An Efficient Feature Extraction Method for Classification of Power Quality Disturbances

H.Erişti, Y.Demir

ABSTRACT—This paper presents an efficient feature extraction method based on discrete wavelet transform for the classification of the power quality disturbances. Firstly, the decomposition coefficients are obtained by applying 10-level wavelet multi resolution analysis to the signals (normal, sag, swell, outage, harmonic, sag with harmonic and swell with harmonic) generated by using the parametric equations. Secondly, a combined feature vector is obtained from standard deviation of these features after distinctive features for each signal are extracted by applying the energy, the Shannon entropy and the log-energy entropy methods to decomposition coefficients. Finally, the support vector machine (SVM) classifier is used for classification performance of proposed feature extraction method. The regularization parameter and kernel parameter of the SVM are determined by 10-fold cross validation. Simulation results indicate that the combined feature vector has more high classification accuracy with regard to the other feature vectors.

Keywords: Disturbance classification, support vector machines, wavelet transform, features extraction.

I. INTRODUCTION

POWER quality (PQ) disturbances occur following events, such as line energizing, reactor and capacitor switching, faults, lightning, large load switching, etc., in the power systems. These disturbances are a serious problem for power system equipments and customers used especially sensitive electronic loads. If the sources, effects, causes and types of such disturbances are determined using a suitable monitoring system, an effectively solution can be performed for mitigation actions. In order to achieve this, monitoring equipments must have functions which involve detection, localization and classification of transient events and should be installed at suitable locations of the power system [1]. Monitoring PQ disturbances are carried out most often by the short time discrete Fourier transform (STFT). The STFT does not recognize the signal dynamics because of the limitation of a fixed window width. On the other hand, the wavelet transform (WT) having some advantages according to the STFT provides a better framework for power quality monitoring.

The WT technique is a very efficiently tool for detection and classification of short-time non-stationary signals thanks to its time-frequency multiresolution analysis (MRA) property. In [2], the WT-MRA technique to detect and localize different PQ disturbances is used. In [3], the continuous WT (CWT) to estimate the disturbance time duration and the discrete WT (DWT) to estimate the disturbance amplitude are proposed. In [4], the wavelet probabilistic network algorithm which is combined the properties of the WT and the probabilistic neural network (PNN) to detect disturbances is presented. There are several studies [5]-[10] where firstly, the WT is used for extracting distinctive features of PQ disturbances and then, classification of disturbances performs using the artificial intelligent techniques. As seen in these studies, the nearest neighbor pattern recognition technique [5], the fuzzy expert system [6], the artificial neural networks (ANNs) [7], the adaptive neuro-fuzzy inference system [8], the self-organizing mapping neural network [9], the self organizing learning array and the support vector machine (SVM) [10] have been used for classifying PQ disturbances.

In the classification system of disturbances, the feature extraction stage is very important to have high classification accuracy, reduce of the feature vector dimension and have less computing time at both the training and testing processes of classifier. In [11], the feature vector is created using rms values and total harmonic distortion values of disturbances. In [10], it is calculated the energy at each the WT decomposition level. In [12], a wavelet norm entropy-based feature extraction method is presented. In [13], the disturbance common features such as total harmonic distortion, number of peaks of the wavelet coefficients, energy of the wavelet coefficients, lower harmonic distortion, are gained using the Fourier transform and the WT. In [14], the relevant features are extracted from the S-transform that can be derived from the CWT choosing a specific mother wavelet and multiplying a phase correction factor.

This paper presents an efficient feature extraction approach for the classification of PQ disturbances using the SVM and the DWT. In this approach, distinctive features for each disturbance signal are extracted by WT-MRA and several feature extraction methods. Then, a combined feature vector is obtained from standard deviation of features belonging to these methods. Finally, classification results for each feature vector are obtained by using the SVM classifier.

The rest of this paper is organized as follows. The proposed algorithm is detailed in Section II. In Section III, the feature extraction stage of proposed disturbance recognition algorithm is presented. A short review to SVM classifiers and parameter selection are given in Section IV. Simulations and results are given in Section V and Section VI concludes this paper.

This work was supported by Firat University Scientific Research Unit. (FUBAP). The Project No: FUBAP-1605.

