Automatic Appliance Classification for Non-Intrusive Load Monitoring

Po-An Chou, Chi-Cheng Chuang and Ray-I Chang

***Abstract*-- This paper is based on non-intrusive load monitoring (NILM), which uses low-frequency sensor in power circuit. Traditional process must establish a database with features before identifying what the circuit is. If the system wants to add new feature of appliances into database, it must relearn electrical data. Therefore, this paper proposes a method, which can identify appliances status and whether new appliances exist or not. It can also learn feature of appliances automatically at the same time. The proposed method combines statistics with classification techniques to simplify the feature extraction. The consequent is quite valid in the economy, accuracy and feasibility. In addition, if NILM system does not identify successfully, it might contain the unknown appliances. The unknown appliances can thus be identified. The system will be able to expand its appliances amount in the database automatically. Experiment performed with a variety of single or multiple classifications which include the unknown appliances.**

***Index Terms*— Multiple signal classification, Data mining, Feature Extraction, NILM**

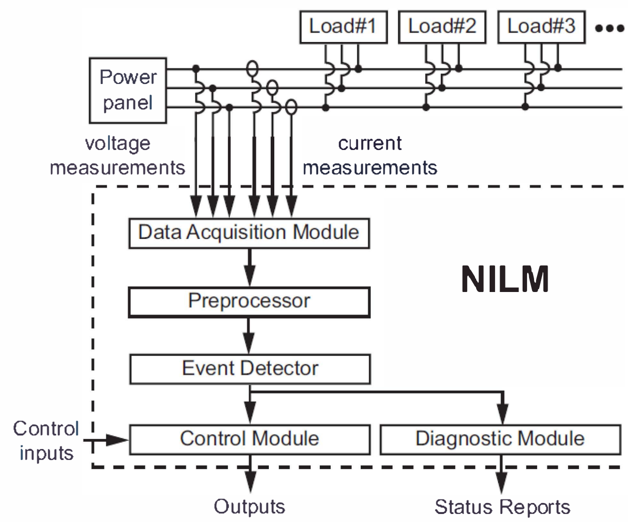

Fig. 1. NILM block diagram [9]

I. Introduction

With the gradual depletion of resources, energy is the one of the most important issue for the world. Related research indicates that if the household is able to master appliances usage and energy consumption in the house that will achieve a warning effect and reduce the energy consumption about 5% to 15% [1]. However, normal resident does not have sufficient expertise and time to keep detailed records of all kinds of electrical parameters and status information. Load monitoring system identifies the object and its status by a various power parameters, which provides more detail electricity information as well as intelligent application.

In 1980s, non-intrusive load monitoring (NILM) or non-intrusive appliance load monitoring was developed [2] as a low cost alternative to attaching individual monitors on each appliance. It can analyze what appliances are used in the house as well as their individual energy consumption. The block diagram of NILM is shown in Fig. 1.

According to the frequency of record, it can be divided into high-resolution and low-resolution. Generally speaking, the frequency of the electricity parameters is implemented at 120Hz or more in high-resolution. Because the high frequency sampling rate is high, it can reflect the electrical waveform of transient change with higher accuracy [3]. Because high frequency sensor is expensive and produce mass data records with more computing resource. Therefore, some researches support relatively slow sampling time. Although slightly decrease in recognition accuracy, greatly reduces the cost for low-resolution identification method. Such as Mario [4] who made real power and reactive as parameters to identify appliance state with four different recognition algorithms, including KNN (K-Nearest Neighbor). Although the recognition accuracy is not quite perfect, but practical.

There are two kinds of identification methods, transient and steady-state identification. Transient identification method is used when the electricity signal become changing, it has to detect the event that occur load differences. That way the system must to monitor the situation of the circuit in time. And load monitor system will compare signal with one in database which is learned [5]. But if it exist unstable appliances or noise, the method may lose its accuracy [6]. Steady-state identification method collects data in fixed-interval, so the size of the sampling is particular important.

No matter what it uses, it has some drawback. It must to execute learning process a period of time, not completely non-invasive. Thus, NILM reveal several degrees of Non-Intrusive,

Po-An Chou is with Department of Engineering Science and Ocean Engineering, National Taiwan University, Taipei, Taiwan, ROC (e-mail: r00525049@ntu.edu.tw).

Chi-Cheng Chuang is with Department of Engineering Science and Ocean Engineering, National Taiwan University, Taipei, Taiwan, ROC (e-mail: polon.chuang@gmail.com).

