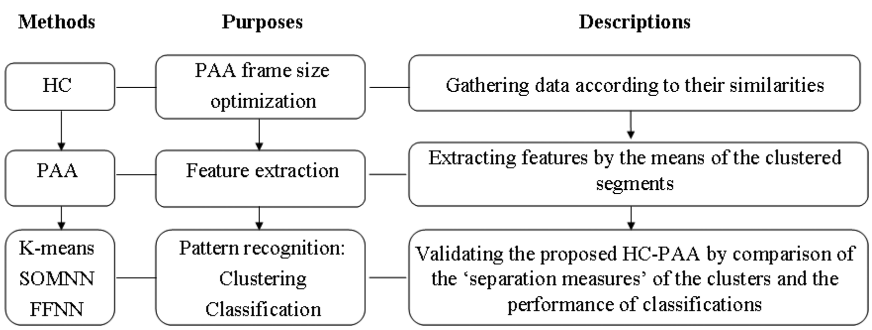


Fig. 1. Outline concept.

k-means and a Self-Organizing Map Neural Network (SOMNN) are also used as alternatives for clustering case, and a Feed-Forward Neural Network (FFNN) is applied as an alternative for the classification problem.

In summary, clustering and classification techniques are therefore applied to the features extracted by PAA, APCA and hybrid HC-PAA in order to show the benefits of the proposed hybrid HC-PAA solution compared to APCA and traditional PAA. From the results of the experimental trials it is shown that hybrid HC-PAA provides better FE performance than traditional PAA and the more recent APCA.

2. Methodology

2.1. *Traditional PAA*

PAA (alternatively termed segmented means[17]) subdivides a sequence $\mathbf{x}$ (a $1 \times N$ vector) into n equal sized segments, $\mathbf{g}_{\rm i}$ ($i = 1{:}n$), and uses the mean of each segment as an extracted feature to provide the resultant sequence $\mathbf{y}$ (a $1 \times n$ vector):

$$\mathbf{y} = [\mathrm{mean}(\mathbf{g}_1), \ldots, \mathrm{mean}(\mathbf{g}_n)]. \tag{1}$$

Each segment is therefore comprised of (N/n) data points of $\mathbf{x}$.[21] An example signal possessing a half bell shape, shown in Fig. 2, outlines the process, and highlights the underlying issues with the traditional method — it has 1000 samples that are

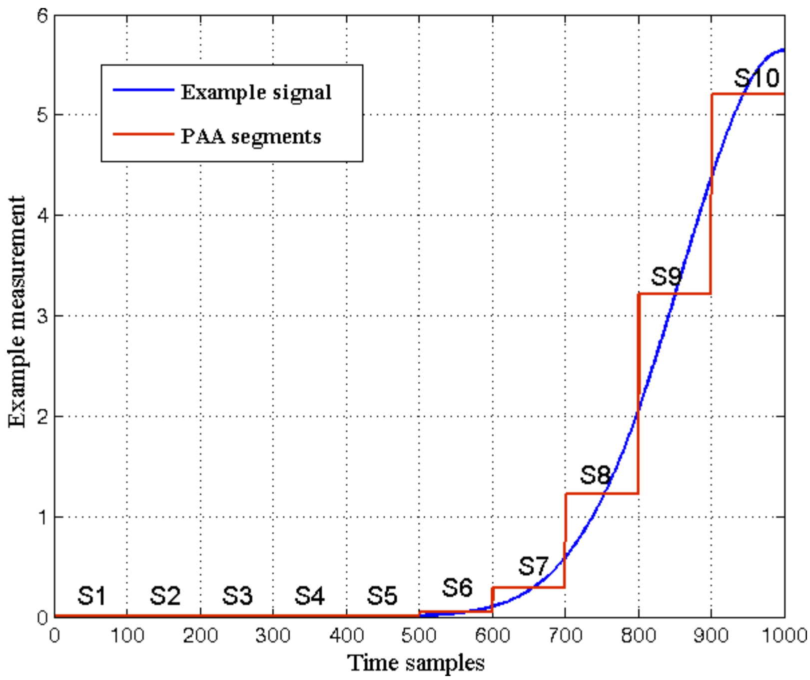


Fig. 2. Example signal and its PAA segmentation and representation (S = Segment).

separated into 10 equally spaced segments and which are represented by the mean of the data within each segment (traditional PAA).

It can be seen that, from 700 to 1000 samples, which is considered an information rich region, PAA segments are too coarse to capture the important features, whilst from 1 to 700 (a region less information rich), adjacent PAA segments provide relatively little added detail, and could be reasonably combined to provide further dimensionality reduction. It is a computationally efficient method of addressing this issue that is considered here, through use of HC.

2.2. *Hierarchical clustering*

HC provides a convenient visual hierarchy/clustering of datasets according to their similarity. The underlying concept of agglomerative HC is to assemble a set of objects into a hierarchical tree, where similar objects join in lower branches, which are further joined based on object "similarity". Objects with the smallest "distance" are joined by a branch of the tree (i.e., a cluster). Further clusters are then formed from merged subclusters, and the hierarchical process iterates until only one cluster remains.[28] The resulting cluster tree is classically depicted as a dendrogram. The resulting hierarchical tree can then be dissected according to either the linkage-distance or cluster number, and in so doing provide cluster classification or novelty detection.[29]

Here, to keep computational overhead low, the Euclidean distance is used as a measure of similarity:

$$d(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2}, \tag{2}$$

where $\mathbf{x}$ and $\mathbf{y}$ are two $1 \times N$ vectors, i.e., the signals $(x_1, x_2, \ldots, x_N)$ and $(y_1, y_2, \ldots, y_N)$. A cluster is formed when the data from two measurements has the minimum Euclidean distance. The first iteration provides the lowest ranking cluster. The procedure is subsequently iterated, including previously constructed clusters, to link higher ranking clusters. Again to limit computational overhead, an average linkage measure is used to calculate the mean distance between all pairs of objects in clusters m and n:

$$D(m,n) = \frac{1}{N_m N_n}\sum_{j=1}^{N_m}\sum_{k=1}^{N_n} d(\mathbf{x}_{mj}, \mathbf{y}_{nk}). \tag{3}$$

where $j = 1, 2, \ldots, N_m$ and $k = 1, 2, \ldots, N_n$. $d(\mathrm{x}_{mj}, \mathbf{y}_{nk})$ is the distance between two objects in the two clusters. N_m is the number of objects in cluster m, and N_n is the number of objects in cluster n.

For the example shown in Fig. 2, HC is applied to the 1000 time samples, and the samples are clustered according to their similarities. The resulting dendrogram is

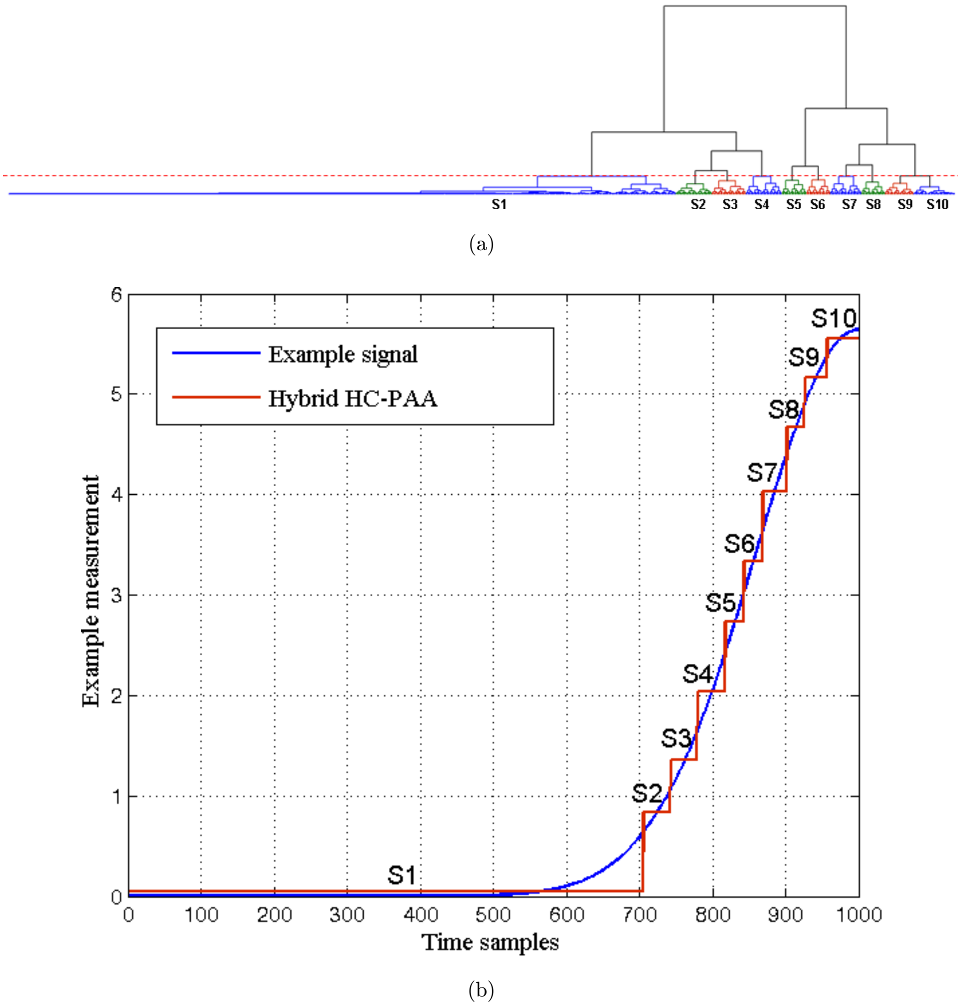


Fig. 3. (a) HC tree and 10 subclusters; (b) hybrid HC-PAA segments shown with the original example signal (S = Segment).

shown in Fig. 3(a). The linkage distance threshold (shown in red) enters from above to capture the highest 10 clusters in the dendrogram. Now, PAA is applied to the resulting 10 unequal segments, with the frame sizes being dictated by the size of the respective dendrogram branches. Again, the mean of each segment is used to

Table 1. Samples included in the original PAA and the hybrid HC-PAA segments (S = Segment).