H.Erişti is with the Tunceli Vocational School, Tunceli University, Tunceli, Turkey (e-mail of corresponding author: heristi@firat.edu.tr).
Y.Demir is with the Department of Electrical Engineering, Firat University, Elazıg, Turkey (e-mail: ydemir@firat.edu.tr).

Paper submitted to the International Conference on Power Systems Transients (IPST2009) in Kyoto, Japan June 3-6, 2009

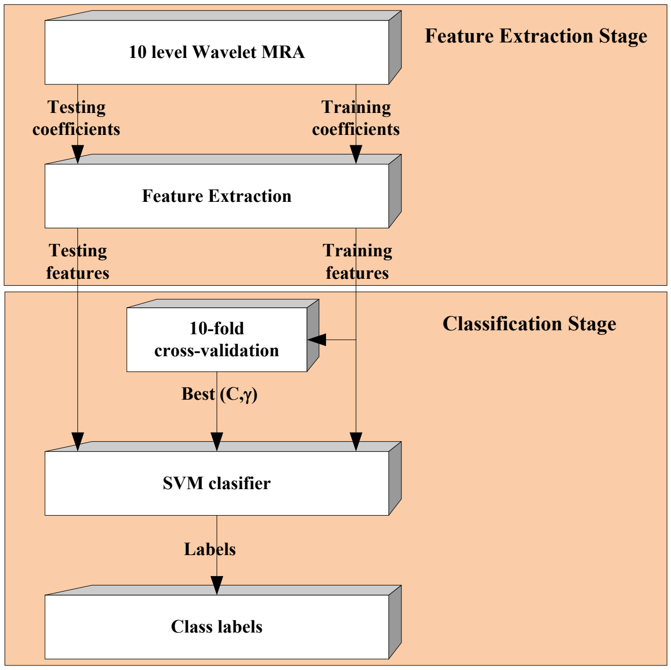

Fig. 1. The block diagram of proposed disturbance recognition algorithm.

II. Proposed Algorithm

The proposed algorithm involves two stages: feature extraction and classification (Fig. 1).

The first stage of disturbance recognition is to extract the distinctive features from disturbance signals. The feature extraction is carried out through the WT-MRA technique. The decomposition and approximation coefficients are obtained by applying 10-level MRA to the signals. Distinctive features for each both testing signal and training signal are extracted by applying the energy, the Shannon entropy and the log-energy entropy methods to detail coefficients belonging to each level and 10^{th} final level approximation coefficients. Additionally, a combined feature vector is obtained from standard deviation of features belonging to these three methods.

In the classification stage, disturbance types are determined by using the SVM classifier. The SVM classifier parameters are firstly selected by 10-fold cross-validation. By scanning the chosen parameter range, the parameters resulting in the lowest classification error are determined. Secondly, the SVM classifier is trained according to these parameters. Then, the feature vector obtained from the feature extraction stage is applied to the SVM input.

III. The Feature Extraction Stage

In the pattern classification problems, the dimensionality of the pattern representation at the network input is desired to keep small as possible to obtain higher classification accuracy and lower computational load and time [15]. Therefore the feature extraction process comes into prominence for a classification system.

A. The Wavelet Transform

The WT technique is a powerful tool to capture the time of transient occurrence and extract frequency features of disturbances.

The CWT of a continuous time signal *x(t)* is defined as,

$$CWT_{\psi}x(a,b)=\int_{-\infty}^{+\infty}x(t)\psi_{a,b}^{*}(t)\,dt, \qquad a,b\in R, \quad a\neq 0 \tag{1}$$

$$\psi_{a,b}^{*}(t)=\frac{1}{\sqrt{a}}\psi^{*}\left(\frac{t-b}{a}\right) \tag{2}$$

where *Ψ(t)* is the mother wavelet, and asterisk denotes a complex conjugate. *a* and *b* are scaling and translating parameters, respectively. The DWT is discrete counterpart of the CWT. In practical applications, the DWT of the sampled signal *x(k)* is replaced by the CWT of *x(t)* such that

$$DWT_{\psi}x(m,n)=\sum_{k}x(k)\psi_{m,n}^{*}(k) \tag{3}$$

$$\psi_{m,n}^{*}(k)=\frac{1}{\sqrt{a_0^m}}\psi^{*}\left(\frac{k-nb_0a_0^m}{a_0^m}\right) \tag{4}$$

m and *n* are scaling and sampling numbers, respectively. *m* indicates frequency localization and *n* indicates time localization. Generally, scaling and translating parameters can be chosen as $a_0 = 2$ and $b_0 = 1$. This choice provides a dyadic-orthonormal WT and the basis for the MRA.

