Time–Frequency Analysis for Efficient Fault Diagnosis and Failure Prognosis for Interior Permanent-Magnet AC Motors

Elias G. Strangas, *Member, IEEE*, Selin Aviyente, *Member, IEEE*, and Syed Sajjad H. Zaidi, *Student Member, IEEE*

***Abstract*—The detection of noncatastrophic faults in conjunction with other factors can be used to determine the remaining life of an electric drive. As the frequency and severity of these faults increase, the working life of the drive decreases, leading to eventual failure. In this paper, four methods to identify developing electrical faults are presented and compared. They are based on the short-time Fourier transform, undecimated-wavelet analysis, and Wigner and Choi–Williams distributions of the field-oriented currents in permanent-magnet ac drives. The different fault types are classified using the linear-discriminant classifier and k-means classification. The comparison between the different methods is based on the number of correct classifications and Fisher's discriminant ratio. Multiple-class discrimination analysis is also introduced to remove redundant information and minimize storage requirements.**

***Index Terms*—Choi–Williams, intermittent fault detection, multiple discrimination analysis (MDA), permanent-magnet ac (PMAC) drives, short-time Fourier transform (STFT), undecimated discrete wavelet transform (UDWT), Wigner distribution.**

I. Introduction

In THE LAST decade, there has been extensive work in the field of fault diagnosis of electric drives, using primarily signal-processing methods [1], [2]. Such methods range from the extensive use of artificial intelligence and intelligent supervisory systems [3] to the harmonic analysis of the supply current in the inverter [4]. The problem of failure prognosis is more demanding and has received increased attention only in the last few years [5]. Among the major challenges in failure prognosis are the need for a large amount of training data and increased computational complexity.

Diagnosis refers to identifying the fault that will lead to a failure and estimating its severity. Prognosis, on the other hand, includes a set of tasks and decisions that can be described as the following definitions.

1) Detect and identify the conditions that may lead to a failure of a component or a system before its performance has been disrupted.
2) Predict the time until a failure will occur and identify the conditions that may accelerate or decelerate this progress.

Failure prognosis is often associated with continuous monitoring of variables of operation and diagnosis of early fault manifestation. It is clear that fault diagnosis and failure prognosis have commonalities and can be addressed within the same framework.

Recognizing that a departure of the performance parameters and operational variables from expected is a result of a fault, and identifying this fault and its severity requires an *a priori* knowledge of how a fault manifests itself. In previous work, we have discussed such fault recognition using standard methods [6]. We present here methods to compare the quality of such analysis methods, as well as methods to categorize a fault from the coefficients estimated from this analysis.

Determining the fault type and its severity requires the collection of signals from both healthy and faulted drives, processing of the signals to extract their salient characteristics, and, finally, the detection and classification of faults. In order to train the different classifiers based on the different transform methods, we used data from a set of previous extensive experiments on a permanent-magnet ac (PMAC) motor, discussed in [6]. They include two types of intermittent electrical faults and one mechanical fault which is also treated as intermittent. In this paper, we first study the effect of the analysis methodology on the overall accuracy of the fault detection and categorization. We selected as the four analysis methods to compare the short-time Fourier transform (STFT), undecimated wavelet transform, Wigner distribution, and Choi–Williams distribution (CWD).

Furthermore, we evaluate two different classification methods used for categorizing the faults. They are the linear-discriminant classifier (LDC) and the k-means classifier. Our preliminary results in evaluating the different transform methods indicate that there is a lot of redundant information in the extracted time–frequency feature vectors. To limit the amount of redundant information, we apply a linear transformation of the time–frequency features based on the multiple-class discriminant analysis (MDA). We compare the performance of the different transform methods using Fisher's coefficient and the classification results after MDA.

In Sections II and III, we discuss the different time–frequency analysis and classification methods and introduce Fisher's discriminant ratio as a way to compare them. In Section IV, we present the different types of faults and the experimental setup. In Section V, we discuss details of implementation and experimental results, while in Section VI, we propose MDA as a method to further optimize the feature-extraction process. In Section VII, we compare the different analysis and categorization methods introduced in this paper in terms of computational complexity and classification accuracy.

Manuscript received March 3, 2008; revised September 19, 2008. First published October 31, 2008; current version published December 2, 2008.

The authors are with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824 USA (e-mail: strangas@egr.msu.edu; aviyente@egr.msu.edu; zaidisyl@msu.edu).

Digital Object Identifier 10.1109/TIE.2008.2007529

II. Analysis Methods

In this section, four different time–frequency analysis methods that are used in fault detection and classification are discussed. These distributions are the STFT, undecimated discrete wavelet transform (UDWT), Wigner–Ville distribution (WVD), and CWD.

A. STFT

The STFT [7] is an extension of the Fourier transform for the analysis of nonstationary signals. In STFT, a signal is divided into small time windows, and each is analyzed using the Fourier transform as follows:

$$\text{STFT}(t,\omega) = \int h(t-\tau)s(\tau)e^{-j\omega\tau}d\tau \tag{1}$$

where s is the signal, h is the window function, and t and ω are the time and frequency parameters, respectively. This formulation provides localization in time while simultaneously capturing frequency information. The time–frequency tiling for the STFT is uniform across time and frequency. In the implementation of the STFT, a design tradeoff must be made between time and frequency resolution. A short-duration window provides good time resolution at the expense of poor frequency resolution, whereas a long-duration window provides good frequency resolution at the expense of reduced-time resolution.