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Original PAA	1–100	101–200	201–300	301–400	401–500	501–600	601–700	701–800	801–900	901–1000
HC-PAA	1–705	706–743	744–779	780–817	818–843	844–868	869–901	902–926	927–956	957–1000

represent the underlying "feature" of each segment according to (1). The resulting hybrid HC-PAA output, along with the original, is shown in Fig. 3(b). It is now evident that through application of the hybrid approach the regions that are information rich have a higher density of segments. For completeness, the segment regions of the original PAA representation and that of the proposed hybrid approach are given in Table 1, where the nonlinear mapping of segment length is clearly evident.

2.3. *Clustering and classification*

Once the segments have been determined, the underlying features can be clustered in order to provide identification or detect "novelty" (and hence the emergence of faults, for instance). In this case, for simplicity, k-means clustering[30] is used, using Euclidean distance to determine centroids. Since the k-means is known to be sensitive to the initial conditions, 20 executions are initiated and the optimized solution is used to reduce the impact of any anomalous results. For comparison purposes the "separation measure" is taken as the distance between cluster centers — a higher separation index therefore indicates improved cluster performance (with increased confidence that misclassification has not occurred).

To provide a more generic performance comparison for the proposed HC-PAA, Artificial Neural Networks (ANNs) are also considered in the example trials that follow. Specifically, a SOMNN is applied for clustering[31] with a "measure of separation" being used as a metric to compare relative performance; and a two-layer FFNN is used for classification,[32] where, with target classes, the Mean Squared Errors (MSEs) are calculated as a measure of performance. In this case then, higher cluster separation measures indicate improved cluster performance, and lower MSE values indicate improved classification performance.

3. Methodology

Performance of the proposed hybrid HC-PAA technique is demonstrated through application in experimental trials, firstly, as a means of detecting emerging faults on a sub-15MW industrial gas turbine based on rundown vibration sig-natures, and also as a biometric identification system based on face recognition using ultrasonic echo signals. Both case studies are pattern recognition problems, while the former one is for fault detection through clustering methods, and the latter one is a classification problem for feature/face recognition.

3.1. *Fault detection on industrial gas turbines*

Vibration signatures taken during the rundown periods of industrial gas turbines are considered as information-rich for determining the health of the underlying units. During a typical rundown, the unit will normally pass through at least one

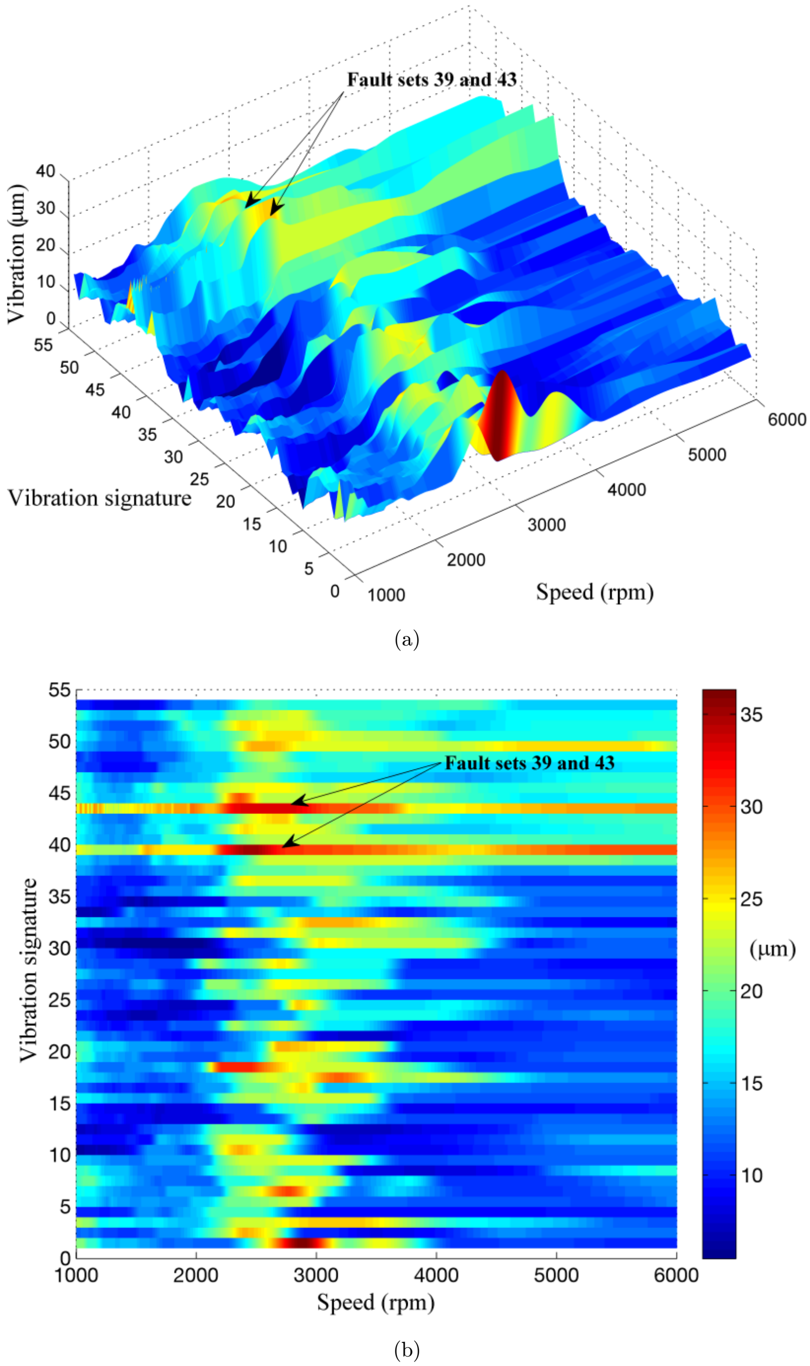


Fig. 4. (a) 3D plot and (b) 2D contour of the vibration signatures.

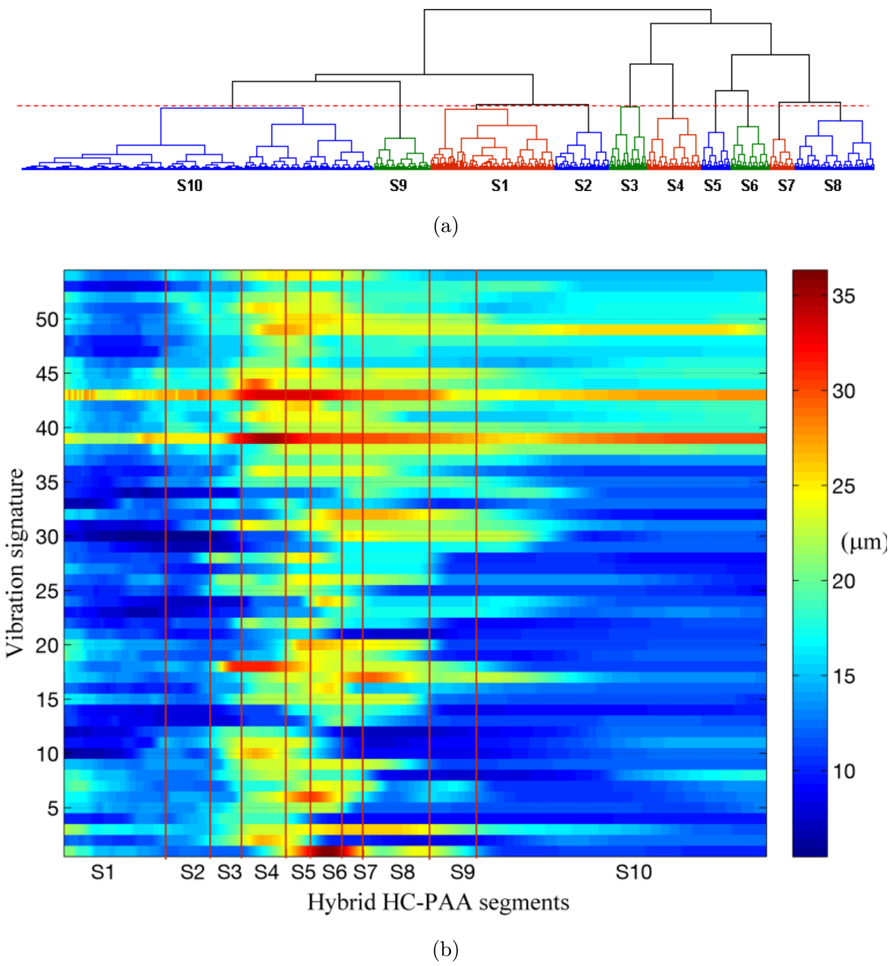


Fig. 5. (a) HC tree and 10 subclusters; (b) hybrid HC-PAA segments applied to the contour map of the vibration signatures (S = Segment).

rotor critical frequency. The objective is to group rundown signatures in order to identify those that show "novel" characteristics, and thereby act as an early warning of emerging fault conditions. Given the volume of data and the need to perform real-time similarity searches, the proposed hybrid HC-PAA approach is

Table 2. Segment regions from traditional PAA, APCA and the proposed hybrid HC-PAA approach (S = Segments).