The MRA decomposes the original signal into several other signals with different levels of resolution by means of high-pass filters (*HP*) and low-pass filters (*LP*) [16]. The approximation and detail coefficients are given as

$$c_j(n)=\sum_{k}LP(k-2n)c_{j-1}(k)\,, \tag{5}$$

$$d_j(n)=\sum_{k}HP(k-2n)c_{j-1}(k)\,. \tag{6}$$

Thus, the signal is mapped by the following set of coefficients

$$x_{signal}=[c_j,d_j,d_{j-1},\ldots,d_1]\,. \tag{7}$$

B. The Feature Extraction

The detail coefficients or approximation coefficients are not directly used as the classifier inputs. In order to reduce the feature dimension, the feature extraction methods are generally implemented to these coefficients at each decomposition level. In this study, the methods of the energy, the Shannon entropy and the log-energy entropy are used as the features extractors. All methods are individually applied to the detail coefficients of each level and the approximation coefficients at 10^{th} level and the features are extracted.

The energy at each decomposition level is calculated using following equations:

$$ED_i=\sum_{j=1}^{N}\left|d_{ij}\right|^2, \qquad i=1,\ldots,\ell \tag{8}$$

$$EC_{\ell}=\sum_{j=1}^{N}\left|c_{\ell j}\right|^2 \tag{9}$$

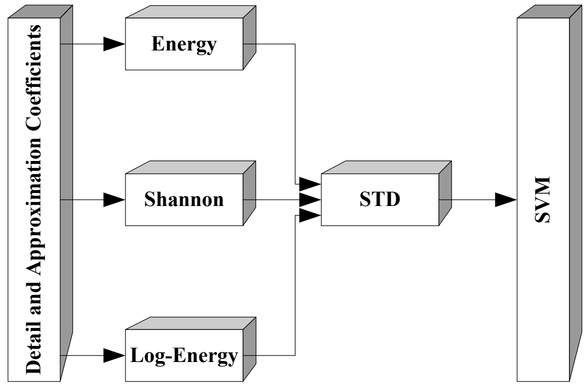

Fig. 2. The block diagram of the combined feature extraction method.

where i is the wavelet decomposition level from level 1 to level ℓ. N is number of the coefficients of detail or approximate at each decomposition level. In this way, for a ℓ level wavelet decomposition, a (ℓ+1) dimensional feature vector is constructed. For both each level detail coefficients and final level approximation coefficients, the Shannon entropy is given as

$$SD_i = -\sum_{j=1}^{N} d_{ij}^2 \log\left(d_{ij}^2\right), \qquad i = 1,\ldots,\ell \tag{10}$$

$$SC_\ell = -\sum_{j=1}^{N} c_{\ell j}^2 \log\left(c_{\ell j}^2\right). \tag{11}$$

The log-energy entropy is calculated using following equations.

$$LD_i = \sum_{j=1}^{N} \log\left(d_{ij}^2\right), \qquad i = 1,\ldots,\ell \tag{12}$$

$$LC_\ell = \sum_{j=1}^{N} \log\left(c_{\ell j}^2\right). \tag{13}$$

In this study, a combined feature vector which represents these three feature extraction methods is applied to classifier input as shown in Fig. 2. As seen in the simulation results, the combined feature vector has a high accuracy rate. The Fig. 3 illustrates the variations in the combined feature vector for six classes of disturbances used for training and analyzed with the Daubechie's 4 (db4) wavelet filter for 10-levels. These variations are obtained by subtracting normal signal from disturbance signals. By means of this figure, it can be said that the distinctive feature levels for sag, swell and harmonic are level 7 and 8, level 7 and 8, and level from 1 to 5, respectively.