B. Undecimated-Wavelet Analysis

Another common tool used for nonstationary signals is the discrete wavelet transform (DWT) [8], which has greater flexibility than the STFT. Different basis functions, or mother wavelets, may be used in wavelet analysis while the basis functions for Fourier analysis are always complex exponentials. Unlike sinusoids, wavelets have finite duration, and their energy is localized around a point in time and, thus, are more suited in detecting transient activity. One can choose or design a wavelet to achieve the best results for a specific application. The UDWT [9] is a shift-invariant version of the DWT and is computed as

$$\begin{aligned} c_{j+1}[l] &= \sum_k h[k]c_j[l+2^j k] \\ d_{j+1}[l] &= \sum_k g[k]c_j[l+2^j k] \end{aligned} \tag{2}$$

where d_j are the wavelet coefficients and c_j are the approximation coefficients at scale j and h and g are the low- and high-pass filters corresponding to a particular wavelet family, respectively. c_0, the approximation coefficients at the lowest scale, correspond to the samples of the signal.

C. Wigner Distribution

Wigner or WVD is another nonstationary signal processing method belonging to Cohen's class and is defined as [7]

$$W(t,\omega) = \int s\left(t+\frac{\tau}{2}\right)s^*\left(t-\frac{\tau}{2}\right)e^{-j\omega\tau}d\tau \tag{3}$$

where s is the signal and t and ω are the time and frequency parameters, respectively. Wigner distribution gives the energy distribution of the signal as a function of time and frequency by taking the Fourier transform of the local autocorrelation function of the signal. It is known to have high time–frequency resolution and does not suffer from the temporal versus frequency-resolution tradeoff encountered in STFT. Wigner distribution satisfies both the time and frequency marginals and preserves the energy of the underlying signal. The major shortcoming of Wigner distribution occurs for multicomponent signals in terms of the cross terms. The cross terms occur due to the bilinear nature of the Wigner distribution and can sometimes obstruct the actual energy distribution.

D. CWD

CWD is a filtered version of the Wigner distribution that reduces the cross terms between the different signal components and is defined as

$$C(t,\omega) = \iiint \phi(\theta,\tau)s\left(u+\frac{\tau}{2}\right)s^*\left(u-\frac{\tau}{2}\right) \times e^{j(\theta u-\theta t-\tau\omega)}du\,d\theta\,d\tau \tag{4}$$

where $\phi(\theta,\tau) = \exp(-((\theta\tau)^2/\sigma))$ is the kernel function that acts as a filter on the signal's autocorrelation function. This distribution can be thought of as a filtered/smoothed version of the Wigner distribution, and the amount of smoothing is controlled by σ. This smoothing removes the cross terms seen in the Wigner distribution at the expense of reduced resolution. CWD also satisfies the time and frequency marginals and preserves the energy of the underlying signal.

III. Pattern Recognition

Once the analysis coefficients have been computed, the presence of a fault is detected based on thresholding the energy of the coefficients across the different frequency bands. These energy values are formed into vectors that provide the input features for the classification algorithm. The next step is to categorize these feature vectors, with the categories corresponding to one of several known fault types.

In this paper, we consider two different classification methods: LDC and k-means classifier.

A. LDC

The LDCs are trained on the input feature vectors, since no *a priori* knowledge of a probability distribution for the sample points is assumed. The feature space is divided into K regions, each having its own weighting coefficients. Based on the training data, the linear-discriminant functions (5) [10] are defined as

$$D_k(\mathbf{x}) = x_1\alpha_{1k} + x_2\alpha_{2k} + \cdots + x_N\alpha_{Nk} + \alpha_{N+1,k}, \quad k = 1,2,\ldots,C \tag{5}$$

where $\mathbf{x}$ is the N-dimensional feature vector and α are the normalized weighting coefficients for the kth class.

Linear-discriminant functions were chosen for the algorithm, since they are computationally efficient. A sample vector belongs to a particular class if the discriminant function is greater for that class than for any other class, i.e., $\mathbf{x}$ belongs to class C_j if

$$D_j(\mathbf{x}) > D_k(\mathbf{x}), \qquad \text{for every } k \neq j.$$

The weighting coefficients are adjusted from their initial guess through a training procedure using sample vectors for which the proper classification is known. The algorithm for this procedure makes adjustments to the weighting coefficients until each sample vector is correctly classified.

Young and Calvert [10] show that this training algorithm will converge in a finite number of steps. When a sample vector $\mathbf{x}$ is correctly classified, no adjustment to the weighting coefficients is made. When a sample vector is incorrectly classified, or

$$D_j(\mathbf{x}) \leq D_l(\mathbf{x})$$

where

$$D_l(\mathbf{x}) = \max_{l \neq j} [D_1(\mathbf{x}), \ldots, D_C(\mathbf{x})]$$

adjustments are made to $\boldsymbol{\alpha}_j$ and $\boldsymbol{\alpha}_l$ as shown in

$$\boldsymbol{\alpha}_j(i+1) = \boldsymbol{\alpha}_j(i) + a\mathbf{x} \tag{6}$$
$$\boldsymbol{\alpha}_l(i+1) = \boldsymbol{\alpha}_l(i) - a\mathbf{x} \tag{7}$$

where a is a gain constant.

Discriminant functions have minimal storage requirements after the training phase, since for each class, only a single vector of weighting coefficients need to be stored. Storage of training samples is no longer required during the classification phase. For multiclass problems ($C > 2$), it can be said that the classes are linearly separable if linear-discriminant functions $D_1(\mathbf{x}), \ldots, D_C(\mathbf{x})$ exist, such that (8) is true

$$D_j(\mathbf{x}) > D_k(\mathbf{x}), \qquad \text{for every } \mathbf{x} \text{ in } C_j \text{ and all } k \neq j. \tag{8}$$

B. k-Means Classification

k-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed *a priori*. The main idea is to define k centroids, one for each cluster. These centroids can be chosen in different ways, such as choosing k random points or using the means from training data as centroids. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. In pattern recognition, k-means is a method for classifying objects based on closest training examples in the feature space. The distance is usually measured using the Euclidean distance. When all of the data points have been assigned to clusters, this procedure is repeated by measuring the means of each cluster and using these as the centroids for clustering. This procedure is repeated until there are no changes in the k centroids.

