Disturbance Classification Utilizing Dynamic Time Warping Classifier

A. M. Youssef, T. K. Abdel-Galil, *Student Member, IEEE*, E. F. El-Saadany, *Member, IEEE*, and M. M. A. Salama, *Fellow, IEEE*

***Abstract*—The application of deregulation policies in electric power systems results in the absolute necessity to quantify power quality. This fact highlights the need for a new classification strategy which is capable of tracking, detecting, and classifying power-quality events. In this paper, a new classification approach that is based on the dynamic time warping (DTW) algorithm is proposed. The new algorithm is supported by the vector quantization (VQ) and the fast match (FM) techniques to speed up the classification process. The Walsh transform (WT) and the fast Fourier transform (FFT) are adopted as feature extraction tools. The application of the combined fast match-dynamic time warping (FM-DTW) algorithms provides superior results in speed and accuracy compared to the traditional artificial neural networks and fuzzy logic classifiers. Moreover, the proposed classifier proves to have a very low sensitivity to noise levels.**

***Index Terms*—Dynamic time warping, pattern classification, power quality, vector quantization, Walsh transform.**

I. Introduction

RECENTLY, distribution and transmission systems have been witnessing a vast increase in the installed power-quality monitors. Manual inspection of the data that is collected from these systems involves high expenses and manpower. An automated classification of power quality events appears to be a viable solution to these problems. Acknowledging this fact, many researchers have proposed automated systems for power quality disturbance recognition. Most of the proposed systems use either the Fourier transform (FT) [1] or the wavelet transform [2], [3] for feature extraction and the artificial neural network (ANN) [4], [5] or fuzzy logic (FL) [6] for event classification. ANNs have attracted a great deal of attention because of their inherent pattern recognition capabilities and their ability to handle noisy data. However, ANNs have several drawbacks, including their inherent need of a large numbers of training cycles. The key benefit of FL is that its knowledge representation is explicit in utilizing simple "IF-THEN" relations. At the same time though, the use of FL is limited because power quality disturbances, especially transients and flickers, cannot be simply described by artificial explicit knowledge. The wavelet transform is a relatively new and powerful tool for analyzing power quality disturbances. The wavelet transform has the capability to extract information from the signal in both time and frequency domains simultaneously. Recently, the wavelet transform has been applied in the detection and classification of power quality events [7], [8]. However, the wavelet transform exhibits some disadvantages such as its computational burden, sensitivity to noise levels, and the dependency of its accuracy on the chosen basis wavelet.

A major concern in the classification process of power-quality disturbances, which may lead to a poor classification accuracy in the ANN and the FL methods, is the nonuniform time alignment between the test signal and the prestored templates. This arises from having several magnitudes, frequencies, and durations for each type of power quality disturbance. The time alignment issue can be most efficiently handled by applying the DTW algorithm , which has been extensively used in speech recognition [9]–[12]. The DTW is a template matching algorithm derived from dynamic programming. Conceptually, template matching is based on the comparison of the test signal against all of the stored templates in the dictionary. A measure of similarity is calculated, and then used to achieve a recognition decision.

In this paper, the most severe and common power-quality disturbances are simulated. In addition, simulations are performed for random changes in the magnitude, duration, and frequency of the generated signals to mimic a wide range of possible disturbance patterns. Both the FFT and the WT are used as feature extraction tools that collect important features from the power-quality disturbance signals, which are then provided to the classifier. Since the DTW algorithm is computationally demanding, both the VQ and the FM algorithms are introduced to speed up the DTW operation. Fig. 1 depicts a simplified block diagram for the proposed recognition system. The rest of this paper is organized as follows: first, the architecture of the proposed system is detailed in Section II; the FFT and the WT are presented in Section III; in Section IV, the new DTW algorithm is described; the utilization of the VQ and the FM algorithms for speeding up the DTW algorithm is presented in Section V; data generation and the results of this study are given in Section VI, and Section VII concludes the paper.

II. Architecture

The proposed power-quality automated recognition system is mainly divided into the following stages: segmentation, feature extraction, data reduction, and template matching.

A. Segmentation and Power-Quality Event Capturing

After the voltage and current signals are measured by potential and current transducers, they are converted to their

Manuscript received September 25, 2002.

A. M. Youssef is with the Combinatorics and Optimization Department, University of Waterloo, Waterloo, ON N2L 3G1 Canada.

T. K. Abdel-Galil, E. F. El-Saadany, and M. M. A. Salama are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1 Canada (e-mail: tkabdelg@hivolt.uwaterloo.ca).