IV. The Classification Stage

For automatic classification of PQ disturbances, the WT is integrated with the artificial intelligent methods or the expert systems. The ANNs have a great using area for classification of the disturbances thanks to their high noise tolerance, their inherent pattern recognition capabilities and their ability to recognize nonlinear functions. However, ANNs have several important disadvantages such as determining a proper architecture problem, local optimum problem, bad convergence property, over-fit or under-fit problem, et al. On the other hand, the SVM classifiers have been receiving a big interesting of power systems researchers because of producing single, optimum and automatic sparse solution by minimizing both generalization and training error and separating data by the large margin at high dimensional space [17, 18].

A. The Support Vector Machines

The SVM is a powerful tool for solving pattern classification problems [18, 19]. Given the training data $(x_1, y_1),\ldots,(x_\ell, y_\ell), \mathbf{x} \in R^M, y_i \in \{-1,+1\}$ for a two-class problem, the SVM constructs the decision functions of form $\mathrm{sgn}((\mathbf{w}^T \mathbf{x}_i) + w_0)$ by the maximum margin, where w is the normal vector of the separating hyperplane in the canonical form and w_o is a bias term [18]. The distances of the point closest to the hyperplanes of both -1 and +1 are calculated as $1/\|\mathbf{w}\|$. The separating margin is defined to be $2/\|\mathbf{w}\|$.

In many practical cases in which data is corrupted by noise,

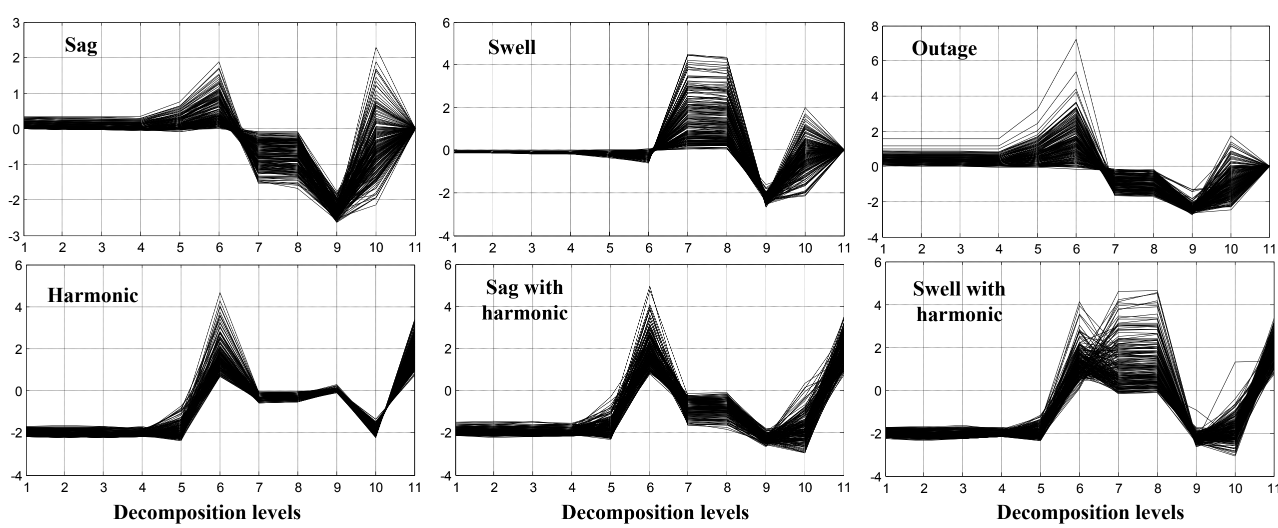

Fig. 3. Variations in the combined feature vector for different disturbance signals.

the data may not be separable by a linear hyperplane. To allow to deviations from margin, the slack variables $\xi_i \geq 0$ are introduced,

$$y_i((\mathbf{w}^T\mathbf{x}_i) + w_0) \geq 1 - \xi_i, \quad i = 1,\ldots,\ell . \tag{14}$$

For the training data x_i, if $0< \xi_i <1$, the data do not have the maximum margin but are still correctly classified. If $\xi_i \geq 1$, the data are misclassified by the optimal hyperplane. Thus the separation margin is increased by leaving intramargin the noise points occurring to near the boundaries or outlier points or both, so that generalization performance is improved.