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Original PAA	1001–1500	1501–2000	2001–2500	2501–3000	3001–3500	3501–4000	4001–4500	4501–5000	5001–5500	5501–6000
APCA	1001–1318	1319–1815	1816–2118	2119–2386	2387–2678	2679–3035	3036–3407	3408–3922	3923–4693	4694–6000
HC-PAA	1001–1725	1726–2045	2046–2265	2266–2585	2586–2757	2758–2985	2986–3133	3134–3602	3603–3936	3937–6000

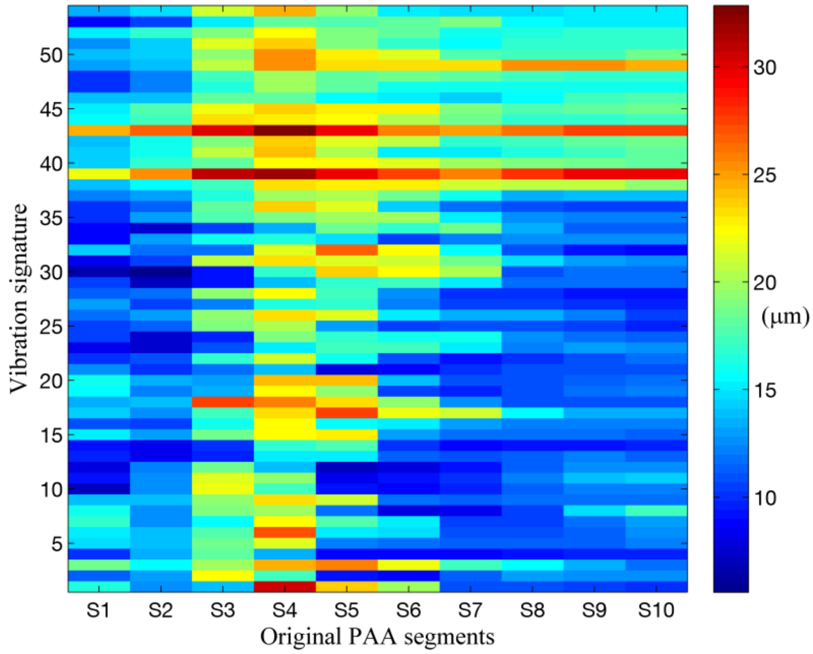
used for the detection of emerging faults. As an exemplar, a series of 54 rundown characteristics from a sub-15MW gas turbine are considered that include 52 "normal" rundowns and 2 rundowns that were subsequently considered to be "abnormal" and therefore indicative of an emerging fault; specifically datasets 39 and 43 shown in Fig. 4 (to provide a consistent reference datum between the rundown series, only vibration data between speeds of 6000 rpm and 1000 rpm are considered). Notably, novelty is not associated with rundowns that simply containing the highest resonant vibration peak, for instance, but are more associated with the relative features of the overall individual signature and how it compares with the collective (see Fig. 4(b) which shows a 2D contour plot of the collective vibration signatures).

Each of the 54 rundown datasets contains 5000 samples (rpm). A HC of the 5000 data samples (of rotor speed) is shown in Fig. 5(a), where the largest 10 clusters are selected according to a threshold of the HC distance measure. Each of the clusters is referred back to the original dataset, so that 10 segments for the 5000 speed samples can be found according to the HC dendrogram threshold, as shown in Fig. 5(b). It is seen from Fig. 5(b) that the regions of particular interest do have the highest density of segments, as required.

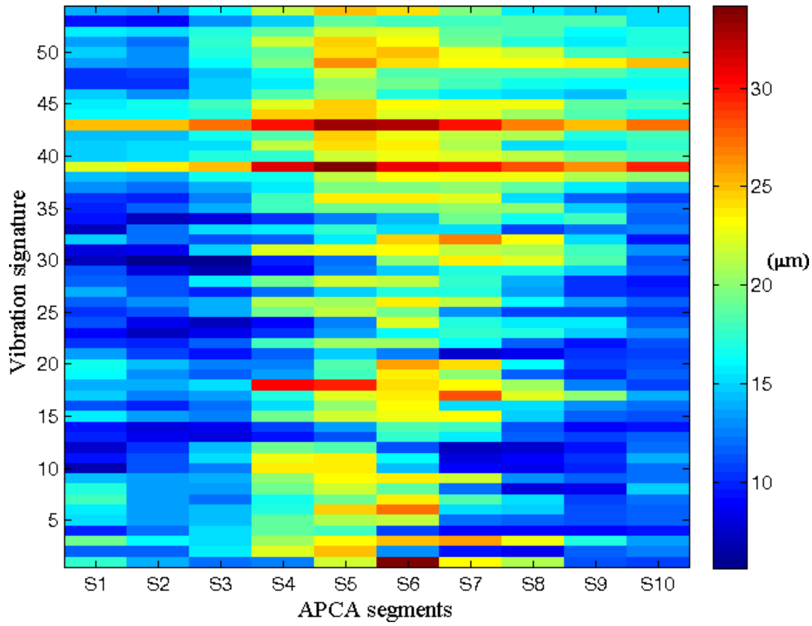
For comparison purposes, the segmentation resulting from traditional use of PAA/APCA (PAA modified using Haar DWTs[9]) and the proposed hybrid HC-PAA approach are shown in Table 2 along with the resulting vibration contour segments in Figs. 6(a)–6(c), respectively. A traditional k-means clustering is now applied to the extracted segmented features for classification, and hence novelty detection. The resulting clusters from the results of PAA, APCA and the hybrid HC-PAA are shown in Figs. 7(a)–7(c) respectively, with the set numbers included in the clusters and the cluster separation measure given in Table 3.

From Table 3, it can be seen that, whilst in all cases the faulted sets are clustered correctly, the hybrid HC-PAA provides a higher separation index compared to traditional PAA, with slight improvements also being evident compared to APCA. Notably, APCA and the hybrid HC-PAA required comparable computation times.

A SOMNN is considered a competitive learning ANN, using unsupervised learning to produce a discretized representation (typically in two dimensions) of an input space.[31] Here, SOMNN training is performed using the extracted features from the traditional PAA, APCA and hybrid HC-PAA results, using 10 elements and 54 samples in the network. The SOMNN is trained with the output space depicted as 2×2 hexagonal grids, using the MATLAB Neural Network Toolbox.[33] The 54 samples of the 10D data are projected onto the four neurons (clusters) that form a map in a 2D topologically (see the 2×2 hexagonal grids shown in Fig. 8, i.e., four elements/clusters). Through training, the reference vector of each neuron moves closer to the cluster center according to the samples that are clustered in the neuron, and the neighboring neurons also act to move closer to one another, eventually forming the final SOM after iteration. Sample hits, i.e., how many samples (out of the

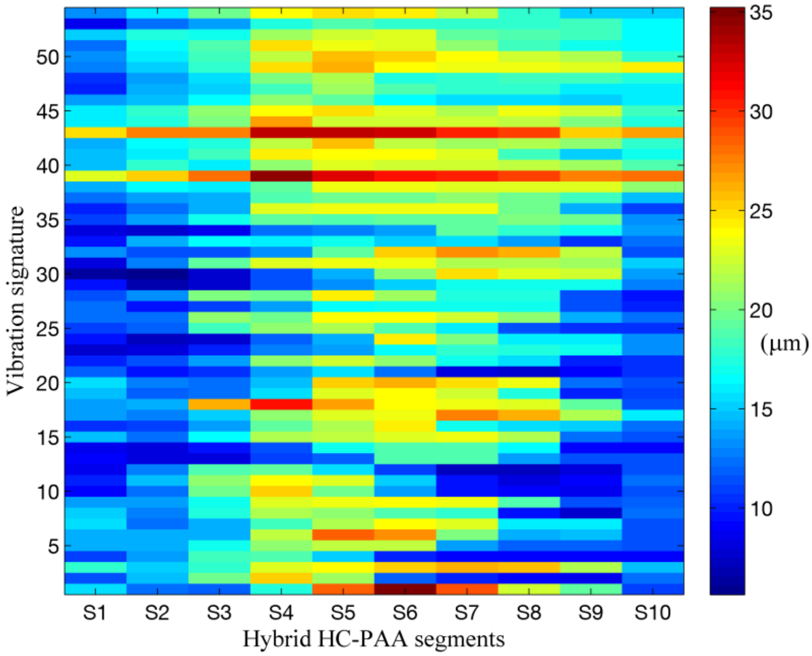


(a)



(b)

Fig. 6. (a) Traditional PAA representation, (b) APCA representation and (c) hybrid HC-PAA representation of the rundown vibration signatures (S = Segment).