Digital Object Identifier 10.1109/TPRWD.2003.820178

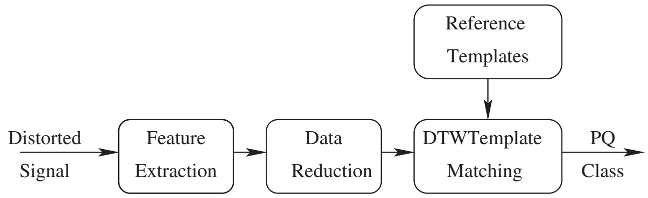

Fig. 1. Proposed automated power-quality recognition system.

digital counterpart by using a data acquisition board. It is highly desirable to extract the segments from the digitized voltage signal where the power-quality disturbance occurs. This can be accomplished by implementing a tracking and detecting algorithm to achieve a fast triggering of the classification schemes. For power-quality disturbances that are associated with sudden changes, a modified Kalman filter can be employed to achieve this goal [13]; for power-quality disturbances, which are characterized with relatively slow changes such as flickers, the Teager Energy Operator [14] can be utilized. It is worth mentioning that all of the utilized techniques for power-quality segmentation need to adjust a threshold value to detect the presence of power quality disturbances. However, this adjustment requires an expert in the field, and more research effort is needed to overcome this serious drawback. In this paper, the DTW is applied in the classification stage, and it is assumed that the clustering of the data, which contains the disturbance, has already occurred.

B. Feature Extraction

Both the FFT and the WT are used as tools to extract the most salient features representing the power-quality phenomenon. Before the application of these transforms, a preprocessing of the signals is required to normalize them since they are collected from various voltage levels in the distribution system.

C. Data Reduction

The amount of storage space and computation effort can be greatly reduced if the reference patterns and the input phenomena that are to be identified are presented as code sequences utilizing VQ instead of spectral parameters sequences [15]. In this paper, the VQ is also employed as a clustering technique to generate accurate, efficient, and reliable templates.

D. DTW Template Matching

While the design of the codebook and generation of the reference patterns are done offline, the classification of the power-quality events with the DTW algorithm is implemented online. Therefore, the DTW backed by the FM algorithm to speed up the classification process is introduced. Fig. 2 demonstrates the construction details of the proposed recognition system. The operation of each block will be discussed in more detail.

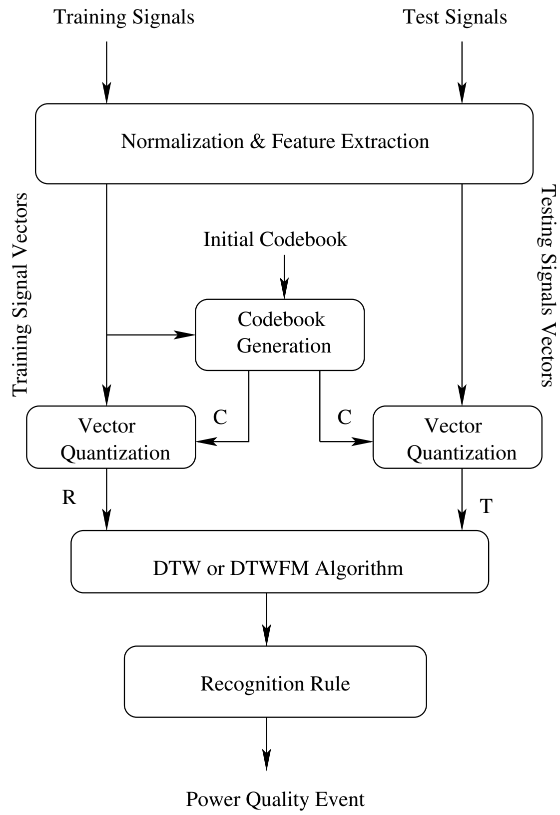

Fig. 2. Proposed system architecture.

III. FEATURE EXTRACTION

A. Discrete Fourier Transform (DFT)

The FT is utilized to convert time-domain waveforms into frequency components [16]. For a discrete time signal $x(n)$ of N points, the DFT is given by

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp\left(\frac{-j2\pi k}{N}\right), \quad 0 \le n \le N-1. \quad (1)$$

The DFT-based analysis/synthesis methods are very common in signal processing literature, primarily due to the existence of the FFT algorithm, which allows the computation of the DFT to be performed in $O(N\log(N))$ rather than $O(N^2)$ computations.