Ray-I Chang is with Department of Engineering Science and Ocean Engineering, National Taiwan University, Taipei, Taiwan, ROC (e-mail: rayichang@ntu.edu.tw).

978-1-4673-2868-5/12/$31.00 ©2012 IEEE

they can be classified in two classes [7]: MS-NILM (Manual Setup) and AS-NILM (Automatic Setup). MS-NILM is more precise than AS-NILM, but is not easy to setup meter for a large-scale.

Electrical signal into the system may contain one or more of electrical signals, the signal analysis and classification has become an important issue. Ethan R.Proper has completed the related research for automatic appliances classification [8]. Since the unknown appliances may not be including in the database. This paper proposes a new method that integrates low-resolution with steady-state identification, the method is able to detect unknown appliances and automatically learn feature of appliances. With individual electrical data of appliances, it can combine different appliances. Reduce the time and process in the learning process. Method belongs to the AS-NILM, including a completely non-invasive and which not requires training samples. In other words, the database of identification is capable of infinitely expand its amount of appliances.

II. Non-Intrusive Load Monitoring

In essence, the status identification is a core section of non-intrusive load monitoring, which consists of three steps in Fig. 2: Feature extraction, event detection, and status identification.

The feature extraction step deals with the mapping from raw electrical data such as real power, reactive power, harmonics and power factor, etc. Changes are then detected and flagged as events in these extracted features. Finally, the events are to be classified by the status identification algorithm as belonging to one appliance or another.

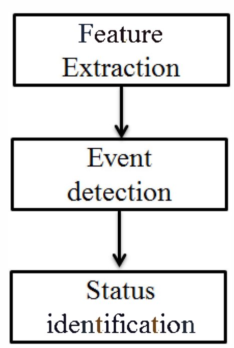

Fig. 2. The step of identification

However, the system usually establishes a database for identification and is divided two parts in Fig. 3: supervised and unsupervised.

Supervised is the machine learning task of inferring a function from supervised training data. The common method is SVM and K-NN [10]. The electrical data is classified and stored, and the events can be designated. Some research use the artificial intelligence methods like BP-ANN [11], and a mass data ensure identification accuracy. But the disadvantage is difficult to achieve completely non-invasive and if a new appliance need to be identified. It has to retrain all data, including old appliances. Unsupervised learning refers to the problem of trying to find hidden structure in unlabeled data. This distinguishes unsupervised learning from supervised learning. Due to lack of training data, the basic method is to establish all data into binary tree. But that way, the database is established for a large. Thus, probability and statistics method have been proposed. It does data statistical processing to retain valid data, can effectively reduce the data storage and computing. In any case, appliances which are not been trained exist in the circuit are difficult to find and learn to database automatically.

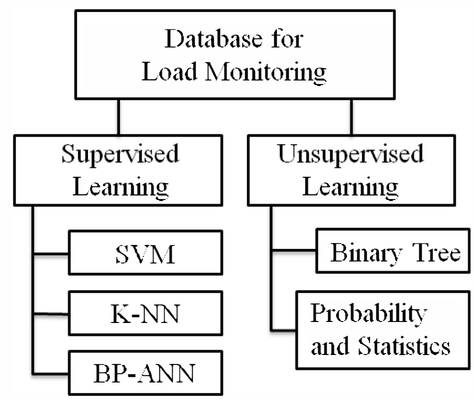

Fig. 3. The different method to establish database

III. Appliance classification and Mathematic model

In order to achieve completely non-invasive, the system establishes a database and extract data classification automatically is necessary. It must quite reduce the storage space of appliance classification, and the simulated appliance classification is method we adopted. Identification method is based on the probability model and the confidence interval, providing a flexible and effective interval.

A. Simulated Classification

In the identification method, all appliance combination we called "classification". To identify whole possibility classification, it must obtain whole classifications by learning in the past. In the paper, the appliance state is divided into two options, "on" and "off". So the appliances quantities and the classification of total number are relation in series. However, increasing appliances quantities will cause trouble on learning, and similarly its classifications will be several times.