(c)

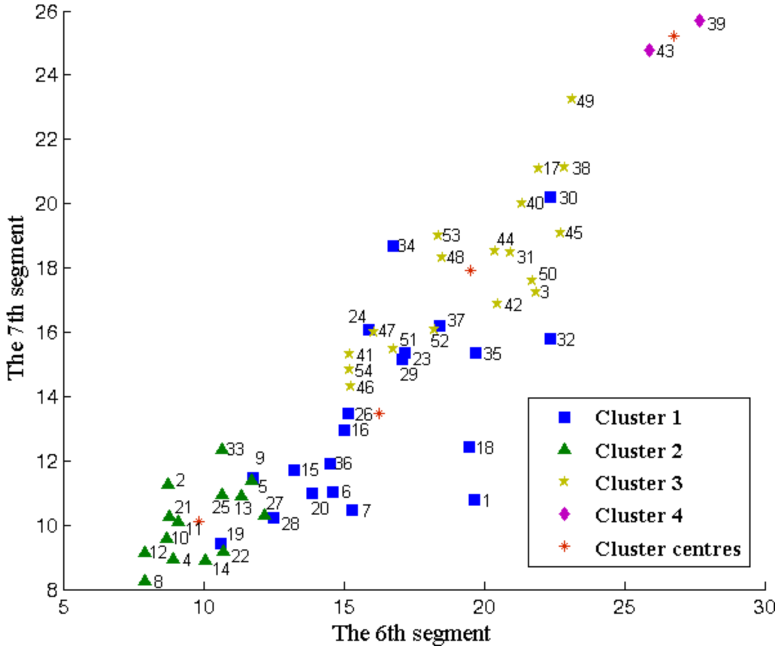
Fig. 6. (*Continued*)

54 samples) are clustered into each neuron, are shown in Figs. 8(a)–8(c) for extracted features using each of the three methods. For instance, for the top left node in Fig. 8(a), 12 samples from the original input data are clustered into the neuron. The set numbers, the clusters and the resulting cluster separation measures are given in Table 4.

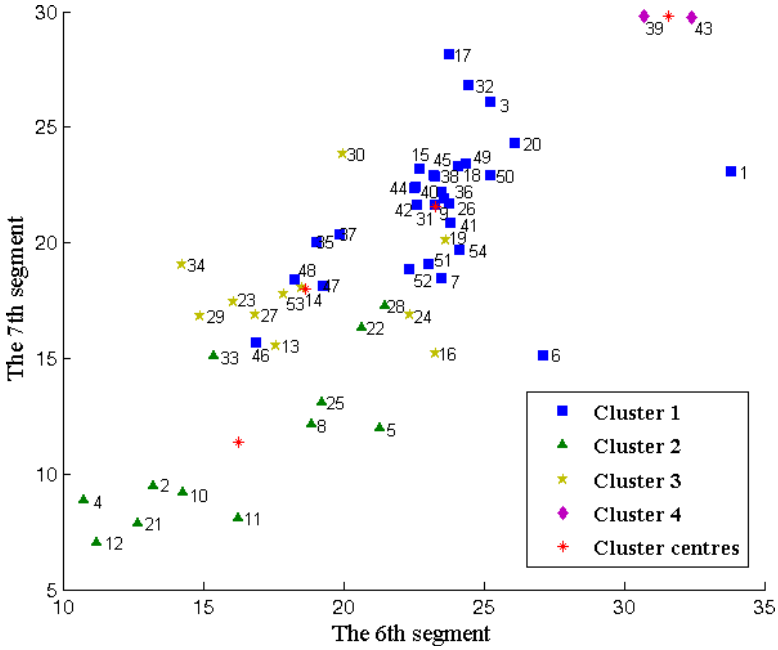
From Table 4 it is evident that whilst the extracted features using both APCA and hybrid HC-PAA have correctly identified the faults, hybrid HC-PAA provides the higher separation index, and hence best performance attributes.

3.2. *Ultrasonic human face identification*

An approach for biometric human face identification based on ultrasonic sensing has previously been reported in Ref. 34 that detects the geometric structure of human faces without being affected by the illumination characteristics of the surrounding environment. Multiple ultrasonic sensors (16 channels arranged in a 4×4 transmitter-receiver combination) are used for data collection, as shown in Fig. 9(a).[34] For this study, data relating to T0-R0 is considered, i.e., transmitter T0 emits one cycle of a Continuous Transmitted Frequency Modulated (CTFM) signal to the target face, and the receiver R0 detects the reflected echo. High Resolution Range Profiles

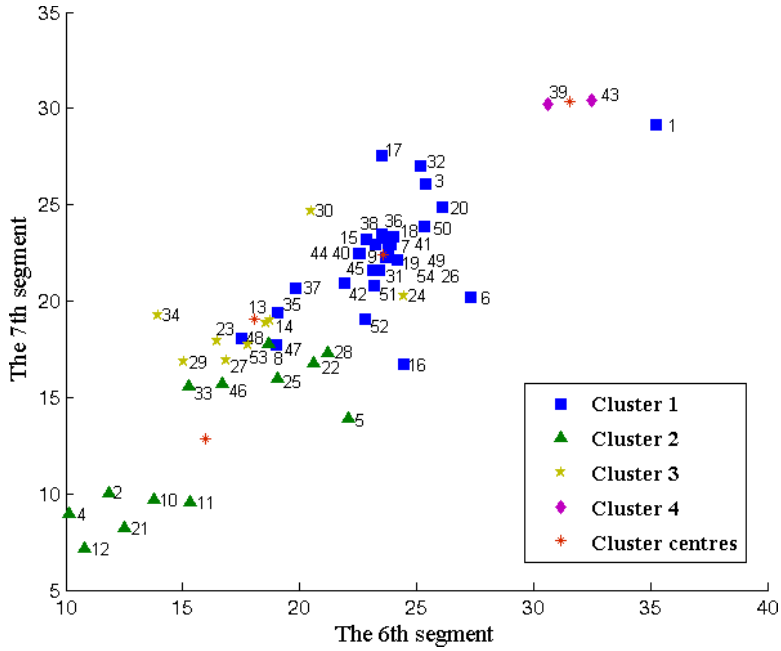


(a)



(b)

Fig. 7. Clustering results and extracted rundown features using (a) traditional PAA, (b) APCA and (c) hybrid HC-PAA.



(c)

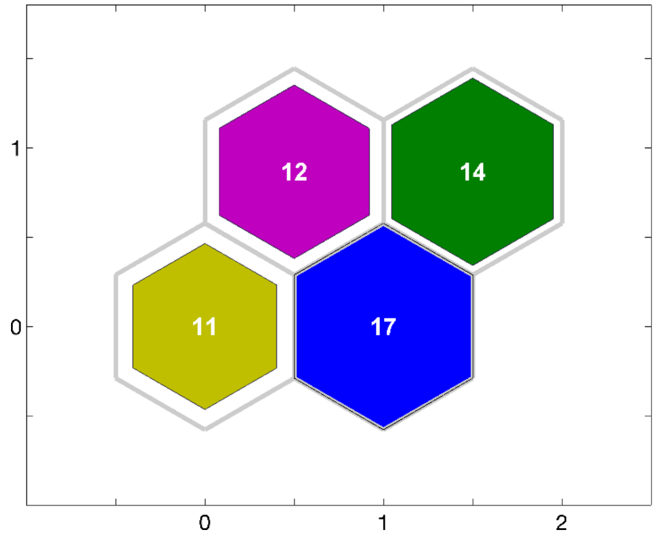
Fig. 7. (*Continued*)

(HRRPs) are obtained from the echo signals, where the normalized energy of different frequency components is calculated using the Fourier transform. A typical HRRP is shown in Fig. 9(b), where the y-axis is the normalized energy at each frequency, and the x-axis is the object (face) distance, which is linearly mapped from the frequency domain. The HRRP result shows the variation of the normalized

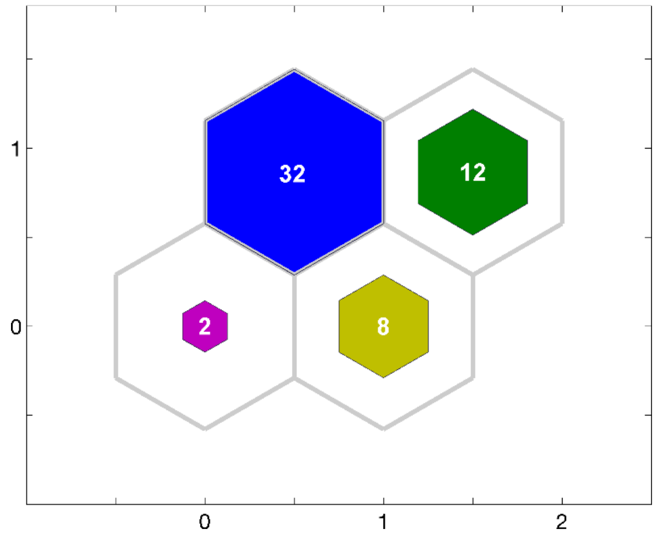
Table 3. Clustering results of extracted features using traditional PAA, APCA and hybrid HC-PAA approach.