Different windowing types (such as the rectangular, Hanning, triangular, and Hamming windows) can be applied in the preprocessing step. However, our experiments indicate that the rectangular window with width equal to one power cycle and zero overlap between consecutive windows, results in a good recognition rate without added computational overhead.

B. Walsh Transform (WT)

Unlike the DFT, which uses a trigonometric basis for the calculation of its coefficients, the Walsh Transform (WT) [17] uses basis vectors constituted of only $\pm 1's$. For a discrete time signal, $x(n)$ of $N = 2^l$ points (l is an integer), the WT is given by

$$W(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)(-1)^{k.n}, \quad 0 \le n \le N-1 \tag{2}$$

where $k.n$ denotes the modulo 2 dot product of the binary representations of k and n. Unlike the FFT, which utilizes multiplication, the mathematical operations encountered in the calculation of the WT coefficients are simple addition and subtraction, which greatly reduces the computational cost. This is the motivation behind exploring the use of the WT for power-quality feature extraction.

IV. Dynamic Time Warping (DTW) Classifier

A. DTW Algorithm

Power-quality disturbances are time-dependent processes. Several occurrences of the same phenomenon are likely to have different durations, frequencies, and magnitudes. Furthermore, the same type of power-quality disturbance with identical duration period tends to vary its magnitude and frequency characteristics, due to the random nature associated with the disturbance. To obtain a global distance measure between two power-quality disturbance patterns (represented as a sequence of vectors), a time alignment must be performed. The challenge of the DTW alignment problem is to discover the optimal warping path, which is a curve relating the j time axis of the reference pattern, represented by $R = [R(1), R(2), \ldots, R(N_R$ to the i time axis of the test pattern represented by $T = [T(1), T(2), \ldots, T(N_T$, where N_T is the number of frames or vectors in the test signal and N_R is the number of frames or vectors in the reference signal. This warping path takes the form $W = \{w(1), w(2), \ldots, w(k), \ldots, w(K)\}$ where each w is a pair of pointers to the samples being matched (i.e., $w(k) = [i(k), j(k)]$). The warping function is required to minimize the overall cost function

$$D = \sum_{k=1}^{K} d[(w(k))] \tag{3}$$

where

$$d[(w(k)] = d(T(i(k)), R(j(k)))$$

is the distance between frame $i(k)$ of the test pattern, and frame $j(k)$ of the reference pattern. Many measures have been proposed in the literature to evaluate d. These measures include, but are not limited to, simple Euclidean distance, covariance weighting, and the LPC log-likelihood measure [10]. In this paper, the Euclidean distance measure is found to be sufficient for a good recognition accuracy. Fig. 3 illustrates the concept of the dynamic time warping path. The allowable area for the DTW curve is defined by slopes and path constraints, which are added as constraints to the optimization problem [11]. Different formulations of the DTW problem have been proposed [12]. However, only a simple DTW approach is presented in this paper. The idea behind most DTW algorithms is the realization that the solution of (3) is equivalent to finding the best path through a finite grid. Classical path-finding techniques are adopted to solve this problem. If it is noted that the best path from point $(1, 1)$ to any given point (i, j) is independent of what happens beyond that point, a simple recursive technique can be implemented to find the best path in the grid.

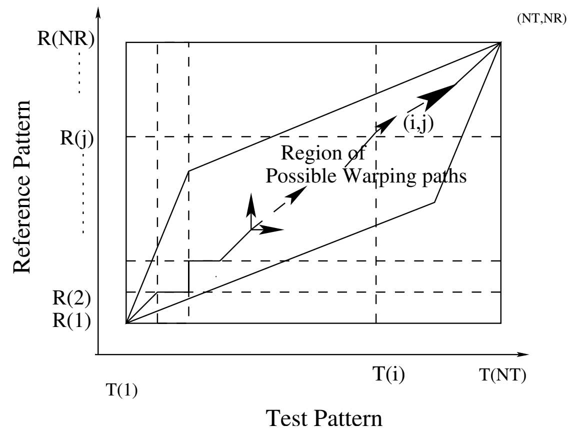

Fig. 3. Dynamic time warping path.