In [13], he confirmed the several signals of appliances can be superimposed with high resolution when alone appliance is running. The low resolution is the accumulation of high-resolution. By simulating with every alone appliance, the system can obtain whole possibility classifications and it can reduce the number and cost of learning. When the appliance is running, resistive load will consume real power, and the reactive power is generated by the inductive and capacitive load. The system uses real power as the main power information. There are N kinds in circuit, which will be N^2 different classifications. It can be expressed as follows.

$$C_i = \{c_{i,1}, c_{i,2}, \ldots, c_{i,N}\} \tag{1}$$

Where $c_{i,1} \sim c_{i,N}$ stands for the states of the appliance 1 to the appliance N. The value is 1 represents the appliance is open or in use and 0 represents offline or turn off. The simulated classifications can be expressed as follows.

$$P_{C_i} = \sum_{j=1}^{N} (c_{i,j} \times P_j) \tag{2}$$

Using these simulated data to the expansion of the database classification, the data of classification is decreasing from N^2 to N in Fig. 4.

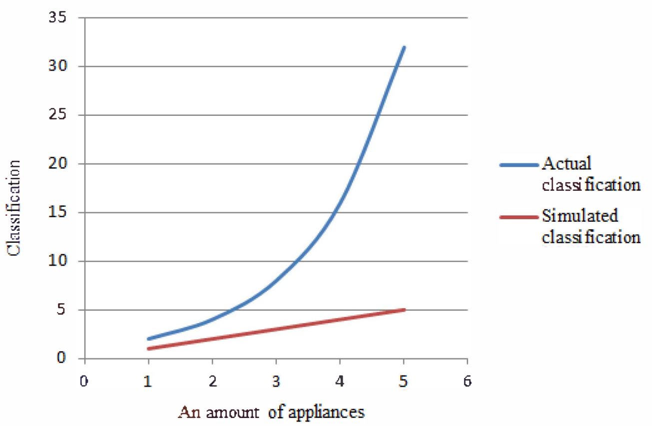

Fig. 4. Requirement for identification database

According to the electrical consumption characteristics, appliance type is divided into two categories: The first is relatively fixed, generally by a single frequency of the motor, thermal resistance, or bulb. The second is electrical consumption changes greatly, including desktop and notebook computers.

B. Preprocess

The system implements low-frequency sampling. Once the signal is diagnosed with no significant change, then the system carried out data acquisition. To make sufficient information to represent the electrical characteristics, which using a large sample about one minute in 1Hz (Using data in Fig. 5 as an example). After the end of the sampling time, it immediately calculates the relevant statistical information in mean, variance. When the system carried out data acquisition every time, can be said for statistical sampling. Therefore the formula is expressed as follows.

$$\bar{X} = \frac{X_1 + \ldots + X_n}{n} \;,\; S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} \tag{3}$$

$X_1 + \ldots + X_n$ stands for sampling data, n stands for sampling size, $\bar{X}$ stands for sample mean and S^2 stands for sample variance.

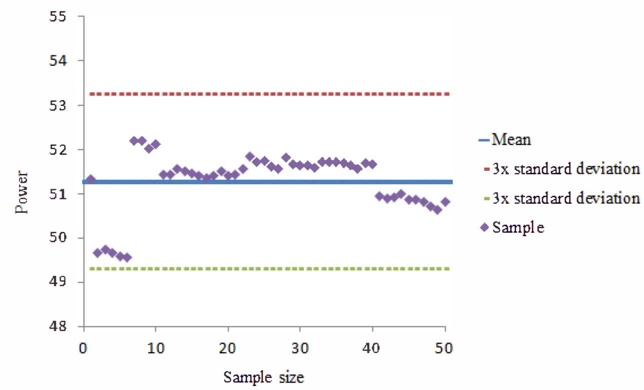

Fig. 5. The sample distribution of LCD1

C. Probability model

In the identification with steady-state, according to central limit theorem (CLT). No matter what population distribution is, sample approximate normal distribution in large sample. An experiment shows in Fig. 6. There are some advantages: the appliances for relatively fixed can narrow the identification interval and for consumption changes greatly also can fuzzy identification interval; a better superimposed computing.

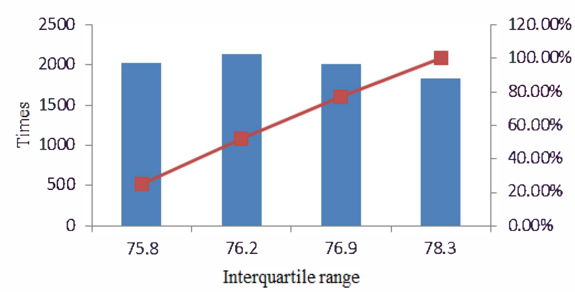

Fig. 6. Histogram for 8,000 times of LCD1 and LCD2

The distribution of a sum of number normally distributed independent variates X_n and with means and variances (μ_i, σ_i^2), respectively is another normal distribution Y.

$$X_i \sim \left(\mu_i, \sigma_i^2\right) \;,\; i = 1, 2, \ldots n$$
$$if \;\; Y = \sum_{i=1}^{n} X_i \;,\; Y \sim \left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right) \tag{4}$$

With this approach, it can establish a confidence interval (CI). In statistics, a particular kind of interval estimate of a population parameter and is used to indicate the reliability of an estimate.