In this paper, we implement an online version of this algorithm. The initial centroids are estimated based on the training data, and each new fault sample is assigned to the class to which it is closest to, i.e., $C_j = \arg\min_{i \in \{1,2,\ldots,C\}} \|\mathbf{x}_j - \mathbf{m}_i\|$, where $\mathbf{m}_i$ is the different class averages obtained from the training data.

C. Evaluation of Features: Fisher Discriminant Ratio

The Fisher discriminant ratio is a measure that quantifies the discrimination capacity of features regardless of the classifier. In this paper, we are considering four different signal transforms, i.e., four different ways of extracting features from the fault signals. The classification results obtained by the different classifiers give information about which class each fault belongs to. However, these results are not necessarily informative of the separation of the different fault classes and how the time–frequency coefficients cluster with the different methods. The goodness of the features can be quantified through the classification accuracy or through Fisher discriminant ratio, which is independent of the classifier. In this paper, we will use both methods and show that the results are consistent.

Fisher's discriminant ratio [11] can be used both as a classification method and as a class-separability criterion. In this paper, we will adopt the latter approach and use it as an indicator of the class separability for the four different time–frequency analysis methods. For multiclass data, Fisher's discriminant ratio is defined as

$$F(\mathbf{X}) = \frac{S_B}{S_W} = \frac{\left\| \sum_{i=1}^{C} K_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^{\mathrm{T}} \right\|_2^2}{\sum_{i=1}^{C} s_i^2}$$

$$\mathbf{m_i} = \frac{1}{K_i} \sum_{j \in C_i} \mathbf{x}_j$$

$$s_i^2 = \frac{1}{K_i} \sum_{j \in C_i} \|\mathbf{x}_j - \mathbf{m}_i\|_2^2$$

$$\mathbf{m} = \frac{1}{K} \sum_{i=1}^{K} \mathbf{x}_i \tag{9}$$

where $\mathbf{X}$ is the matrix of the extracted features, K_i is the number of samples in each class, $\mathbf{m}_i$ is the mean of each class, s_i^2 is the variance within each class, and K is the total number of samples. Fisher's discriminant ratio can be interpreted as the ratio of the interclass distance to inner class scatter. Fisher's criterion is motivated by the intuitive idea that the discrimination power is maximized when the spatial distribution of different classes are as far away as possible from each other and the spatial distribution of samples from the same class are as close as possible to each other.

In the application proposed in this paper, this criterion will be applied to the time–frequency features extracted by the four methods to compare the discrimination power of the different transforms. In this case, the number of classes $C = 5$, and $\mathbf{X}$ is the matrix of the time–frequency coefficients.

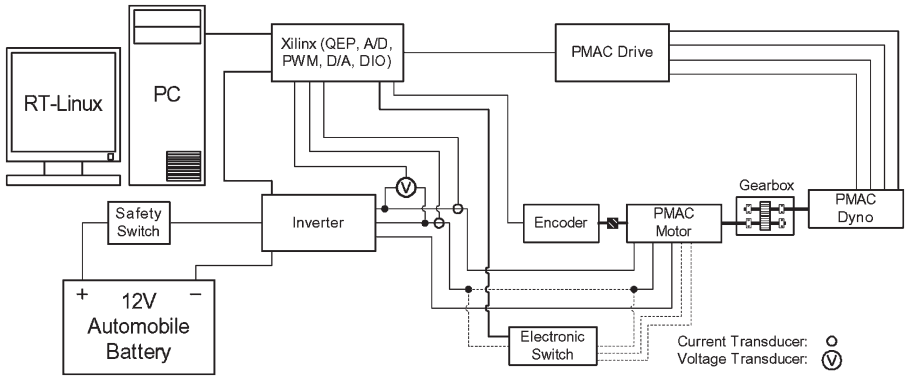


Fig. 1. Experimental setup.

IV. Types of Faults and Experimental Setup

To collect data from healthy and faulted drives, it is common to test a drive in its healthy state and, subsequently, test it under a set of faulted conditions. This generally requires permanent and repeatable modification or damage to it, in order to change its characteristics.

As interest in drive reliability has led to increased work in fault diagnosis and failure prognosis, it also has resulted in highly reliable motors. This development, however desirable, makes it difficult to collect in the field the data needed to develop the necessary database for fault diagnosis. It makes it even more difficult to collect the historic data that will help determine the model necessary to track the evolution of faults and failure prognosis.

Alternatives include simply operating the drive until failure [12] or designing experiments where faults could be introduced during operation. The most acceptable method for data collection to develop and test fault-detection strategies remains experimental [6], [12]–[16].

In this paper, we are comparing analysis methods that are useful in detecting the transient phenomena associated with the inception and clearing of faults. In order to collect the data required to design and test the algorithms, an experiment was developed to mimic the electrical and mechanical faults under study. Data from these faults were extracted from continuously sampled signals. Issues specific to transient fault detection include methods to make the detection invariant to the duration of the fault and to the starting point of the sample with respect to the fault inception. The algorithms compared are based on the analysis of the torque-producing component of the field-oriented stator currents in PMAC motors. The energy of the transform coefficients is thresholded to extract low-dimensional feature vectors. These vectors are then used in the fault-detection algorithms. These algorithms can be implemented in an online system, using extra processing time in the motor controller. Such an online implementation repeats the laboratory setup, with certain variations. The laboratory controller is implemented on a PC rather than on a DSP, but the sensors, measured quantities, control algorithm, cycle duration, and the control variables (e.g., i_q and i_d) are the same.