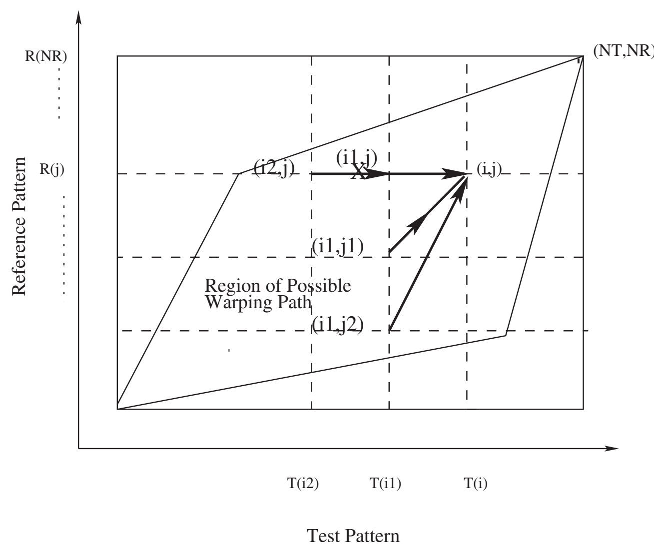

Fig. 4. Possible movements to the generic grid point (i, j).

Let $D_{\text{Acc}}(i, j)$ denote the minimum accumulated distance function from point $(1, 1)$ to point (i, j). A recursion equation for the accumulated distance can be written in the form

$$D_{\text{Acc}}(i, j) = d(T(i), R(j)) + \min_{(q_i, q_j)}[D_{\text{Acc}}(q_i, q_j)] \tag{4}$$

where $(q_i, q_j) \in$ the set of the grid points such that a path exists between (q_i, q_j) and (i, j).

Equation (4) shows that the minimum accumulated distance of the grid point (i, j) consists of the local distance d between the feature sets $T(i)$ and $R(j)$, in addition to the minimum accumulated distance to its legal predecessors.

Fig. 4 presents an example in which there are only three valid paths to the generic point $c(k) = (i, j)$. These paths originate from $(i-1, j-2)$, $(i-1, j-1)$, and $(i-1, j)$. Also, nonlinear constraints can be imposed in solving the recursion equation. For example, suppose that no path can originate from $(i-1, j)$

to the grid point (i,j), if the best path to the grid point $(i-1,j)$ comes from the grid point $(i-2,j)$ as shown in Fig. 4. Such local continuity constraint on the path can be expressed as

$$j(k)-j(k-1)=\begin{cases}0,1,2, & \text{if } j(k-1)\neq j(k-2)\\ 1,2, & \text{if } j(k-1)=j(k-2)\end{cases} \quad (5)$$

which yield the modified form of (4). See equation (6) at the bottom of the page, where $A=D_{\mathrm{Acc}}(i-1,j)$, $B=D_{\mathrm{Acc}}(i-1,j-1)$, and $C=D_{\mathrm{Acc}}(i-1,j-2)$.

In this study, we have imposed the following constraints:

- The function must be monotonic, that is
$$i(k)\geq i(k-1) \text{ and } j(k)\geq j(k-1).$$
- The function must match the end points of T and R, that is
$$i(1)=j(1)=1,\quad i(K)=N_T,\ \ j(K)=N_R.$$
- The function must not skip any points, that is
$$i(k)-i(k-1)\leq 1 \text{ and } j(k)-j(k-1)\leq 1.$$

The first and third constraints above imply that there are only three legal predecessors for $c(k)$. For example, if $c(k)=(i,j)$, then these points are $(i,j-1)$, $(i-1,j)$, and $(i-1,j-1)$. Hence, only three possibilities per point need to be considered. The final desired solution is given as

$$D^*=D_{\mathrm{Acc}}(N_T,N_R). \quad (7)$$

The optimum warping path is determined by backtracking from the "end of the path" back to the beginning [11], [12].

B. Decision Rule of Recognition

The last major step in the pattern-recognition model is the decision rule that chooses which reference pattern or patterns most closely match the unknown test pattern. Although a variety of approaches are applicable here, only two decision rules have been used in most practical systems, namely, the nearest neighbor rule (the NN rule) and the K-nearest neighbor rule (the KNN rule). In this study, the NN rule has been utilized. The NN rule operates as follows:

Given

- V reference patterns, R^i, $i=1,2,\ldots,V$;
- the average distance score D^i for each pattern from the DTW algorithm.

Then, the NN rule is simply given by

$$i^*=\arg\min_i[D^i]. \quad (8)$$

This means that the pattern R^{i*} with the smallest distance is chosen as the recognized pattern. In some applications, the explicit choice of i is not required; instead, an ordered (by distance) list of recognition candidates is used. In this case, the set of distances D^i is recorded to give a new set $D^{[i]}$ such that

$$D^{[1]}\leq D^{[2]}\ldots\leq D^{[V]}. \quad (9)$$

V. Speeding up the DTW Process

DTW is computationally demanding. Therefore, in order to speed up the DTW process, the following combination of techniques is adopted: working with quantized vectors, instead of continuous ones, and using a fast match procedure before performing the detailed DTW algorithm. The computational gains achieved by these two techniques are discussed below.