The case is unknown population variance in large sample. When the population variance is unknown, the sample variance is as the estimator of the population variance, and it is a t-distribution.

$$\left(\bar{X} - t_{\frac{\alpha}{2}(n-1)} \frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}(n-1)} \frac{S}{\sqrt{n}} \right) \tag{5}$$

Although, when the great freedom (greater than 30), distribution will be close to Z-distribution, then Z value instead of value.

$$\left(\bar{X} - Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + Z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right) \tag{6}$$

t stands for parameter of t-distribution probabilities and Z stands for parameter of Z-distribution probabilities, $1-\alpha$ is confidence level.

D. Identification

By confidence interval, the calculated mean which fall into what the interval of classifications in Fig. 7, and get it. If not, it maybe exist unknown appliances in circuit and the system records its power information to unknown storage space, called unknown pool. Later if it captured similar power information in pool, it would analyze by each other.

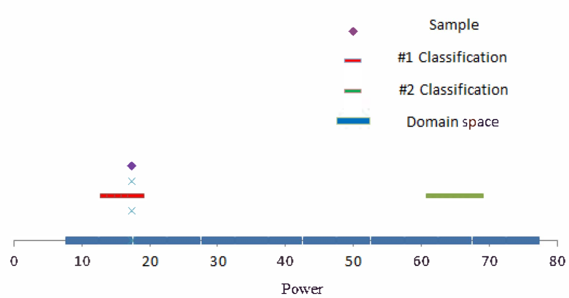

Fig. 7. Identification by different confidence interval

Compared to the power information whether approximate normal or not, ours pay more attention to confidence interval, and use it to control the identification interval.

IV. Process of Algorithm

The algorithm is divided into two parts: Identification and Detect unknown. Fig. 8 is the show chart of main processing. It has works as follows: When the system starts, it detects the signal and sampling immediately, and preprocesses the data and calculates its electrical simulated combination. Then, identify by system, including an appliance and combinations. If it happens to fails, it will store the electrical sampling into unknown pool. In unknown pool, multiple data sampling is identified by each other. If it appear two or than more which has same electrical features and same characteristics. The system determines it has new signal into circuit, and stores the data into identical database.

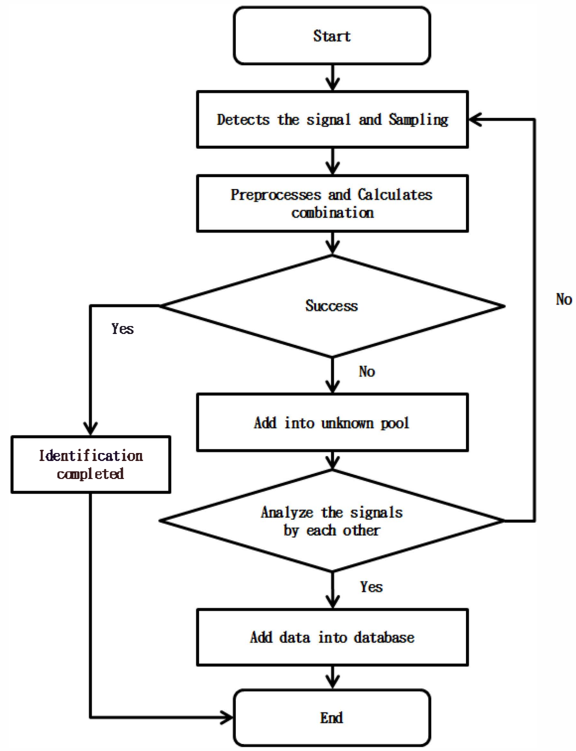

Fig. 8. Algorithm flow diagram

V. Experiment and Results Analysis

A. Experiment Design

Experiment is setup for Residential appliances, and the appliances are usually less than 100 watt in Table.1. In order to simulate the actually non-invasive situation, we randomly extract ten sampling of a day. And the system randomly selects three amounts of appliances as known appliances, then does preprocess and computes their classifications. This is set the database for identification. One or more appliances are selected in their classifications randomly, including relatively fixed and in changing greatly. In addition, select one out of them to test detection unknown appliances algorithm. In this experiment, we set confidence level to control the identification range.