The test machine used in this paper is a six-pole surface-mounted PMAC machine for a 12-V automotive application, operated in field orientation. Its rated power and no-load speed are approximately 1 hp and 3000 r/min, respectively. A custom Xilinx FPGA-based I/O board was used as the drive controller. The experimental setup, shown in Fig. 1, is described in detail by Zanardelli *et al.* [6].

Three different types of faults were applied in the laboratory.

1) A short-circuit was imposed momentarily between a point in the coil and a terminal. This fault occurred with the machine operating under various load conditions. The imposed short-circuit was with different values of fault resistance. Two distinct events were recognized, the fault inception and the fault clearing. The advantage of two separate classes for each fault is invariance of the categorization to the duration of the fault. Additionally, a post-processing algorithm to verify subsequent classification of the beginning and end of the same fault type would further confirm the existence of the fault.
2) Similarly, a resistor was inserted in series with a terminal in the winding to simulate a momentary breaking of contact. Again, fault inception and fault clearing were identified as separate events, and the severity of the fault was primarily determined by the value of the inserted resistance. In both of these cases, the duration of disruption can be an additional measure of severity.
3) A mechanical fault was introduced, in the form of worn gears. This was realized by removing portions of gear teeth. It must be noted here that, in this situation, the fault manifests itself through a short high-frequency transient (corresponding to the time of contact of the damaged tooth) repeating at a lower frequency (defined by speed and gear ratio). Here, we study only the high-frequency part of the event.

V. Implementation

A. Fault Detection and Classification

The detection and classification of the faults described in Section IV are based on the analysis of the stator currents of the machine. Rather than analyzing the three phase currents independently of each other, the field-oriented currents i_{qs} and i_{ds} are used. Together, i_{qs} and i_{ds} are a complete representation of the stator currents; however, it has been determined experimentally that, through analysis of i_{qs} only, accurate fault detection and classification can be achieved. This can be explained by the fact that most faults manifest themselves in the torque of the motor, which is directly proportional to i_{qs}. The sampling rate for the current signal i_{qs} is set at 20 kHz.

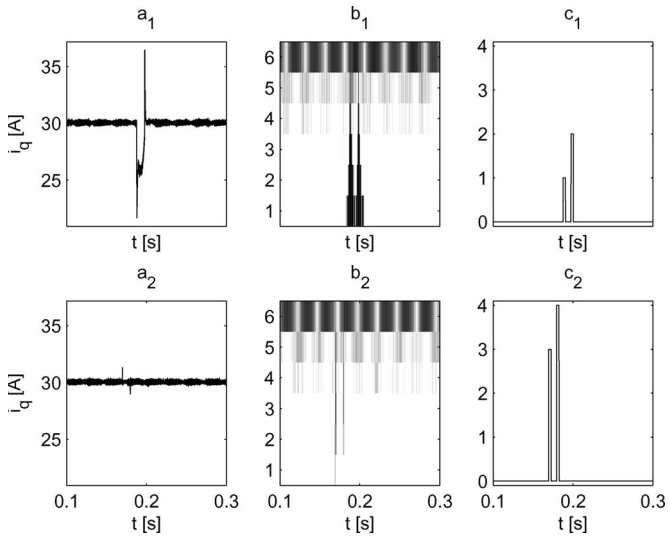


Fig. 2. Typical results for electrical faults based on analysis of the UDWT coefficients. a_1 shows the current for a momentary increase of contact resistance to 1.8 p.u., b_1 the corresponding UDWT coefficients, and c_1 the categorization of the inception and clearing of the faults. a_2, b_2, and c_2 show the same for a turn-to-turn short.

For the STFT, nfft $= 64$, noverlap $= 48$, and a 64-point rectangular window was used. The resultant STFT had 33 frequency bands. The two outermost bands corresponding to the dc and 10-kHz components of i_{qs} were discarded. The algorithm was based on the remaining 31 frequency bands. For WD and CWD, nfft $= 64$ is used, the negative frequency bands are discarded due to symmetry, resulting in 33 frequency bands. For the UDWT, the Daubechies D4 mother wavelet was used, and decomposition was performed for six levels.

The algorithm has two parts: a detection phase and a classification phase. The detection phase of the algorithm is based on thresholding the energy of the analysis coefficients. The classification phase is implemented when the criterion for detection is met.

To train the algorithms, thereby determining the weighting coefficients, data were used from eight electrical-fault experiments from each of the following operating conditions: healthy, increased contact resistance, decreasing turn-terminal resistance, and increasing number of teeth missing in the gears. The analysis coefficients begin at the index of the local maxima of the energy where the threshold was exceeded. The developed algorithm recognizes the fault inception and clearing as distinct events with different characteristics, while the duration of the fault plays no role in the recognition of the fault.

Following the training of the classification algorithm, data that had not been used in the training algorithm were tested. Two data sets from each of the following mechanical and electrical fault conditions were tested: healthy; 1, 1.5, 2, 2.5 missing teeth; increased contact resistance (2.14, 2.80, 4.03, 6.33, and 15.84 p.u.); and turn-terminal short.

B. Experimental Results

Results from two typical test cases for the electrical faults based on analysis of the coefficients resulting from UDWT are shown in Fig. 2. Similar results were obtained for Wigner, Choi–Williams, and STFT analysis.