	PAA	APCA	Hybrid HC-PAA
Cluster 1	1, 6, 7, 9, 15, 16, 18, 19, 20, 23, 24, 26, 28, 29, 30, 32, 34, 35, 36, 37	1, 3, 6, 7, 9, 15, 17, 18, 20, 26, 31, 32, 35, 36, 37, 38, 40, 41, 2, 44, 45, 46, 47, 48, 49, 50, 51, 52, 54	1, 3, 6, 7, 9, 15, 16, 17, 18, 19, 20, 26, 31, 32, 35, 36, 37, 38, 40, 41, 42, 44, 45, 47, 48, 49, 50, 51, 52, 54
Cluster 2	2, 4, 5, 8, 10, 11, 12, 13, 14, 21, 22, 25, 27, 33	2, 4, 5, 8, 10, 11, 12, 21, 22, 25, 28, 33	2, 4, 5, 8, 10, 11, 12, 21, 22, 25, 28, 33, 46
Cluster 3	3, 17, 31, 38, 40, 41, 42, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54	13, 14, 16, 19, 23, 24, 27, 29, 30, 34, 53	13, 14, 23, 24, 27, 29, 30, 34, 53
Cluster 4	39, 43	39, 43	39, 43
Cluster Separation index (mean)	167.5	181.7	182.0

energy with respect to object distance, where the peaks and troughs show the distribution of the scattering effect from the target face. For instance, the first peak in Fig. 9(b) represents the nose, and the highest peak represents the forehead (since the forehead has a wide reflection area, and hence provides higher energy). In this way, the geometrical features of the face can be represented by the HRRPs.

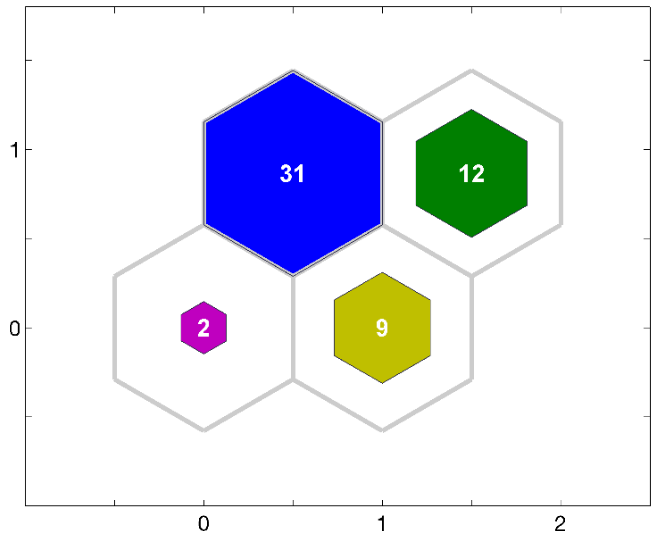


(a)



(b)

Fig. 8. SOMNN neuron sample hits from 54 run-down samples of extracted features using (a) traditional PAA, (b) APCA and (c) hybrid HC-PAA.



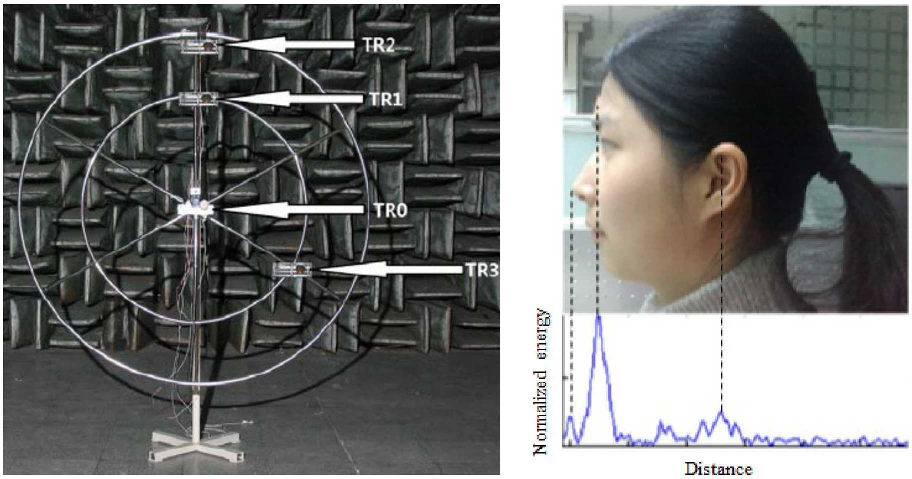
(c)

Fig. 8. (*Continued*)

Table 4. SOMNN clustering results of extracted features using traditional PAA, APCA and hybrid HC-PAA approach.

	PAA	APCA	Hybrid HC-PAA
Cluster 1	1, 6, 7, 9, 15, 16, 18, 19, 20, 26, 28, 36, 41, 46, 51, 52, 54	1, 3, 6, 7, 9, 15, 16, 17, 18, 19, 20, 26, 28, 31, 32, 35, 36, 37, 38, 40, 41, 42, 44, 45, 47, 48, 49, 50, 51, 52, 53, 54	1, 3, 6, 7, 9, 15, 16, 17, 18, 19, 20, 26, 28, 31, 32, 35, 36, 37, 38, 40, 41, 42, 44, 45, 47, 48, 49, 50, 51, 52, 54
Cluster 2	2, 4, 5, 8, 10, 11, 12, 13, 14, 21, 22, 25, 27, 33	2, 4, 5, 8, 10, 11, 12, 21, 22, 25, 33, 46	2, 4, 5, 8, 10, 11, 12, 21, 22, 25, 33, 46
Cluster 3	3, 17, 38, 39, 40, 42, 43, 44, 45, 49, 50	13, 14, 23, 24, 27, 29, 30, 34	13, 14, 23, 24, 27, 29, 30, 34, 53
Cluster 4	23, 24, 29, 30, 31, 32, 34, 35, 37, 47, 48, 53	39, 43	39, 43
Cluster Separation (mean)	Wrongly clustered	181.5	181.8

Here then, three faces are considered, and 30 HRRPs for each of the three faces are collected; respectively, sets 1–30, 31–60 and 61–90. Each HRRP (90 in total) has 540 sample points (which are the distance measures after pre-processing the raw data) as shown in Fig. 10. The HRRPs are approximated using a nominal selection of 10 segments using (a) traditional PAA with equally spaced sample regions, (b) APCA and (c) the proposed hybrid HC-PAA approach. A nominal cluster number of 3 is used since there are known to be three objects (faces).



(a) (b)

Fig. 9. Data collection for biometric ID: (a) arrangement of ultrasonic sensors; (b) example HRRP.[34]

The HC results of the 540 data samples are shown in Fig. 11(a). Notably in this case, S1 is closely associated with S10 as a subcluster. Since S1 and S10 need to be separated for FE purpose, in this case, the largest nine clusters are selected according the threshold of the HC distance measure, and S1 is further separated as an individual segment. In practice, this is simply accomplished using a basic "loop structure" in the HC-PAA algorithm, such that if a particular cluster involves multiple sample series, the cluster number (in this case, the original 10 clusters) from original HC results is decremented (in this case to 9, and S1 is clustered out separately).

Referencing each of the clusters back to the original measurements, the 10 segments for each of the 540 point datasets can be found, as shown in Fig. 11(b). It is evident that the highest density of the segments lay around the energy-rich characteristics, which S10 contains the majority of the low energy characteristic. The segmented regions resulting from traditional PAA, APCA and the hybrid HC-PAA are given in Table 5 for completeness.

The resulting representations from (traditional) PAA, APCA and hybrid HC-PAA applied to the HRRPs (from each of the three faces) are shown in Fig. 12(a)–12(c), respectively. It can be seen that significant differences in the results are evident. To provide a performance comparison, k-means clustering of the extracted features from each method is shown in Fig. 13(a)–13(c) respectively, and the HRRPs included in the clusters and the cluster separation measures are given in Table 6.

In this case, the clustering results for traditional PAA could not identify the three classes correctly, whilst APCA and the hybrid HC-PAA both correctly identified the three faces. Notably, again, the computation time of APCA and HC-PAA is comparable, however, HC-PAA gave significantly higher cluster separation and is therefore considered to provide a more robust solution.