A. Vector Quantization (VQ)

The DTW computational cost can be significantly reduced if the computations required for evaluating the distance between the spectral feature vectors are replaced by a simple lookup table operation. However, in order to be able to construct such a lookup table, the continuous feature vectors must be transformed into discrete ones. Constructing the table can be performed offline, before the recognition process. To achieve this task, the VQ is required to map each continuous observation vector into a discrete codebook index. Once the codebook of these vectors has been obtained, the mapping between the continuous vectors and codebook indices becomes a simple nearest neighbor computation. More precisely, the VQ problem can be stated as follows:

Given

- the feature vector set: $\tau=[x_1,x_2,\ldots,x_N], x_i\in R^k$;
- the codebook consisting of M codeword vectors;
- the distance function $d(x,c_i)$ defined between feature vector $x\in R^k$ and codeword vector $c\in R^k$.

Find

- The particular codeword vector index i^*, which is associated with the codeword vector $c_i\in R^k$ that possesses the minimum distance from x is given as
$$i^*=\arg\min\ d(x,c_i). \quad (10)$$
Instead of dealing with the feature vector $x\in R^k$, the focus is on the index i^*, which means a large savings in calculation time. The major problem in the VQ algorithm is the design of a proper codebook for the quantization. This procedure divides the training vectors into M separated sets where M is the size of the codebook. Then, each such set is represented by a single vector c_i where $1\leq i\leq M$, which is generally the centroid of the vectors in the training set assigned to the ith region. Subsequently, the partition and the codebook (i.e., the centroid of each partition), are iteratively optimized. The design problem can be abbreviated as follows:

$$D_{\mathrm{Acc}}(i,j)=\begin{cases}d(T(i),R(j))+\min[A,B,C], & \text{if } j(k-1)\neq j(k-2)\\ d(T(i),R(j))+\min[B,C], & \text{if } j(k-1)=j(k-2)\end{cases} \quad (6)$$

Given

- the training set $\tau = [x_1, x_2, \ldots, x_N], x_i \in R^k$;
- the size of codebook M.

Find

- the codebook $C = [c_1, c_2, \ldots, c_M], c_i \in R^K$;
- the partition of space $P = [s_1, s_2, \ldots, s_M]$. Where s_m represents the encoding region associated with the codebook vector c_m, which minimizes the average distortion index

$$D_{\text{ave}} = \frac{1}{NK} \sum_{n=1}^{N} \|x_n - c_m\| \tag{11}$$

such that the following two conditions are fulfilled:

- First

$$s_n = \left\{ \|x_n - c_m\|^2 \leq \|x_n - c_{m'}\|.^2,\ m' = 1, 2, \ldots, M \right\}. \tag{12}$$

This condition implies that the encoding region s_m should consist of all the vectors that are closer to c_m than any of the other codebook vectors.

- Second

$$c_m = \left\{ \frac{\sum\limits_{x_n \in s_m} x_n}{\sum\limits_{x_n \in s_m} 1},\ m = 1, 2, \ldots, M \right\}. \tag{13}$$

This condition implies that the codebook vector c_m should be the average of all the training vectors that are in the encoding region.

In this study, the LBG-VQ technique is adopted to find the desired codebook. The details of this algorithm are discussed in [18]. Furthermore, the training sequence is obtained by recording the feature vectors that are extracted from several cycles of the different power-quality disturbances. Associated with the VQ is a distortion penalty since an entire region of vector space is represented by a single vector. Clearly, it is advantageous to keep the distortion penalty as small as possible. However, this implies a large codebook size, and consequently, larger storage requirements. Although the distortion steadily decreases as M increases, our experimental result shows that only small decreases in distortion occur beyond $M = 64$.