TABLE I
ALGORITHM PERFORMANCE FOR ELECTRICAL FAULTS BASED ON LINEAR-DISCRIMINANT CATEGORIZATION

Test Description	STFT Total / Detected / Classified Correctly	UDWT Total / Detected / Classified Correctly	Wigner Total / Detected / Classified Correctly	C-W Total / Detected / Classified Correctly
Contact Resistance $(2.14pu)$ inception	2 / 2 / 2	2 / 2 / 2	2 / 2 / 0	2 / 2 / 0
Contact Resistance $(2.14pu)$ clearing	2 / 2 / 0	2 / 2 / 1	2 / 2 / 0	2 / 2 / 0
Contact Resistance $(2.80pu)$ inception	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2	2 / 2 / 0
Contact Resistance $(2.80pu)$ clearing	2 / 2 / 0	2 / 2 / 2	2 / 2 / 0	2 / 2 / 0
Contact Resistance $(4.03pu)$ inception	2 / 2 / 1	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2
Contact Resistance $(4.03pu)$ clearing	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2
Contact Resistance $(6.33pu)$ inception	2 / 2 / 1	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2
Contact Resistance $(6.33pu)$ clearing	2 / 2 / 1	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2
Contact Resistance $(15.84pu)$ inception	2 / 2 / 0	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2
Contact Resistance $(15.84pu)$ clearing	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2
Turn-Terminal Short inception	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2
Turn-Terminal Short clearing	2 / 2 / 2	2 / 2 / 2	2 / 1 / 0	2 / 2 / 2

The top row of Fig. 2 shows results from one of the increased-contact-resistance faults. The bottom row has results from the turn-terminal short. The left columns show the measured value of i_{qs} for each test case. The center columns show the corresponding coefficients. The right columns are the output of the detection and classification algorithm with $0 =$ healthy, $1 =$ beginning of increased-contact-resistance fault, $2 =$ end of increased-contact-resistance fault, $3 =$ beginning of turn-terminal short, and $4 =$ end of turn-terminal short. The results presented are based on new data not used in the training set. It can be seen that the amplitude of the analysis coefficients is increased at the points of inception and clearing of the faults. The increased energy at these points meets the criterion for detection. The results shown indicate correct classification of the inception and clearing of both faults explored in this paper. The performance of the algorithm for the different transform methods with the different classifiers can be observed in Tables I and II and are compared in Table III. From Table III,

TABLE II
ALGORITHM PERFORMANCE FOR ELECTRICAL FAULTS BASED ON k-MEANS CATEGORIZATION

Test Description	STFT Total / Detected / Classified Correctly	UDWT Total / Detected / Classified Correctly	Wigner Total / Detected / Classified Correctly	C-W Total / Detected / Classified Correctly
Contact Resistance $(2.14pu)$ inception	2 / 2 / 0	2 / 2 / 2	2 / 2 / 2	0 / 2 / 0
Contact Resistance $(2.14pu)$ clearing	2 / 2 / 0	2 / 2 / 0	2 / 2 / 0	2 / 2 / 0
Contact Resistance $(2.80pu)$ inception	2 / 2 / 0	2 / 2 / 2	2 / 2 / 2	1 / 2 / 2
Contact Resistance $(2.80pu)$ clearing	2 / 2 / 0	2 / 2 / 2	2 / 2 / 0	2 / 2 / 0
Contact Resistance $(4.03pu)$ inception	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2
Contact Resistance $(4.03pu)$ clearing	2 / 2 / 1	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2
Contact Resistance $(6.33pu)$ inception	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2
Contact Resistance $(6.33pu)$ clearing	2 / 2 / 1	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2
Contact Resistance $(15.84pu)$ inception	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2
Contact Resistance $(15.84pu)$ clearing	2 / 2 / 0	2 / 2 / 2	2 / 2 / 2	2 / 2 / 2
Turn-Terminal Short inception	2 / 2 / 2	2 / 2 / 2	2 / 2 / 1	2 / 2 / 2
Turn-Terminal Short clearing	2 / 2 / 2	2 / 2 / 2	2 / 1 / 0	2 / 2 / 2

it can be seen that UDWT performs the best as compared to the other transform methods and that the Fisher's discriminant coefficient is consistent with the classification results, i.e., the feature coefficients cluster better for transforms with higher classification accuracy. It is also seen that the classification results obtained with LDC and k-means are close to each other, with LDC giving higher accuracy in some cases. This is due to the fact that LDC uses the whole set of training coefficients to find the best linear discrimination planes between the different clusters while k-means only compares each new sample to the cluster mean.

VI. MDA

The analysis of the faults with different time–frequency transforms indicates that the time–frequency feature vectors contain redundant information and that not all regions of the time–frequency plane contain discriminative information. Therefore, we need data-reduction methods to choose the "optimal" lower dimensional transformation of the time–frequency feature vector. In this paper, we propose to use a linear transformation of the time–frequency feature matrix based on the MDA. MDA aims at finding the "best" linear transformation from the high-dimensional feature space to a low-dimensional space in terms of maximizing the discrimination between different classes.

MDA is an extension of Fisher's discriminant ratio that maximizes the ratio of intraclass scatter to interclass scatter. Given the time–frequency feature vectors $\mathbf{x}$ and the class labels, we compute the intraclass and interclass scatter matrices. Intraclass scatter matrix is defined as $\Sigma_w = \sum_{i=1}^{C} \sum_{\mathbf{x}\in C_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^{\mathrm{T}}$ and interclass scatter matrix is defined as $\Sigma_b = \sum_{i=1}^{C} K_i(\mathbf{m}_i - \bar{\mathbf{m}})(\mathbf{m}_i - \bar{\mathbf{m}})^{\mathrm{T}}$, where $\mathbf{m}_i$ is the class average, $\bar{\mathbf{m}}$ is the total average, and K_i is the number of samples in Class i. The optimal linear transformation Φ is defined such that the following ratio is maximized:

$$J(\Phi) = \frac{\Phi^{\mathrm{T}}\Sigma_b\Phi}{\Phi^{\mathrm{T}}\Sigma_w\Phi}. \tag{10}$$

The optimal Φ is found by solving the generalized eigenvalue problem, $\Sigma_b\Phi = \lambda\Sigma_w\Phi$. For C classes, the linear transform projects to a $C-1$-dimensional space.