By accordingly, the SVM constructs the constraint primal quadratic optimization problem that minimizes the training and generalization error by

$$\min_{\mathbf{w},\xi} \quad \Phi(\mathbf{w},\xi) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{\ell}\xi_i \tag{15}$$

$$s.t. \quad y_i\left(\mathbf{w}^T\mathbf{x}_i + w_0\right) \geq 1 - \xi_i\,,\ \xi_i \geq 0,\ \ i = 1,\ldots,\ell\,, \tag{16}$$

where C is the regularization parameter which controls the penalty incurred by each misclassified point in the training set. Generally, larger C values generate SVM models with smaller margin and better training accuracy as relatively smaller C values produce larger margin and better generalization accuracy.

To solve the primal problem in (15), Lagrange function is firstly formulated [20]. Then its derivatives with respect to the primal variables w, ξ and w_o are calculated and KKT conditions are satisfied. Finally the obtained dual optimization problem is solved,

$$\max_{\alpha} \ Q(\alpha) = -\frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T\mathbf{x}_j + \sum_{i=1}^{\ell}\alpha_i \tag{17}$$

$$s.t. \ \sum_{i=1}^{\ell}\alpha_i y_i = 0 \quad 0 \leq \alpha_i \leq C\,, \qquad i = 1,\ldots,\ell\,, \tag{18}$$

where α_i is nonnegative Lagrange multipliers.

The SVM maps the inputs x into some higher dimensional space by means of a nonlinear feature mapping $\varphi(x)$ for solving the classification problem separated by only highly complex decision boundaries in the input space. Thus the problem changes into linearly separable case at the feature space. If only scalar product $\mathbf{x}_i^T\mathbf{x}_j$ in (17) is replaced by the kernel function $K(\mathbf{x}_i,\mathbf{x}_j) = \varphi^T(\mathbf{x}_i)\varphi(\mathbf{x}_j)$ assumed to be symmetric and positive definite [18], the dual problem subject to constraints in (18) is rewritten as

$$Q(\alpha) = -\frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} y_i y_j \alpha_i \alpha_j K\left(\mathbf{x}_i,\mathbf{x}_j\right) + \sum_{i=1}^{\ell}\alpha_i \ . \tag{19}$$

$K(\mathbf{x},\mathbf{x}_i) = \mathbf{x}^T\mathbf{x}_i$ is a linear kernel. One of the most common kernels is Gaussian kernel defined as

$$K(\mathbf{x},\mathbf{x}_i) = \exp(-\|\mathbf{x}-\mathbf{x}_i\|^2 / 2\sigma^2) \tag{20}$$

where σ is the width parameter of Gaussian function.

The decision surface of the SVM is obtained by using only the training data x_i with $\alpha_i \neq 0$ lying closest to the decision boundary called as support vector,

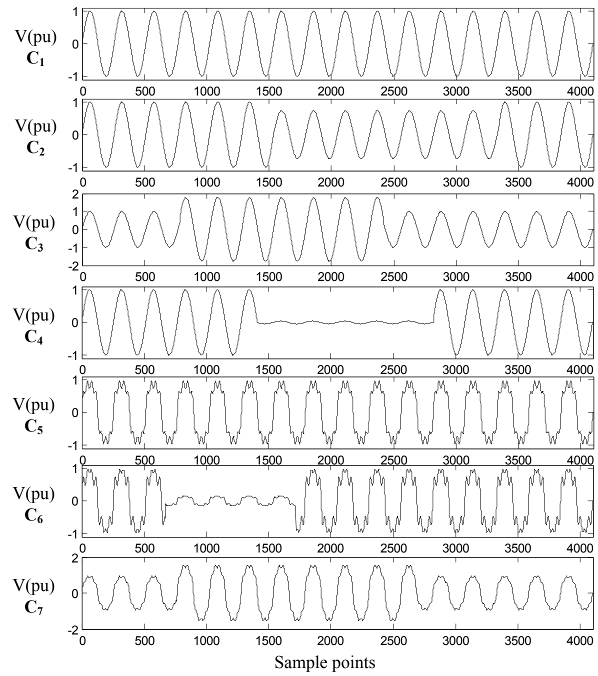

Fig. 4. Power system disturbances.

$$g(\mathbf{x}) = \mathbf{w}^T\varphi(\mathbf{x}_i) + w_0 = \sum_{\alpha_i \neq 0} y_i \alpha_i K(\mathbf{x},\mathbf{x}_i) + w_0\,, \tag{21}$$

where threshold can be obtained averaging over unbounded named support vectors with $0<\alpha_i<C$.

$$w_0 = \frac{1}{|U|}\sum_{i \in U} y_i - \mathbf{w}^T\varphi(\mathbf{x}_i)\,, \tag{22}$$

where U is the set of unbounded support vector indices.