B. Fast Match

The idea behind the FM procedure is to count the number of occurrences of each codeword vector in the reference patterns of each phenomenon. These counts are then normalized by dividing them by the total number of vectors in each phenomenon. If the codebook size is M, then, for each phenomenon i, a normalized frequency of occurrence vector $(f_{i1}, f_{i2}, \ldots, f_{iM})$ is obtained during the training phase where f_{ij} denotes the normalized frequency of occurrences of the codeword vector with index j in phenomenon i. Any test pattern of length l can be represented by a sequence $t = \{t_1, t_2, \ldots, t_l\}$, corresponding to the indices of its quantized vectors. An ad-hoc measure for the probability that a test pattern is generated by certain phenomenon i is given by

$$P_t(i) = \prod_{k=1}^{l} f_{it_k} \tag{14}$$

where $P_t(i)$ is the probability that a test signal t belongs to class i, or more efficiently

$$\log(P_t(i)) = \sum_{k=1}^{l} \log(f_{it_k}). \tag{15}$$

Here, it is assumed that $\log(f_{i1}), \log(f_{i2}), \ldots, \log(f_{iM})$, for each phenomenon has been calculated and stored offline during the training phase. Only the reference pattern corresponding to the two phenomena with the highest fast match score is passed on to the detailed DTW matching process. Therefore, if p phenomena are present, then the computational cost is reduced by a factor of $[1 - 2/p]$. In this study, with six phenomena represented, a savings of 66.6% in computational time is achieved by applying the fast match algorithm. The implementation details of the FM algorithm as a tool for power-quality classification are introduced in [19].

VI. Application and Results

A. Signal Modeling

In order to obtain representative signals for the most common power-quality disturbances to serve the purpose of training, as well as the testing of the DTW classifier, power-quality disturbance signals are simulated using Matlab. Six categories of disturbances are simulated, namely, undisturbed sinusoid, sudden swell, sudden sag, harmonics, voltage flicker, and oscillatory transient.

The disturbances are based on ten cycles of voltage waveform. These waveforms are generated at a sampling rate of 256 samples/cycle for a total of 2560 points. This sampling rate can detect up to 7.6 kHz for power frequency equal to 60 Hz. However, in order to avoid aliasing, this sampling rate is used for the accurate detection of frequencies around 4 kHz. In order to create different disturbance cases, some unique parameters for each disturbance type have been introduced and allowed to change randomly. Parameters such as the starting time, magnitude, duration, frequency, and damping are implemented. This random generation of signals renders the testing of the DTW more reliable since none of these attributes is fixed for real distribution system disturbances. Table I provides a detailed summary of all the disturbance types along with the controlling parameters, definition, and equations. For a sufficient size and variety in the training set, 200 examples of each disturbance class are generated. A large variety of examples is needed to test the proposed algorithm. Thus, an additional 200 examples of each disturbance class, different from those used in training, are generated randomly for testing the DTW classifier.

B. Results

Substantial computer simulations are conducted to optimize the feature extraction algorithms and the clustering of the input

TABLE I
DISTURBANCE SIGNAL MODELING

Event	Symbol	Controlling Parameters	Equation
Pure Sinusoidal	M_1	N/A	$v(t) = \sin(\omega t)$
Sudden Sag	M_2	$0.1 \leq \alpha \leq 0.9, T \leq t_2 - t_1 \leq 9T$	$v(t) = A(1 - \alpha(u(t_2) - u(t_1)))\sin(\omega t)$
Sudden Swell	M_3	$0.1 \leq \alpha \leq 0.8, T \leq t_2 - t_1 \leq 9T$	$v(t) = A(1 + \alpha(u(t_2) - u(t_1)))\sin(\omega t)$
Harmonics	M_4	$0.1 \leq \alpha_3 \leq 0.9, 0.1 \leq \alpha_5 \leq 0.9,$ $0.1 \leq \alpha_7 \leq 0.9, \Sigma\alpha_i^2 = 1$	$v(t) = A(\alpha_1 \sin(\omega t) + \alpha_3 \sin(3\omega t) +$ $\alpha_5 \sin(5\omega t) + \alpha_7 \sin(7\omega t))$
Flicker	M_5	$0.1 \leq \alpha \leq 0.2, 0.1 \leq \beta \leq 0.5$	$v(t) = A(1 + \alpha\sin(\beta\omega t))\sin(\omega t)$
Oscillatory Transient	M_6	$0.1 \leq \alpha \leq 0.8, 0.5T \leq t_2 - t_1 \leq 3T$ $0.1ms \leq \tau \leq 0.2ms$	$v(t) = A[\sin(\omega t) + ...$ $\alpha \exp^{-(t-t_1)/\tau} \sin \omega_n(t - t_1)(u(t_2) - u(t_1))]$