A. Results With MDA

In this section, we compare the four time–frequency transforms using the MDA approach. The classification algorithm is trained using take-one-out approach. The optimal transformation of the time–frequency feature vector Φ is found based on this training set and is applied on the remaining sample. Each sample is assigned to the nearest cluster using k-means classification algorithm described in Section III. This procedure is repeated to account for all signal samples and to compute the classification accuracy. In this paper, the number of signal classes is equal to five, which means that Φ projects to a 4-D space.

Table IV reports the classification accuracy of the four algorithms using the projected feature vectors in the 4-D space. It can be seen that, although UDWT still performs the best, CWD performs almost equally well. We also note that projection to a low-dimensional space benefits Choi–Williams and Wigner distributions the most, indicating the higher degree of redundancy in these transforms. For these transforms, MDA reduces the redundancy and extracts more discriminant features. Since UDWT is already not as redundant as Choi–Williams and Wigner distributions, the benefit of using MDA in this case is minimal. Moreover, the results for Choi–Williams show a greater improvement as compared to Wigner distribution, since CWD does not suffer from cross terms. Figs. 3–6 show the clustering of the different fault classes in the 3-D space for the four transforms.

VII. COMPARISON OF ANALYSIS AND CATEGORIZATION METHODS

In this section, we offer a critical comparison between the different analysis and categorization methods in terms of fault detection and classification accuracy and computational complexity.

TABLE III
COMPARISON OF CATEGORIZATION METHODS APPLIED TO FOUR ANALYSIS METHODS. FOURTEEN FAULT EVENTS ARE ANALYZED, EACH CONSISTING OF A FAULT INCEPTION AND CLEARING, TO A TOTAL OF 28 EVENTS

Distribution	Detection		Fisher's Coefficient	Linear Discriminant Analysis					k-Means Analysis				
	Correct	Incorrect		Inception		Clearing		percent correct	Inception		Clearing		percent correct
				correct	incorrect	correct	incorrect		correct	incorrect	correct	incorrect	
UDWT	28	0	2.98	14	0	13	1	96	14	0	12	2	93
Choi-Williams	28	0	1.79	10	4	12	2	79	12	2	10	4	79
Wigner	27	1	1.6788/ 1.6798*	12	2	9	5	75	10	4	9	5	68
STFT	25	3	1.01/ 1.4*	10	4	8	6	64	10	4	8	6	64

* the Fisher coefficient for the first number used all events, categorized in their appropriate cluster; for the second number, incorrect categorizations were discarded.

TABLE IV
CLASSIFICATION RESULTS WITH MDA

Distributions	Classification Accuracy
UDWT	94.64 %
Choi-Williams	92.54 %
Wigner	80.35 %
STFT	58.92 %

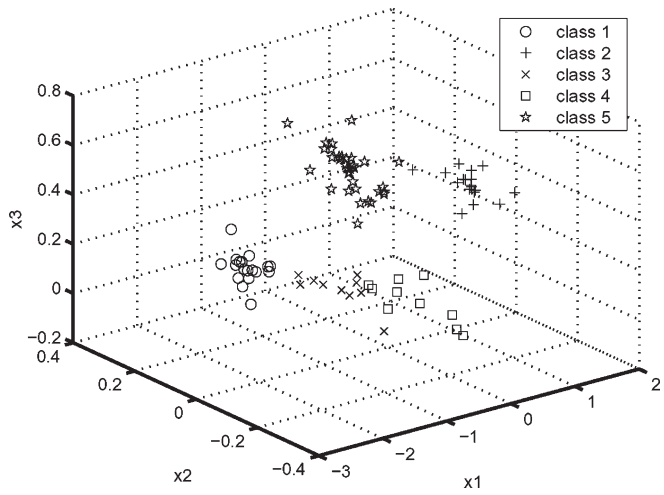


Fig. 3. Performance of different transforms with MDA–STFT.

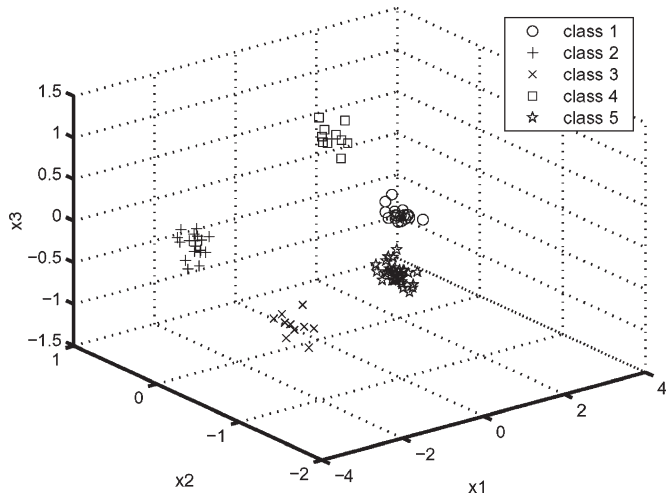


Fig. 4. Performance of different transforms with MDA–UDWT.