B. The Parameter Selection

An important problem in the SVM training is to select the parameter C and the kernel parameters. This is known as model selection. Kernel parameters are referred to as hyper parameters. Choosing hyper parameters involves minimizing an estimate of generalization error or some related performance measures. In this paper, the parameter C and the kernel parameter are selected by using k-fold cross-validation. In k-fold cross validation, training data is randomly split into k mutually exclusive subsets or folds of approximately equal size. The SVM decision function is obtained using k-1 of the subsets and tested on the subset left out. This is repeated k times. Averaging over the k trials gives estimate of the expected generalization error.

V. Simulations and Results

A. Disturbance Signal Generation Using Parametric Equations

The recognition problem of the PQ disturbances was considered as the classification problem with seven classes consisting of normal signal and the disturbance signals called as sag, swell, outage, harmonics, sag with harmonic and swell with harmonic given in Table 1. The simulation data was

TABLE II

CLASSIFICATION RESULTS BASED ON ENERGY FEATURE VECTOR

True Class	C_1	C_2	C_3	C_4	C_5	C_6	C_7	Accuracy (%)
C_1	200	0	0	0	0	0	0	100
C_2	1	167	0	32	0	0	0	83.5
C_3	1	0	199	0	0	0	0	99.5
C_4	0	12	0	188	0	0	0	94
C_5	0	0	0	0	200	0	0	100
C_6	0	0	0	0	3	197	0	98.5
C_7	0	0	0	0	1	0	199	99.5
Overall success rate (%) : 96.42								

TABLE III

CLASSIFICATION RESULTS BASED ON THE SHANNON ENTROPY FEATURE VECTOR

True Class	C_1	C_2	C_3	C_4	C_5	C_6	C_7	Accuracy (%)
C_1	200	0	0	0	0	0	0	100
C_2	0	173	0	27	0	0	0	86.5
C_3	0	0	200	0	0	0	0	100
C_4	0	36	0	164	0	0	0	82
C_5	0	0	0	0	200	0	0	100
C_6	0	0	0	0	4	196	0	98
C_7	0	0	1	0	2	1	199	98
Overall success rate (%) : 94.93								

TABLE IV

CLASSIFICATION RESULTS BASED ON THE LOG-ENERGY ENTROPY FEATURE VECTOR

True Class	C_1	C_2	C_3	C_4	C_5	C_6	C_7	Accuracy (%)
C_1	200	0	0	0	0	0	0	100
C_2	0	176	1	23	0	0	0	88
C_3	0	4	196	0	0	0	0	98
C_4	0	4	0	196	0	0	0	98
C_5	0	0	0	0	200	0	0	100
C_6	0	0	0	0	0	194	6	97
C_7	0	0	0	0	0	2	198	99
Overall success rate (%) : 97.14								

TABLE V

CLASSIFICATION RESULTS BASED ON COMBINED FEATURE VECTOR

True Class	C_1	C_2	C_3	C_4	C_5	C_6	C_7	Accuracy (%)
C_1	200	0	0	0	0	0	0	100
C_2	0	195	0	5	0	0	0	97.5
C_3	0	0	200	0	0	0	0	100
C_4	0	1	0	199	0	0	0	99.5
C_5	0	0	0	0	200	0	0	100
C_6	0	0	0	0	0	200	0	100
C_7	0	0	0	0	0	0	200	100
Overall success rate (%) : 99.57								

generated using MATLAB and the parametric equations in [10, 17, 21]. The advantage of using parametric equations is that a better generalization performance can be obtained by collecting a few different signals belonging to same class.

Two hundred disturbance signals of each class were randomly generated for training and testing at interval of their control parameters. These signals were sampled at 256 points/cycle and generated for a total of 4096 points which contain the disturbances. This sampling rate can detect up to 6.4 kHz for power frequency equal to 50 Hz. A set of sample voltage waveforms given in Fig. 4 demonstrates the characteristics of the PQ disturbances.