TABLE II
FFT- AND WT-BASED DTW CLASSIFICATION RESULTS FOR CODEBOOK SIZE EQUAL TO 64

	Fast Fourier Transform						Walsh Transform					
	$M_1\%$	$M_2\%$	$M_3\%$	$M_4\%$	$M_5\%$	$M_6\%$	$M_1\%$	$M_2\%$	$M_3\%$	$M_4\%$	$M_5\%$	$M_6\%$
$M_1\%$	100	0	0	0	0	0	100	0	0	0	0	0
$M_2\%$	0	95	0	0	5	0	0	94	0	0	6	0
$M_3\%$	0	0	98	0	2	0	0	0	92	0	8	0
$M_4\%$	0	0	0	100	0	0	0	0	0	100	0	0
$M_5\%$	0	4	8	0	88	0	0	3	11	0	86	0
$M_6\%$	0	0	0	0	0	100	0	0	0	0	1	99
Accuracy	96.83%						95.17%					

TABLE III
FFT- AND WT-BASED FM-DTW CLASSIFICATION RESULTS FOR CODEBOOK SIZE EQUAL TO 64

	Fast Fourier Transform						Walsh Transform					
	$M_1\%$	$M_2\%$	$M_3\%$	$M_4\%$	$M_5\%$	$M_6\%$	$M_1\%$	$M_2\%$	$M_3\%$	$M_4\%$	$M_5\%$	$M_6\%$
$M_1\%$	100	0	0	0	0	0	100	0	0	0	0	0
$M_2\%$	0	97	0	0	3	0	0	94	0	0	6	0
$M_3\%$	0	0	98	0	2	0	0	0	92	0	8	0
$M_4\%$	0	0	0	100	0	0	0	0	0	100	0	0
$M_5\%$	0	4	7	0	89	0	0	3	10	0	87	0
$M_6\%$	0	0	0	0	0	100	0	0	0	0	1	99
Accuracy	97.3%						95.3%					

data by using the VQ algorithm. Different codebook sizes namely, 8, 16, 32, and 64 are adopted in testing the classifier. Classification accuracy is found to increase if the codebook size increases. Moreover, it is observed that saturation is achieved at the codebook size of 64. Further increase in the utilized codebook size does not result in classification improvement, but rather it may worsen it due to overfitting. The results obtained for the codebook size of 64 are reported since this size yields the best classification accuracy. The results for the six-class classification utilizing the DTW algorithm alone with codebook size 64 and two different extraction techniques are given in Table II. The results indicate high and comparable accuracy for both utilized extraction tools. Table III presents the results for codebook size of 64 for the WT and the FFT as feature extraction tools, with implementing the fast match algorithm in order to speed up the DTW process. Again, high classification accuracy is achieved. The main revenue from applying the combined FM-DTW is the reduction of computational effort without scarifying the classification accuracy. Although the computation speed and construction simplicity are among the main factors in choosing a successful classifier, performance accuracy remains the prime issue, especially in a noisy environment. Therefore, another study is conducted to investigate the effect of including random noise on the classified signals. A white noise varying between 1% and 3% is added to the generated signals to reflect the practical noise levels in the distribution systems. The results in Table IV show that the classification accuracy of the FM-DTW algorithm is not affected by a noisy environment. This is in contrast to the previous study where the classification rate dropped below 90% at a noise level $= 3\%$ when a simple matching technique combined with wavelet multiresolution analysis is utilized [20]. The proposed classifier is superior over other classification techniques and is immune against noisy data since the classification accuracy does not drop under 95% in all of the cases.

TABLE IV
CLASSIFICATION ACCURACY WITH VARIABLE NOISE LEVELS

Noise Level	FM-DTW Classifier Accuracy	
	FFT	WT
0%	97.3%	95.3%
1%	96.0%	96.1%
2%	96.1%	97.1%
3%	95.3%	96.8%

VII. CONCLUSION

The identification and classification of power-quality disturbances are a vital task for the precise monitoring of distribution systems. A new classification methodology that is based on the DTW algorithm is proposed in this paper. The computational effort and the storage requirements in the classification process are substantially decreased by combining the FM algorithm with the DTW. Further simplification is achieved by the use of the labels of the quantized vectors, instead of the feature vectors. Two different feature extraction techniques, namely the FFT and the WT, are used to extract the salient features of the studied signals. The degree of success of the new classifier varies, depending on the chosen codebook size, with success rates up to 97% for a codebook size that is equal to 64. The proposed algorithm outperforms other classification techniques and shows a very low sensitivity to data noise levels. The proposed algorithm shows great potential for the future development of fully automated monitoring systems with online classification capabilities.