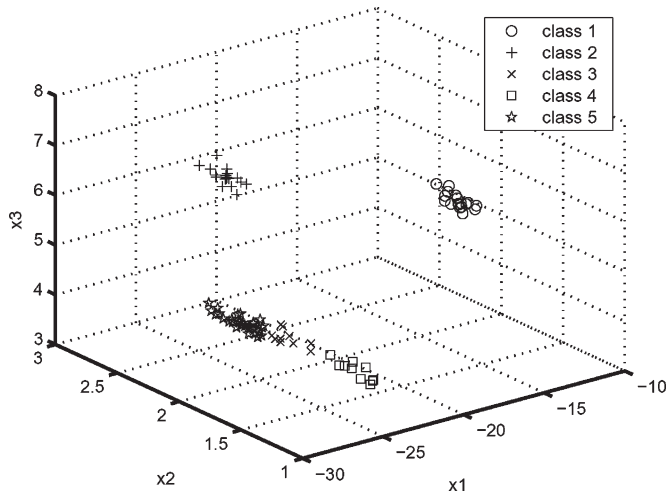


Fig. 5. Performance of different transforms with MDA–Wigner–Ville.

A. Analysis Methods

In terms of real-time implementation, the proposed four different time–frequency analysis methods have important differences. STFT and UDWT are linear transforms and thus require a lower number of computations, $O(N \log N)$ for STFT, and $O(LN)$ for UDWT, where L is the number of scales. WVD and CWD, on the other hand, are bilinear transforms of the signal and thus require more computations, $O(N^2 \log N)$. Depending on the application and the limitations of the system, linear-transform methods such as STFT and UDWT may be preferred over the bilinear ones. Table V shows the computational complexity of each analysis method.

In terms of the classification accuracy, UDWT performs the best for all classification methods and based on Fisher's discriminant ratio (Table III). This is due to the fact that wavelets are well suited in detecting transient phenomena in a signal such as the faults considered in this paper. Time–frequency distributions such as the Choi–Williams and Wigner distributions, on the other hand, have a lot of redundancy which reduces the classification accuracy. However, when a transform that reduces the amount of redundancy is applied, such as MDA, their performance gets close to UDWT. This indicates that both CWD and WVD contain discriminative information and that these

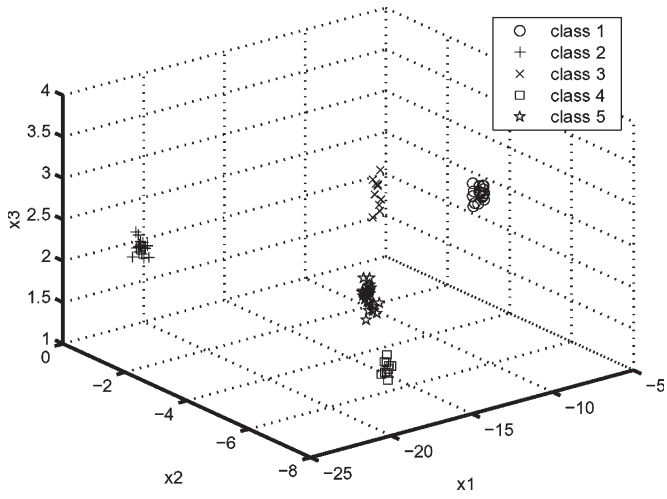


Fig. 6. Performance of different transforms with MDA–Choi–Williams.

TABLE V
COMPUTATIONAL COMPLEXITY OF THE UNDECIMATED WAVELET, STFT, WIGNER–VILLE, AND CHOI–WILLIAMS TRANSFORMS

Transform	Computational Complexity (Length N sequence)
UDWT	$O(LN)$, L is the number of scales
STFT	$O(N \log N)$
WVD	$O(N^2 \log N)$
CWD	$O(N^2 \log N)$

distributions are suitable for high-accuracy classification only after data reduction. CWD performs better than WVD owing to the reduction of the cross terms which may interfere with the identification of the different faults. STFT performs the worst among all of the tested distributions due to the time–frequency tradeoff and the dependence of the feature vector on the window function. It is possible to improve STFT results by optimizing the window function for the underlying signals.

B. Categorization Methods

In terms of the two categorization methods considered in this paper, LDC performs better than k-means. This is due to the fact that LDC uses the whole set of feature vectors from the training data to determine the discriminant surfaces, whereas k-means only uses the cluster-mean information.

In terms of computational complexity, k-means is easier to implement both in terms of training and the testing algorithms. On the training side, it requires the computation of cluster means, and on the testing side, it only requires the computation of a distance measure. LDC, on the other hand, requires more computational power for both learning the discriminant coefficients and for computing the online discriminant function during the testing stage.

VIII. CONCLUSION

In this paper, four different time–frequency analysis methods were compared for use in an algorithm for the detection and categorization of faults in electrical drives. The algorithm consists of analysis and categorization of the inception and clearing of intermittent faults. An exhaustive set of such conditions is necessary to develop a robust algorithm. Since the time–frequency feature vectors are highly redundant, we proposed an MDA-based method to reduce the time–frequency feature vectors to low-dimensional space prior to classification. It was shown that classification results with MDA in a 3-D space are comparable to classification results using the whole time–frequency feature vector with increased computational efficiency.

Two categorization algorithms are presented, trained using a set of operating conditions that include healthy drives and samples of faulted drives. The four candidate transform methods are applied to the same training and test data sets and categorization methods. The results are compared in terms of three measures: the number of correct and false categorizations, the confidence in their results as measured by Fisher's discriminant ratio, and the ease of implementation in hardware. The proposed methods are then discussed and evaluated. Finally, a framework for extending fault-diagnosis algorithms to failure prognosis is proposed, based on a Hidden Markov Model.