TABLE I
The Appliances and Their Consume

Appliance	Power (W)
LCD 1	49 ~ 53
LCD 2	25 ~ 27
Fan (strong speed)	43 ~ 46
Fan (soft speed)	35 ~ 38
Lamp	20 ~ 22
Hair drier (Speed 1)	23 ~ 25
Hair drier (Speed 3)	574 ~ 579
CRT Television	114 ~ 125

B. Results and Analysis

The results show that identification rate is quite high in single appliance in Fig. 9, despite it is relatively fixed or changing greatly, and the failure always occurs that into wrong simulated classification. And the rate of the multi appliances, success probability is some lower due to be not real appliances classification in Fig. 10. Domain space is still quite loose in three amounts of appliances, we can adjust the confidence level to seek better identification rate. Once confidence level set large, the measured value will fall into more kinds of appliances classification interval. It cause value with unknown appliances may be divided into known classification. Therefore, we set Z value to 5 that is satisfactory results between identification and detection.

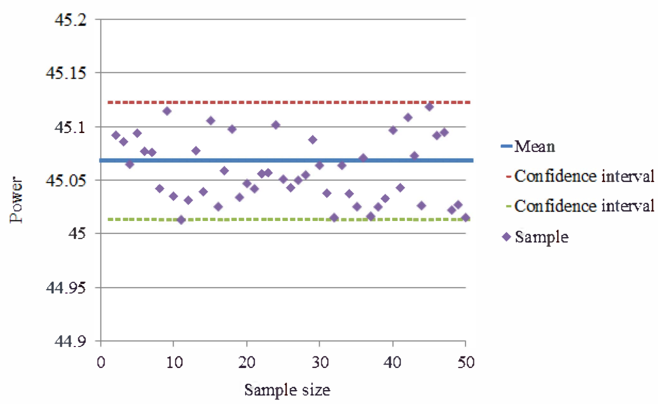

Fig. 9. The result of Fan (strong speed)

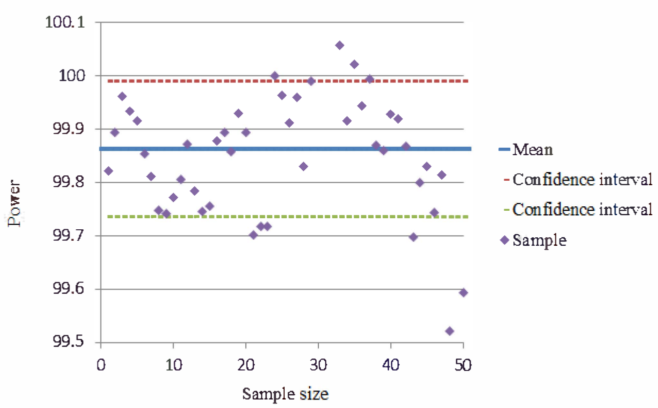

Fig. 10. The result by simulated classification of LCD1 , LCD2 and Lamp

TABLE II
PROBABILITY FOR AN APPLIANCE

Appliance Classification	Probability
LCD 1	96 %
LCD 2	90 %
Fan (strong speed)	100 %
Fan (soft speed)	100 %
Lamp	87 %
Hair drier (Speed 1)	85 %
Hair drier (Speed 3)	100 %
CRT Television	100 %

TABLE III
PROBABILITY FOR APPLIANCES BY SIMULATED CLASSIFICATION

Appliance Classification	Amount	Probability
LCD 1 + Fan (soft speed)	2	85 %
LCD 2 + Lamp	2	74 %
Fan (strong speed) + Lamp	2	87 %
Fan (soft speed) + Hair drier (Speed 3)	2	100 %
Lamp + Hair drier (Speed 1)	2	78 %
Hair drier (Speed 3) + CRT Television	2	92 %
LCD 2 + Hair drier (Speed 1) + Lamp	3	68 %
LCD 1 + Lamp + CRT Television	3	75 %

Table 2 3 is the experiment success probability with different classification.