B. Simulation Results

The LIBSVM which is an efficient tool for the SVM related training was used to evaluate the classification performance of extracted feature vectors for PQ disturbance signals [22]. The RBF kernel was chosen, because it can be like a linear kernel or a sigmoid kernel under different parameter settings. The values of parameter *C* and RBF kernel parameter were determined by using a 10-fold cross validation process for obtaining minimum classification error. The db4 wavelets were used as the wavelet function and disturbance signals were analyzed with a 10-level WT-MRA. All feature extraction methods, the energy, the Shannon entropy and the log-energy, were individually applied to the detail coefficients of each level and the approximation coefficients at 10^{th} level and the features were firstly extracted. Then, the obtained features by using each feature extractor were asunder scaled to be having the same mean and standard deviation.

Table II-IV gives the simulation result for seven-class PQ disturbance problem based on the energy feature vector, the Shannon entropy vector and the log-energy entropy feature vector, respectively. In these tables, the correct classification results are tabulated at diagonal elements. The misclassification results are tabulated at non-diagonal elements, respectively. As it can see here, average misclassification rates of sag and outage classes in these tables are 11.25, 15.75, and 7, respectively. Due to magnitude sag>10% and magnitude outage<10%, these rates are very bad with respect to rates of other classes. By means of these tables, it can be said that the log-energy technique has the best accuracy with respect to others.

Table V shows classification results for proposed approach. The misclassification problem of sag and outage classes is removed by this approach. Besides, other classes are classified as completely correct. These results indicate that the combined feature vector approach has more high classification accuracy with regard to the other feature vectors.

C. Discussion

Taking into consideration the given results in Table V, the accuracy of the proposed feature extraction method can be evaluated comparatively with the obtained results by using the WT and only energy technique in [10] and [21]. In both paper, the seven classes were generated by same parametric equations. In [10], the test accuracy was obtained as 94.93% by using SOLAR. In [21], the test accuracy was obtained as 90.4% by using the decision tree. On the other hand, in this paper, the test accuracy was obtained as 99.57%. It is clearly seen that the proposed feature extraction approach in this paper classifies effectively the power quality disturbances with different type. Also, test accuracy for the energy feature vector in this paper is 96.42%. This result shows that the used classification system based on the SVM and the DWT has the best classification performance with respect to [10] and [21].

VI. CONCLUSIONS

This paper considers an efficient feature extraction approach to classifying the PQ disturbances, relying on the SVM classifier and the DWT. In this approach, three feature vectors for each disturbance signal are firstly obtained by using different feature extraction methods and are examined for evaluating of classification performance. Then, a combined feature vector is obtained from standard deviation of features belonging to these methods. The experimental results show that the proposed combined feature vector has effectively a classification capability the PQ disturbances. Moreover, the proposed method can reduce the quantity of extracted features of disturbance signal without losing its property. Thus, the classifier system based on proposed feature extraction method needs less memory space and less computing time at both the training and testing processes.

VII. REFERENCES

[1] Z. L. Gaing, "Wavelet-based neural network for power disturbance recognition and classification," *IEEE Trans. Power Delivery*, vol. 19, no. 4, pp. 1560-1568, Oct. 2004.

[2] A. M. Gaouda, M. M. A. Salama, M. R. Sultan, and A. Y. Chikhani, "Power quality detection and classification using wavelet multiresolution signal decomposition," *IEEE Trans. Power Delivery*, vol. 14, no. 4, pp. 1469-1476, Oct. 1999.

[3] L. Angrisani, P. Daponte, M. D. Apuzzo, and T. Testa, "A measurement method based on the wavelet transform for power quality analysis," *IEEE Trans. Power Delivery*, vol. 13, no.4, pp. 990- 998, Oct. 1998.

[4] C. H. Lin and M. C. Tsao, "Power quality detection with classification enhancible wavelet-probabilistic network in a power system," *IEE Proc.-Gener. Transm. Distrib.*, vol. 152, no. 6, pp. 969-976, Nov. 2005.

[5] A. M. Gaouda, S.H. Kanoum, and M.M.A. Salama, "On-line disturbance classification using nearest neighbor rule," *Electric Power Systems Research*, vol.57, pp.1-8, 2001.