REFERENCES

[1] R. R. Schoen, T. G. Habetler, F. Kamran, and R. G. Bartheld, "Motor bearing damage detection using stator current monitoring," *IEEE Trans. Ind. Appl.*, vol. 31, no. 6, pp. 1274–1279, Nov./Dec. 1995.

[2] A. Lebaroud and G. Clerc, "Diagnosis of induction machine by time frequency representation and hidden Markov modelling," in *Proc. IEEE Int. SDEMPED*, Sep. 6–8, 2007, pp. 272–276.

[3] C. H. Lo, Y. K. Wong, and A. B. Rad, "Intelligent system for process supervision and fault diagnosis in dynamic physical systems," *IEEE Trans. Ind. Electron.*, vol. 53, no. 2, pp. 581–592, Apr. 2006.

[4] B. Akin, U. Orguner, H. A. Toliyat, and M. Rayner, "Low order PWM inverter harmonics contributions to the inverter-fed induction machine fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 55, no. 2, pp. 610–619, Feb. 2008.

[5] O. Ondel, E. Blanco, and G. Clerc, "Beyond the diagnosis: The forecast of state system application in an induction machine," in *Proc. IEEE Int. SDEMPED*, Sep. 6–8, 2007, pp. 491–496.

[6] W. G. Zanardelli, E. G. Strangas, and S. Aviyente, "Identification of intermittent electrical and mechanical faults in permanent-magnet AC drives based on time–frequency analysis," *IEEE Trans. Ind. Appl.*, vol. 43, no. 4, pp. 971–980, Jul./Aug. 2007.

[7] L. Cohen, *Time–Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[8] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms, A Primer*. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[9] P. Dutilleux, "An implementation of the "algorithme à trous" to compute the wavelet transform," in *Wavelets: Time–Frequency Methods and Phase Space, Proceedings of the International Conference*. New York: Springer-Verlag, 1987.

[10] T. Y. Young and T. W. Calvert, *Classification, Estimation, and Pattern Recognition*. New York: American Elsevier, 1974.

[11] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Hoboken, NJ: Wiley-Interscience, 2000.

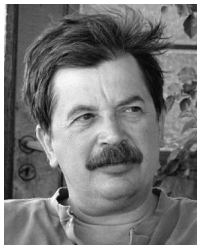
[12] H. S. Liu, B. Y. Lee, and Y. S. Tarng, "Monitoring of drill fracture from the current measurement of a three-phase induction motor," *Int. J. Mach. Tools Manuf.*, vol. 36, no. 6, pp. 729–738, Jun. 1996.

[13] A. Bellini, F. Filippetti, G. Franceschini, C. Tassoni, R. Passaglia, M. Saottini, G. Tontini, M. Giovannini, and A. Rossi, "On-field experience with online diagnosis of large induction motors cage failures using MCSA," *IEEE Trans. Ind. Appl.*, vol. 38, no. 4, pp. 1045–1053, Jul./Aug. 2002.

[14] X. Q. Liu, H. Y. Zhang, J. Liu, and J. Yang, "Fault detection and diagnosis of permanent-magnet DC motor based on parameter estimation and neural network," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 1021–1030, Oct. 2000.

[15] S. Pöyhönen, M. Negrea, P. Jover, A. Arkkio, and H. Hyötyniemi, "Numerical magnetic field analysis and signal processing for fault diagnostics of electrical machines," in *Proc. Int. Conf. Elect. Mach.*, Aug. 2002. paper 365.

[16] S. Rajagopalan, J. A. Restrepo, J. M. Aller, T. G. Habetler, and R. G. Harley, "Wigner–Ville distributions for detection of rotor faults in brushless DC (BLDC) motors operating under non-stationary conditions," in *Proc. IEEE Int. Symp. Diagnostics Elect. Mach., Power Electron. Drives*, Sep. 2005, pp. 313–319.



Elias G. Strangas (M'80) received the Dipl.Eng. degree in electrical engineering from the National Technical University of Greece, Athens, Greece, in 1975, and the Ph.D. degree from the University of Pittsburgh, Pittsburgh, PA, in 1980.

He was with Schneider Electric (ELVIM), Athens, from 1981 to 1983, and the University of Missouri, Rolla, from 1983 to 1986. Since 1986, he has been with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, where he heads the Machines and Drives Laboratory. His research interests include the design and control of electrical machines and drives, finite-element methods for electromagnetics, and fault prognosis and mitigation of electrical-drive systems.

Selin Aviyente (M'02) received the B.S. degree (with high honors) in electrical and electronics engineering from Bogaziçi University, Istanbul, Turkey, in 1997, and the M.S. and Ph.D. degrees in electrical engineering systems from the University of Michigan, Ann Arbor, in 1999 and 2002, respectively.

Since 2002, she has been an Assistant Professor with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing. Her research interests are in the area of statistical signal processing, in particular, nonstationary signal analysis.

Dr. Aviyente was the recipient of the 2005 Withrow Teaching Excellence Award and a 2008 NSF CAREER Award.

Syed Sajjad H. Zaidi (S'08) received the B.S. and M.S. degrees in electrical engineering from the National University of Sciences and Technology, Rawalpindi, Pakistan, in 1998 and 2006, respectively. He has been working toward the Ph.D. degree in the Machines and Drives Laboratory, Michigan State University, East Lansing, since 2006.

He was with Marine Systems Ltd. Pakistan from 2000 to 2002. His research interests include fault prognosis of electric machines and drives, time–frequency distribution, nonstationary signal processing, Bayesian estimation, and tracking.