TABLE IV
DETECTING UNKNOWN APPLIANCES
IN THREE APPLIANCES AMOUNT

Appliance	Times	Error %
LCD1	3	4 %
LCD2	3	10 %
Fan (strong speed)	3	0 %
Lamp	3	13 %
Hair drier (Speed 1)	3	0 %

The rate of detection unknown appliances is more than ninety percentage in the laboratory. The key is how many the signals of classifications with unknown appliances into the system. In other words, it is able to detect the unknown appliances completely, and it means how many appliances exist in the system and how many unknown identifications exist.

VI. CONCLUSION AND PERSPECTIVE

This paper proposes an identification appliance and detection unknown appliances method, the method have some advantages: Steady-state and low-cost, by simple power meters, do not require learning, non-intrusive completely, and multi appliances can be identified.

By simulating with every alone appliance, the system can obtain whole possibility classifications to establish a database and can reduce the number and cost of learning. The sample variance provides reaction in different characteristics. A sum of normal distribution is fairly effective in simulated classification that supports confidence interval. The detection method provides a database expansion.

The distribution of appliance classifications is a Pascal triangle in database, and it means that is less probability to occur only a single appliance or all appliances in circuit at the same time. In fact, the habit is often that two or three different appliances classifications in circuit. It can enhance the rate of detection unknown appliances. The experiment result shows over than half percentages in only two different classifications in three to four appliances.

The results show the algorithm is quite well in identification solution. A few failures would happen in domain overlap. Once more appliances are detected and extended to database, and the data of classifications are growth by series. So that it is squeeze in domain space and decreases some identification rate.

In the future, we are going to develop for different electrical parameters and higher dimension space domain. And we focus on how to solve more different electrical characteristics. The unknown signals are analyzed by multi sample power meter. It can get more application, which is expected to use fewer computing resources to get better results in Large-scale system.

VII. References

[1] S. Darby, "The effectiveness of feedback on energy consumption", 2006.

[2] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80,pp. 1870-1891, 1992.

[3] S. R. Shaw, et al., "Nonintrusive load monitoring and diagnostics in power systems," *Instrumentation and Measurement, IEEE Transactions on*, vol. 57, pp. 1445-1454, 2008.

[4] M. Berges and K. Shao, "Classifying Electrical Appliance State Transitions from Power Metrics Time-Series," 2008.

[5] Gu-Yuan Lin, Shih-Chiang Lee, and Jane Yung-Jen Hsu "Sensing from the panel: Applying the Power Meters for Appliance Recognition", In Proceedings of the 14th Conference on Artificial Intelligence and Applications. October, 2009.

[6] H. Najmeddine and K. El Khamlichi Drissi "State of art on load monitoring methods", *Proc. IEEE 2nd Int. Power Energy Conf. (PECon 2008)*, pp.1256 -1258

[7] Proper, Ethan R. (Ethan Richard), "Automated classification of power signals", Thesis (Nav. E.)--Massachusetts Institute of Technology, Dept. of Mechanical Engineering; and, (S.M.)--Massachusetts Institute of Technology, System Design and Management Program, 2008.

[8] W. Wichakool, A.T. Avestruz, R. W. Cox, S. B. Leeb, "Resolving power consumption of variable power electronic loads using nonintrusive monitoring," *IEEE Power Electronics Specialists Conference, 2007, Orlando, USA.*

[9] Marisa B Figueiredo, Ana De Almeida, Bernardete Ribeiro "An Experimental Study on Electrical Signature Identification of Non-Intrusive Load Monitoring (NILM) Systems", A. Dobnikar, U. Lotrič, and B. Šter (Eds.): *ICANNGA 2011*, Part II, LNCS 6594, pp. 31–40, 2011.

[10] Y. H. Lin and M.S. Tsai, "A novel feature extraction method for the development of nonintrusive load monitoring system based on BPANN," *International Symposium on Computer Communication Control and Automation (3CA)*, pp. 215-218, 2010.

[11] S. Inagaki, et al., "Nonintrusive Appliance Load Monitoring based on Integer Programming," *IEEJ Transactions on Power and Energy, vol. 128*, pp. 1386-1392, 2008.

[12] Khaled Chahinea, Khalil El Khamlichi Drissia, Christophe Pasquiera, Kamal Kerrouma, Claire Faurea, Thierry Jouannetb and Michel Michoub, "Electric Load Disaggregation in Smart Metering Using a Novel Feature Extraction Method and Supervised Classification", *Khaled Chahine et al. / Energy Procedia 6 (2011),* pp.627–632

[13] A. Dobnikar, U. Lotric and B. er (Eds.): *ICANNGA* 2011, Part II, LNCS 6594, pp. 31, 2011.