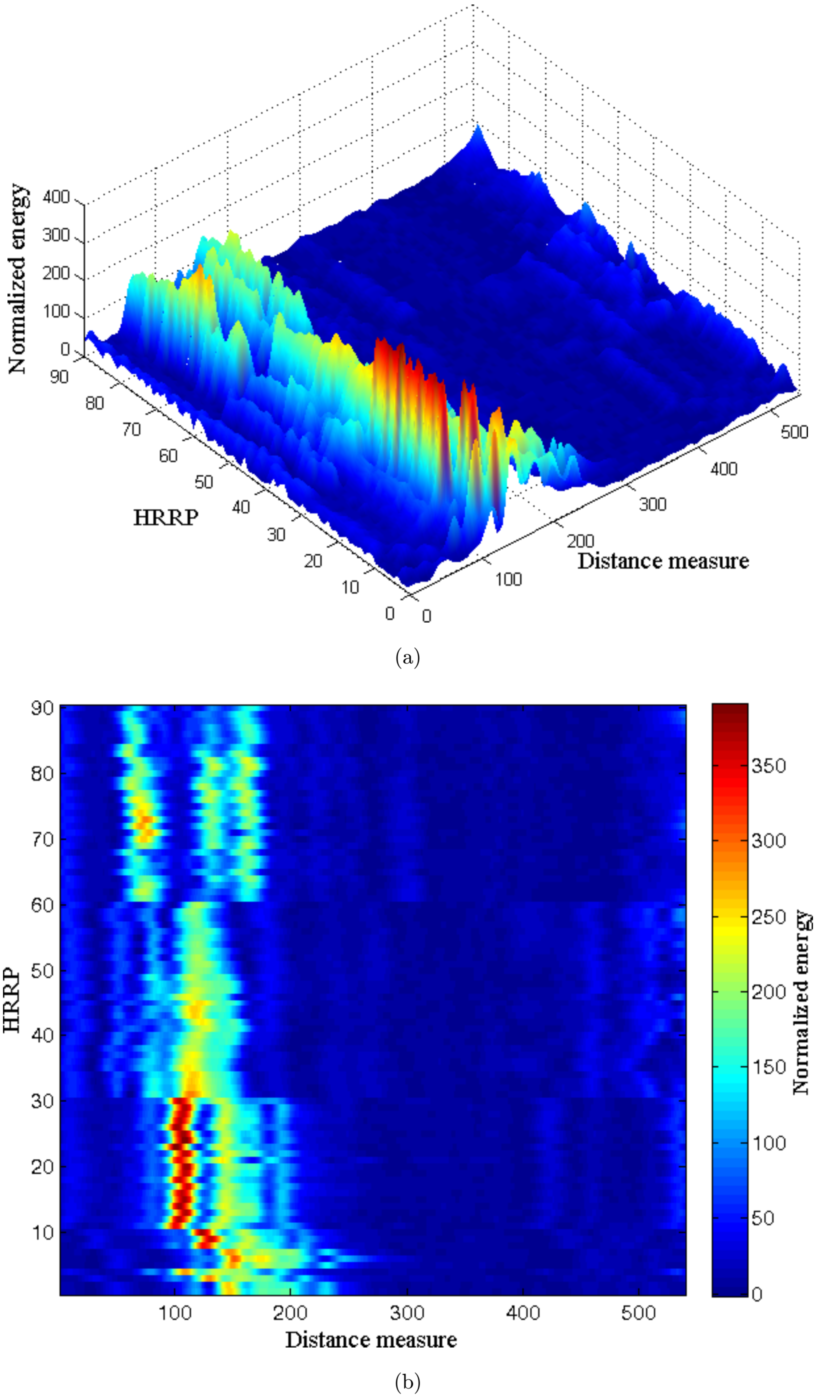
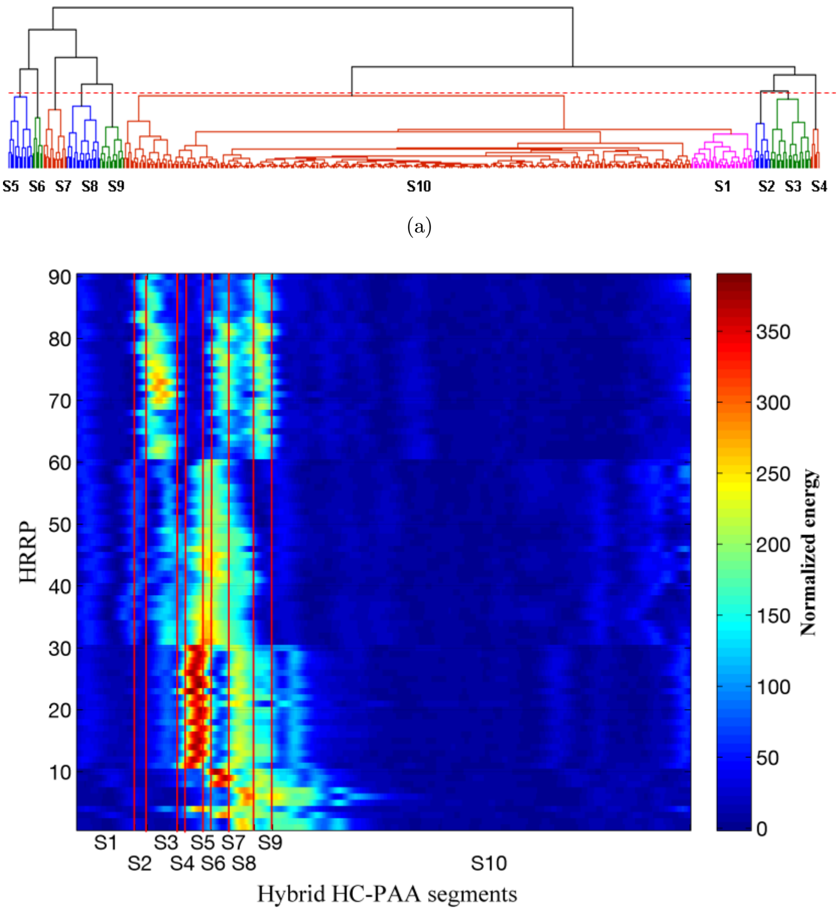


Fig. 10. HRRP representation of the three objects (faces): (a) 3D plot and (b) 2D contours.



(b)

Fig. 11. (a) HC dendrogram with 10 subclusters; (b) contour of hybrid HC-PAA segmentation of HRRPs.

Since the three classes are known in this case, the application can be considered as a classification problem. To provide a more comprehensive performance evaluation, a two-layer FFNN is applied to the extracted features from the traditional PAA, APCA and hybrid HC-PAA results. FFNN can be trained for classifications

Table 5. Distance samples included in the original PAA, APCA and the hybrid HC-PAA segments (S = Segment).

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Original PAA	1–54	55–108	109–162	163–216	217–270	271–324	325–378	379–432	433–486	487–540
APCA	1–55	56–75	76–93	94–108	109–129	130–149	150–170	171–229	230–422	423–540
HC-PAA	1–52	53–63	64–90	91–96	97–112	113–119	120–135	136–157	158–173	174–540