REFERENCES

[1] G. T. Heydt, P. S. Fjeld, C. C. Liu, D. Pierce, L. Tu, and G. Hensley, "Applications of the windowed FFT to electric power quality assessment," *IEEE Trans. Power Delivery*, vol. 14, pp. 1411–1416, Oct. 1999.

[2] D. C. Robertson, O. I. Camps, J. S. Mayer, and W. B. Gish, "Wavelets and electromagnetic power system transient," *IEEE Trans. Power Delivery*, vol. 11, pp. 1050–1058, Apr. 1996.

[3] S. Santoso, W. M. Grady, E. J. Powers, J. Lamoree, and S. C. Bhatt, "Characterization of distribution power quality events with Fourier and wavelet transforms," *IEEE Trans. Power Delivery*, vol. 15, pp. 247–254, Jan. 2000.

[4] A. K. Ghosh and D. L. Lubkeman, "The classification of power system disturbance waveforms using a neural network approach," *IEEE Trans. Power Delivery*, vol. 10, pp. 109–115, Jan. 1995.

[5] S. Santoso, E. J. Powers, W. M. Grady, and A. C. Parsons, "Power quality disturbance waveform recognition using wavelet-based neural classifier. I.Theoretical foundation," *IEEE Trans. Power Delivery*, vol. 15, pp. 222–228, Jan. 2000.

[6] P. K. Dash, S. Mishra, M. M. A. Salama, and A. C. Liew, "Classification of power system disturbances using a fuzzy expert system and a Fourier linear combiner," *IEEE Trans. Power Delivery*, vol. 15, pp. 472–477, Apr. 2000.

[7] A. M. Gaouda, M. M. A. Salama, M. R. Sultan, and A. Y. Chikhani, "Power quality detection and classification using wavelet multi-resolution signal decomposition," *IEEE Trans. Power Delivery*, vol. 14, pp. 1469–1476, Oct. 1999.

[8] A. M. Gaouda, E. F. El-Saadany, M. M. A. Salama, V. K. Sood, and A. Y. Chikhani, "Monitoring hvdc systems using wavelet multi-resolution analysis," *IEEE Trans. Power Delivery*, vol. 16, pp. 662–670, Nov. 2001.

[9] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[10] J. R. Deller, J. H. L. Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals*. Piscataway, NJ: IEEE Press Classic Reissue, 2000.

[11] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," in *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 26, Feb. 1978, pp. 43–49.

[12] L. Rabiner and S. Levinson, "Isolated and connected word recognition—theory and selected applications," *IEEE Trans. Commun.*, vol. COM-29, pp. 621–659, May 1981.

[13] T. K. Abdel-Galil, E. F. El-Saadany, and M. M. A. Salama, "Power quality assessment in deregulated power systems," in *Proc. IEEE Power Eng. Soc. Winter Meeting*, New York, Jan. 2002, pp. 952–958.

[14] ——, "Energy operator for on-line tracking of voltage flicker levels," in *Proc. IEEE Power Eng. Soc. Winter Meeting*, New York, Jan. 2002, pp. 1153–1157.

[15] S. Furui, "A vq-based preprocessor using cepstral dynamic features for speaker-independent large vocabulary word recognition," in *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 36, July 1988, pp. 980–987.

[16] E. Brigham, *The Fast Fourier Transform*. Englewood Cliffs, NJ: Prentice-Hall, 1974.

[17] N. Ahmed and K. R. Rao, *Orthogonal Transforms for Digital Signal Processing*. New York: Springer-Verlag, 1975.

[18] A. Gersho, R. M. Gray, and R. Gallager, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer, 1991.

[19] T. K. Abdel-Galil, E. F. El-Saadany, A. M. Youssef, and M. M. A. Salama, "On-line disturbance recognition utilizing vector quantization based fast match," in IEEE Power Eng. Soc. Summer Meeting, Chicago, IL, July 21–25, 2002.

[20] A. M. Gaouda, S. H. Kanoun, and M. M. A. Salama, "On-line disturbance classification using nearest neighbor rule," *J. Electr. Power Syst. Res.*, vol. 57, pp. 1–8, Jan. 2001.

A. M. Youssef, photograph and biography not available at the time of publication.

T. K. Abdel-Galil (S'00), photograph and biography not available at the time of publication.

E. F. El-Saadany (M'01), photograph and biography not available at the time of publication.

M. M. A. Salama (F'02), photograph and biography not available at the time of publication.