[6] Y. Liao, and J.B. Lee, "A Fuzzy-expert system for classifying power quality disturbances," *Electr.Power Energy and Syst.*, vol.26, no.3, pp.199-205, Mar. 2004.

[7] I. Monedero, C. Leon, J., Ropero, A. Garcia, J.M., Elena, and J.C. Montano, "Classification of electrical disturbances in real time using neural networks," *IEEE Trans. Power Delivery*, vol. 22, no.3, pp. 1288-1296, Jul. 2007.

[8] S. D. Kim, and M. M. Morcos, "An improved method for classifying power quality disturbances," *Electric Power Components and Systems,* vol.32, no.4, pp. 407-420, Apr. 2004.

[9] Y. Y. Hong, and C.W. Wang, "Switching detection/classification using discrete wavelet transform and self-organizing mapping network," *IEEE Trans. Power Delivery*, vol. 20, no. 2, pp. 1662-1668, Apr. 2005.

[10] H. He and J. A. Starzyk, "A Self-organizing learning array system for power quality classification based on wavelet transform," *IEEE Trans. Power Delivery*, vol. 21, no. 1, pp. 286-295, Jan. 2006.

[11] P. G. V. Axelberg, I. Y. Gu, and M. H. J. Bollen, "Support vector machine for classification of voltage disturbances," *IEEE Trans. Power Delivery*, vol. 22, no. 3, pp. 1297-1303, Jul. 2007.

[12] M. Uyar, S. Yildirim, and M. T. Gencoglu, "An effective wavelet-based feature extraction method for classification of power quality disturbance signals," *Electric Power Systems Research*, vol.78, pp.1747-1755, Apr. 2008.

[13] F. F. Zhu, G. S. Hu, and J. Xie, "Classification of power quality disturbances using wavelet and fuzzy support vector machines," Proceedings of the 4th International Conference on Machine Learning and Cybernetics, vol. 7, pp. 3981–3984, Aug. 2005.

[14] I. W. C. Lee, and P. K. Dash, "S-transform-based intelligent system for classification of power quality disturbance signals," *IEEE Trans. on Industrial Electronics*, vol. 50, no. 4, pp. 800–805, Aug. 2003.

[15] A. K. Jain, R. P. W. Duin, J. Mao, "Statistical pattern recognition: a review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4-37, Jan. 2000.

[16] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.* vol. 11, no. 7, pp. 674–693, July 1989.

[17] P. Janik, and T. Lobos, "Automated classification of power quality disturbances using SVM and RBF Networks," *IEEE Trans. Power Deliv.*, vol. 21, no. 3, pp. 1663-1669, July 2006.

[18] V. Vapnik, *Statistical Learning Theory*, New York: Wiley, 1998.

[19] C. Cortes, and V. N. Vapnik, Support vector networks, *Mach.Learn.,* vol. 20, pp. 273–297, Sep. 1995.

[20] D. P. Bertsekas, *Nonlinear Programming*, second ed., Athena Scientific, 1999.

[21] T. K. Abdel-Galil, M. Kamel, A. M. Youssef, E.F. El-Saadany, and M. M. A. Salama, "Power quality disturbance classification using inductive inference approach," *IEEE Trans. Power Delivery*, vol. 19, no. 4, pp. 1812-1818, Oct. 2004.

[22] C. C. Chang, and C. J. Lin, "LIBSVM: a library for support vector machines," Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

VIII. BIOGRAPHIES

Hüseyin Erişti was born in Mersin, Turkey in 1979. He received the B.Sc. and M.Sc. degrees from Fırat University in Elazığ, Turkey, in 2000 and 2005, respectively. He is currently working towards his PhD degree. His research interests are in the areas of power quality, signal processing, artificial intelligent techniques. He has been working as lecturer at Tunceli University since 2000.

Yakup Demir was born in Mersin, Turkey in 1964. He received the B.Sc., M.Sc. and Ph.D. degrees from Fırat University, Elazığ, Turkey, in 1985, 1988, and 1993, respectively. He is currently a professor at the Faculty of Electrical and Electronic Engineering, Firat University. His research interest areas are power quality, artificial intelligent techniques, piecewice-linear systems and switched systems.