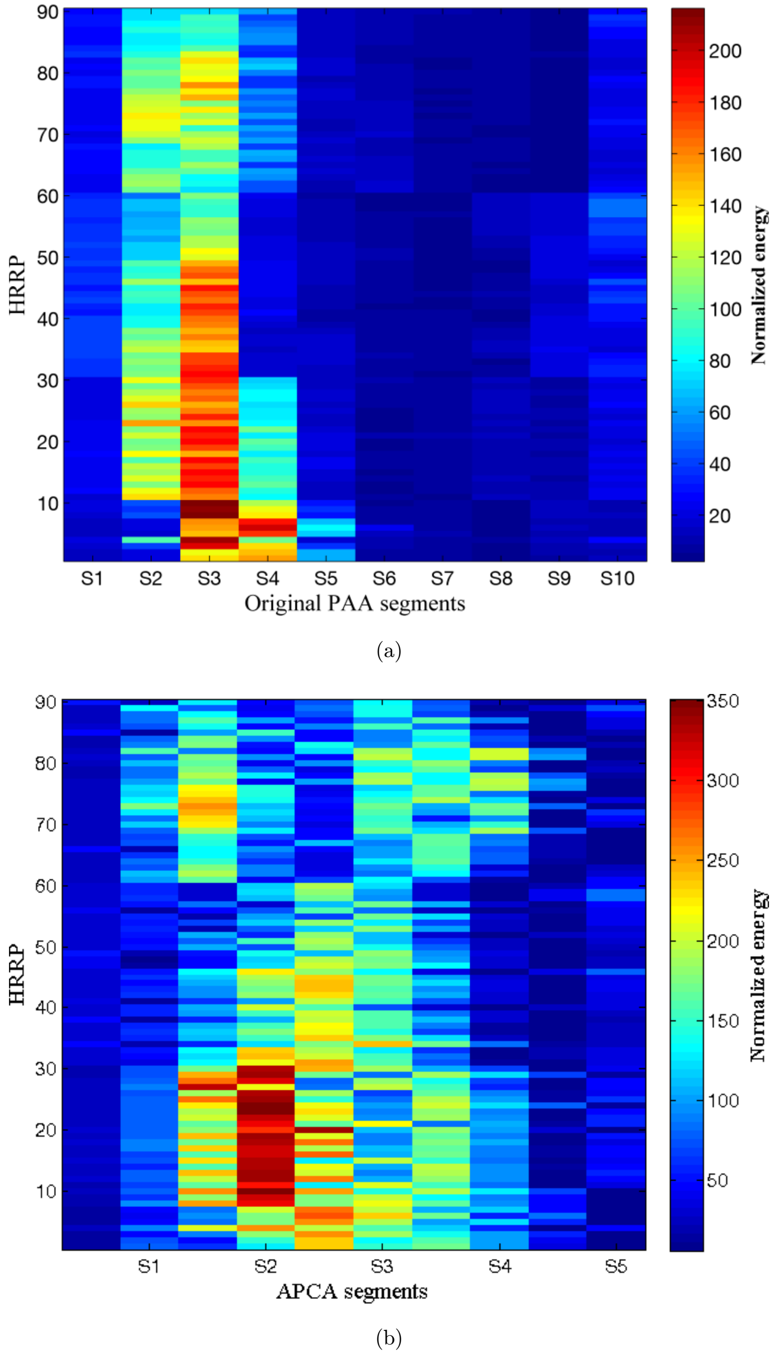
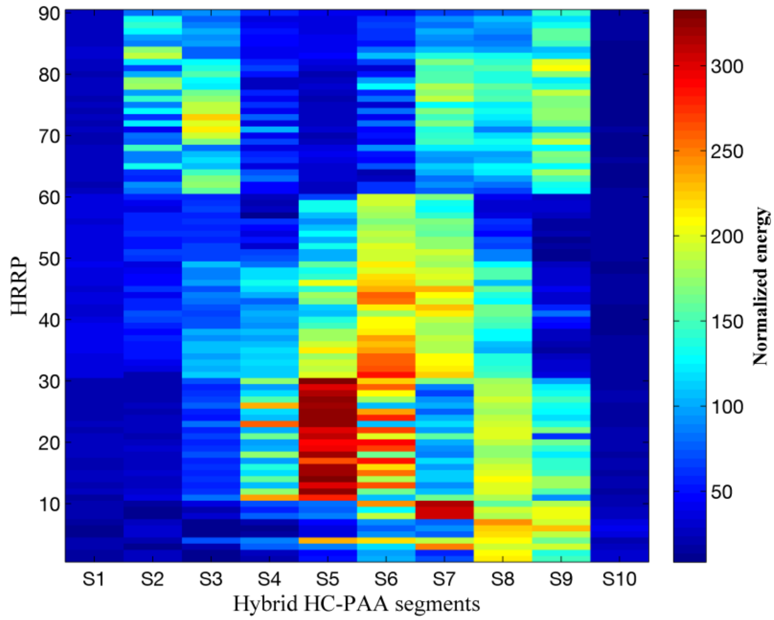
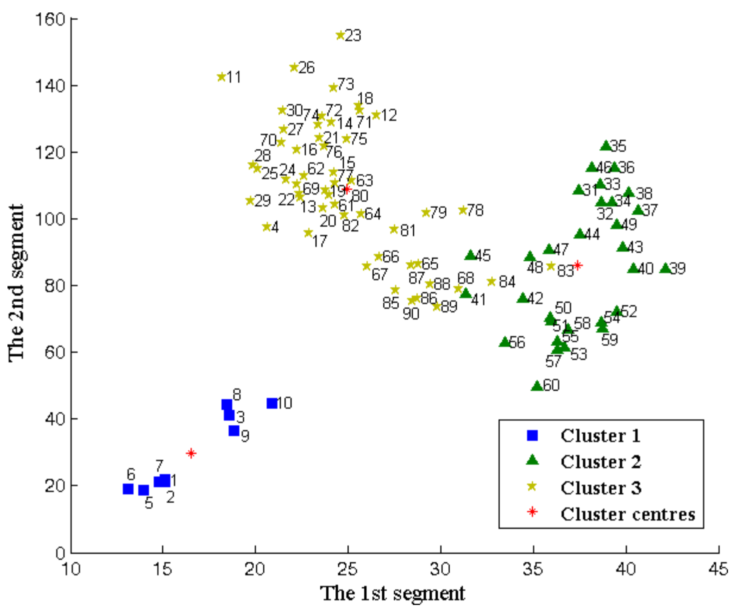


Fig. 12. Segment contours resulting from (a) traditional PAA; (b) APCA; (c) hybrid HC-PAA representation of the HRRPs (S = Segment).



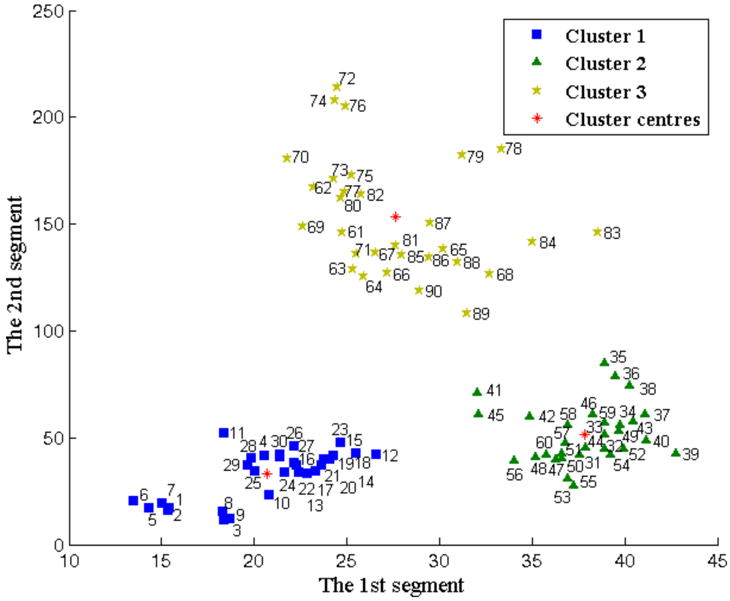
(c)

Fig. 12. (*Continued*)

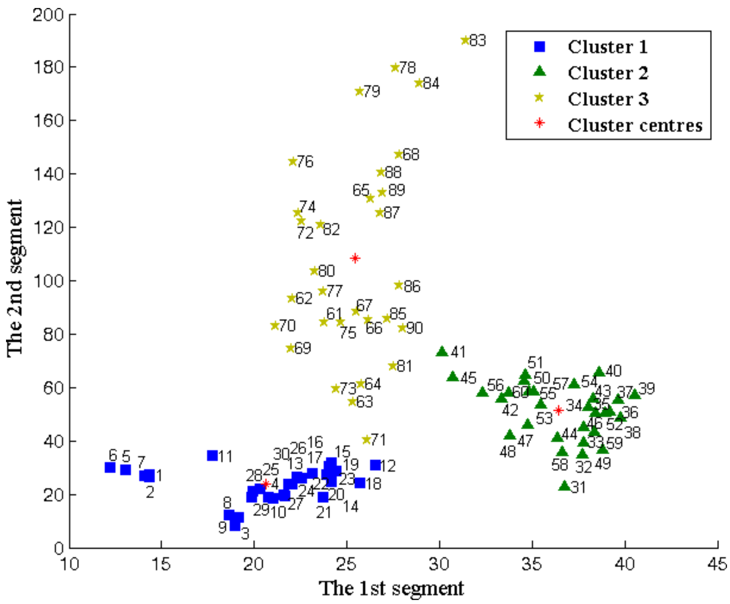


(a)

Fig. 13. Clustering results for extracted features for ultrasonic face identification using (a) traditional PAA, (b) APCA and (c) hybrid HC-PAA.



(b)



(c)

Fig. 13. (*Continued*)

Table 6. Clustering results for the traditional PAA, APCA and the hybrid HC-PAA extracted features for ultrasonic face identification.

	Original PAA	APCA	Hybrid HC-PAA
Cluster 1	1, 2, 3, 5, 6, 7, 8, 9, 10	1–30	1–30
Cluster 2	31–60	31–60	31–60
Cluster 3	4, 11–30, 61–90	61–90	61–90
Cluster separation (mean)	Wrong clusters	806.6	966.7

Table 7. FFNN classification performances for the traditional PAA, APCA and the hybrid HC-PAA extracted features for ultrasonic face identification.

	Original PAA	APCA	Hybrid HC-PAA
Classification performance (average)	3.278×10^{-7}	2.428×10^{-7}	1.011×10^{-7}

according to target classes, and classification performance is monitored through use of MSEs. Again, MATLAB Neural Network Toolbox[33] is employed.

Since performance can be affected by the initial conditions, 20 executions are initiated and the average MSE (performance) is used, with the results shown in Table 7 for (traditional) PAA, APCA and hybrid HC-PAA. From the results it is clear that the extracted features using hybrid HC-PAA provide lower MSEs, again indicating improved classification performance.

4. Conclusion

The paper has presented a basic method to improve the performance of traditional PAA by modifying segment frame sizes through the application of HC. Using the resulting hybrid HC-PAA as a FE methodology, pattern recognition is subsequently accomplished using k-means and ANNs. Two experimental trials have been used to demonstrate the efficacy of the technique, including industrial gas turbine fault detection based on rundown vibration signatures and a biometric face identification based on HRRPs from ultrasonic echo signals. Results show that the proposed hybrid HC-PAA provides the improved PAA FE performance by both increasing the classification performance and increasing the cluster separation distances in order to reduce the chance of misclassification. HC-PAA is also shown to provide improved performance compared to APCA (an improved and commonly used method) by demonstrating greater cluster separation measures and classification performance for the two case studies. Through additional performance comparisons with other well-known techniques, the proposed methodology has been shown to provide a computationally efficient and robust method of FE/novelty detection on large datasets for a diverse spectrum of applications. It should be noted that, whilst the proposed HC-PAA has been developed for FE, the underlying principles are much more widely applicable to other FS, data reduction and the rapid identification of information rich portions of large data series.

Abbreviations

ANN — Artificial Neural Network
APCA — Adaptive Piecewise Constant Approximation
CTFM — Continuous Transmitted Frequency Modulated
DFT — Discrete Fourier Transform
DWT — Discrete Wavelet Transform
FE — Feature Extraction
FFNN — Feed-Forward Neural Network
FS — Feature Selection
HC — Hierarchical Clustering
HRRP — High Resolution Range Profile
ICA — Independent Component Analysis
MSE — Mean Squared Error
PAA — Piecewise Aggregate Approximation
PCA — Principal Component Analysis
PP — Projection Pursuit
SOMNN — Self-Organizing Map Neural Network
SVD — Singular Value Decomposition

Acknowledgments

The authors would like to thank Siemens Industrial Turbomachinery, Lincoln, UK for providing data to support the research outcomes. The authors would also like to thank Prof. Eamonn Keogh for providing the APCA code for performance comparisons.

References

1. T. C. Fu, A review on time series data mining. *Eng. Appl. Artif. Intel.* **24**(1) (2011) 164–181, doi: 10.1016/j.engappai.2010.09.007.
2. J. Ren, J. Zabalza, S. Marshall and J. Zheng, Effective feature extraction and data reduction in remote sensing using hyperspectral imaging, *IEEE Signal Process. Mag.* **31** (4) (2014) 149–154, doi:10.1109/MSP.2014.2312071.
3. D. Garrett, D. A. Peterson, C. W. Anderson and M. H. Thaut, Comparison of linear, nonlinear, and feature selection methods for EEG signal classification, *IEEE Trans. Neural Syst. Rehabil. Eng.* **11**(2) (2003) 141–144, doi: 10.1109/TNSRE.2003.814441.
4. D. Ping Tian, A review on image feature extraction and representation techniques, *Int. J. Multimed. Ubiq. Eng.* **8**(4) (2013) 385–396.
5. Y. Zhang, C. Bingham, Z. Yang, B. W. Ling and M. Gallimore, Machine fault detection by signal denoising — with application to industrial gas turbines, *Measurement* **58** (2014) 230–240, doi: 10.1016/j.measurement.2014.08.020.
6. L. Jiang, J. Xuan and T. Shi, Feature extraction based on semi-supervised kernel Marginal Fisher analysis and its application in bearing fault diagnosis, *Mech. Syst. Signal Process.* **41**(1–2) (2013) 113–126, doi: 10.1016/j.ymssp.2013.05.017.

7. H. Abdi and L. J. Williams, Principal component analysis, *Wiley Interdiscip. Rev.: Comput. Statist.* **2** (2010) 433–459, doi: 10.1002/wics.101.
8. L. O. Jimenez-Rodriguez, E. Arzuaga-Cruz and M. Velez-Reyes, Unsupervised linear feature-extraction methods and their effects in the classification of high-dimensional data, *IEEE Trans. Geosci. Remote Sens.* **45**(2) (2007) 469–483, doi: 10.1109/TGRS.2006.885412.
9. I. K. Fodor, A survey of dimension reduction techniques, Technical Report, Lawrence Livermore National Lab, CA, US (2002), doi: 10.2172/15002155.
10. A. Hyvärinen and O. Erkki, Independent component analysis: Algorithms and applications, *Neural Netw.* **4–5**(13) (2000) 411–430, doi: 10.1016/s0893-6080(00)00026-5.
11. Y. Zhang and Y. Zhao, Modulation domain blind speech separation in noisy environments, *Speech Commun.* **55**(10) (2013) 1081–1099, doi: 10.1016/j.specom. 2013.06.014.
12. J. Wu, M. Liu and L. Jin, A hybrid support vector regression approach for rainfall forecasting using particle swarm optimization and project pursuit technology, *Int. J. Comput. Intel. Appl.* **9**(2) (2010) 87–104, doi: 10.1142/S1469026810002793.
13. E. Rodriguez-Martinez, J. Y. Goulermas, T. Mu and J. F. Ralph, Automatic induction of projection pursuit indices, *IEEE Trans. Neural Netw.* **21**(8) (2010) 1281–1295, doi: 10.1109/TNN.2010.2051161.
14. Z. Li, X. Yan, Z. Tian, C. Yuan, Z. Peng and L. Li, Blind vibration component separation and nonlinear feature extraction applied to the nonstationary vibration signals for the gearbox multi-fault diagnosis, *Measurement* **46**(1) (2013) 259–271, doi: 10.1016/j.measurement.2012.06.013.
15. W. Dai, J. Y. Wu and C. J. Lu, Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes, *Expert. Syst. Appl.* **39**(4) (2012) 4444–4452, doi: 10.1016/j.eswa.2011.09.145.
16. E. Keogh and M. J. Pazzani, A simple dimensionality reduction technique for fast similarity search in large time series databases, in *Proc. Pacific Asia Conf. Knowledge Discovery and Data Mining* (2000), pp. 122–133.
17. B. K. Yi and C. Faloutsos, Fast time sequence indexing for arbitrary lp norms, in *Proc. VLDB '00* (Morgan Kaufmann Publishers, 2000), pp. 385–394.
18. J. Austin, R. Davis, M. Fletcher, T. Jackson, M. Jessop, B. Liang and A. Pasley, DAME: Searching large data sets within a grid-enabled engineering application, *Proc. IEEE* **93**(3) (2005) 496–509, doi: 10.1109/jproc.2004.842746.
19. A. Zifan, S. Saberi, M. H. Moradi and F. Towhidkhah, Automated ECG segmentation using piecewise derivative dynamic time warping, *Int. J. Biomed. Sci.* **1**(3) (2006) 181–185.
20. Y. Zhang and J. R. Glass, A piecewise aggregate approximation lowerbound estimate for posterorgram-based dynamic time warping, in *InterSpeech* (2011), pp. 1909–1912.
21. E. Keogh, K. Chakrabarti, M. J. Pazzani and S. Mehrotra, Dimensionality reduction for fast similarity search in large time series databases, *Knowl. Inf. Syst.* **3**(3) (2001) 263–286, doi: 10.1007/PL00011669.
22. K. Chakrabarti, E. Keogh, S. Mehrotra and M. J. Pazzani, Locally adaptive dimensionality reduction for indexing large time series databases, *ACM Trans. Database Syst.* **27**(2) (2002) 188–228, doi: 10.1145/568518.568520.
23. M. Gallimore, M. J. Riley, C. M. Bingham, Z. Yang and Y. Zhang, Intelligent diagnostics and prognostics for industrial machines using an optimization approach, in *Proc. 4th Int. Conf. IRF'13* (2013), pp. 223–224.
24. F. Murtagh and P. Contreras, Algorithms for hierarchical clustering: An overview, *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.* **2**(1) (2012) 86–97.

25. H. Nishizawa, T. Obi, M. Yamaguchi and N. Ohyama, Hierarchical clustering method for extraction of knowledge from a large amount of data, *Opt. Rev.* **6**(4) (1999) 302–307, doi: 10.1007/s10043-999-0302-2.
26. O. Maqbool and H. A. Babri, Hierarchical clustering for software architecture recovery, *IEEE Trans. Softw. Eng.* **33**(11) (2007) 759–780, doi: 10.1109/tse.2007.70732.
27. A. Barirani, B. Agard and C. Beaudry, Competence maps using agglomerative hierarchical clustering, *J. Intell. Manuf.* **24**(2) (2013) 373–384, doi: 10.1007/s10845-011-0600-y.
28. T. Hastie, R. Tibshirani and J. Friedman, "*14.3.12 Hierarchical Clustering*". *The Elements of Statistical Learning*, 2nd edn. (Springer, New York, 2009).
29. Y. Zhang, C. M. Bingham and M. Gallimore, Applied sensor fault detection, identification and data reconstruction, *Adv. Military Technol.* **8**(2) (2013) 13–26.
30. P. S. Bishnu and V. Bhattacherjee, Software fault prediction using quad tree-based *k*-means clustering algorithm, *IEEE Trans. Knowl. Data Eng.* **24**(6) (2012) 1146–1150, doi: 10.1109/TKDE.2011.163.
31. T. Kohonen, The self-organizing map, *Proc. IEEE* **78**(9) (2002) 1464–1480, doi: 10.1109/5.58325.
32. D. Svozil, V. Kvasnicka and J. Pospichal, Introduction to multi-layer feed-forward neural networks, *Chemometr. Intell. Lab.* **39**(1) (1997) 43–62, doi: 10.1016/S0169-7439(97)00061-0.
33. Matlab version 7.12.0, Natick Massachusetts, The Mathworks Inc. (2012).
34. Y. Xu, J. Y. Wang, B. X. Cao and J. Yang, Multi sensors based ultrasonic human face identification: Experiment and analysis, in *Proc. IEEE MFI'12* (2012), pp. 257–261.