16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

SEGMENTED SELF-ORGANIZED FEATURE EXTRACTION FOR ONLINE FILTERING IN A HIGH EVENT RATE DETECTOR

Eduardo F. Simas Filho[1,2], *José M. Seixas*[1] *and Luiz P. Calôba*[1]

[1] Signal Processing Laboratory, Federal University of Rio de Janeiro, Rio de Janeiro-RJ, Brazil
[2] Federal Center for Technological Education, Simões Filho-BA, Brazil
emails: esimas@lps.ufrj.br, seixas@lps.ufrj.br, caloba@lps.ufrj.br

ABSTRACT

In this work, a novel feature extraction strategy is proposed for the electron/jet channel of ATLAS detector second level trigger. Placed around one of the collision points of LHC (the next generation particle accelerator), ATLAS will be responsible for the selection and recording of relevant information. A huge amount of data will be generated, in spite of that, only a few will be relevant for characterization of new physics. Considering this, an efficient triggering system is very important to maximize the detector performance. A segmented signal processing routine is proposed here in order to make use of different characteristics of each detector layer. Self-Organizing Maps (SOM) were trained for each layer, and further adjusted through Learning Vector Quantization (LVQ) to maximize particle discrimination. Neural classifiers perform electron/jet identification using as inputs the Segmented SOM information. Through the proposed approach, a discrimination efficiency of 97.4% was achieved for a false alarm probability of 2.4%.

1. INTRODUCTION

In particle physics, collider experiments are often used to search for new physics channels. The next-generation collider experiment, LHC (Large Hadron Collider), will be colliding bunches of protons at every 25 ns for a wide research program [1]. Despite this very high event rate, the interesting channels will rarely occur. Therefore, a high-efficient filtering (triggering) system is required to guarantee that most of the background noise will be rejected and valuable information will not be lost.

In order to search for the Higgs particle and other new phenomena a new energy range will be explored by LHC, that, when operating at full capacity, will produce 40×10^6 events/s. The total detector information will be near 60 TB/s (1,5 MB/event). Considering this amount of data, the filtering procedure must be performed online within short latency time.

ATLAS is one of the main detectors placed around the collision points of LHC. The ATLAS online trigger system [2] has three filtering levels and access information from calorimeters, muon chamber and tracking (see Figure 1). The first trigger level (LVL1) shall reduce the event rate to 75 kHz. The decision must be taken on less than 2.5 microseconds and thus LVL1 uses detector data with reduced granularity and is implemented in hardware. The High-Level Trigger (HLT), which comprises the second level trigger (LVL2) and the event filter (EF), receives from the first level the primary location of regions in the detector where interesting events appeared. These detector regions are known as Regions of Interest (RoI). The HLT is implemented in software, processed in parallel by thousands of PCs, and must reduce the event rate to 100 Hz. The HLT overall processing time is about one second (10 ms for LVL2).

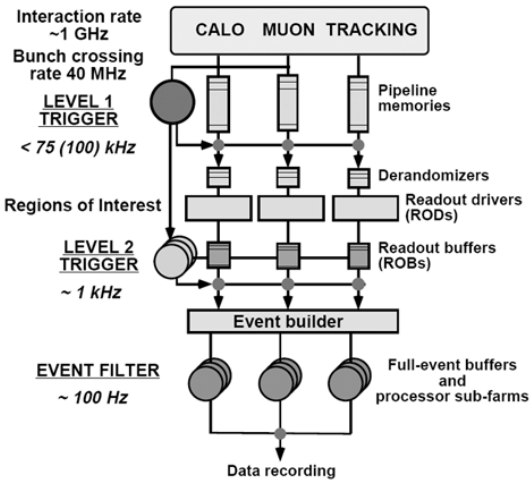


Figure 1: ATLAS trigger system architecture.

Due to their fast response, good linearity and high resolution, calorimeters play an important role in most modern collider experiments. Indeed, the ATLAS online filtering system relies very much on its calorimeter system. Calorimeters [3] measure the energy of the incoming particles by means of total absorption. The energy deposition profile allow high efficient particle discrimination.

For LHC, interesting signatures (Ex: Higgs boson) can be found through decays that produce electrons. Hadronic jets present energy deposition profiles similar to electrons, forming a huge background noise for the experiment. In this work, electron/jet (e^-/j) discrimination discrimination is performed in the second-level (LVL2) trigger of ATLAS by using calorimeter information. As illustrated in Figure 2, ATLAS calorimeter system is segmented into seven layers, four electromagnetic (PS, E1, E2 and E3) and three hadronic (H0, H1 and H2), each one with different physical characteristics and cell granularity [4].

The particle discrimination procedure at LVL2 may be splitted into feature extraction, where relevant information is extracted from the measured signals, and hypothesis testing, where particle discrimination is performed. In this paper, in order to cope with the different levels of granularity present at each detector section, a segmented feature extrac-

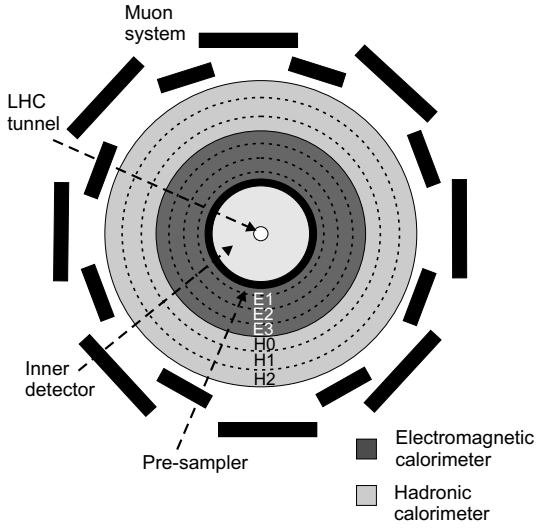


Figure 2: ATLAS detector diagram.

tion procedure is proposed. The information is processed separately, in each calorimeter layer, using Segmented Self Organizing Maps (S-SOM), further adjusted through Learning Vector Quantization (LVQ) algorithm [5] to maximize discrimination performance. Supervised neural classifiers [6] perform the hypothesis testing using as inputs S-SOM information. The proposed approach obtained better performance when compared to ATLAS baseline discriminating algorithm (T2Calo) [2].

2. CALORIMETER SIGNALS FORMATTING

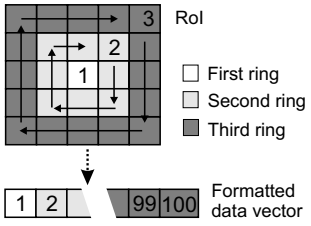
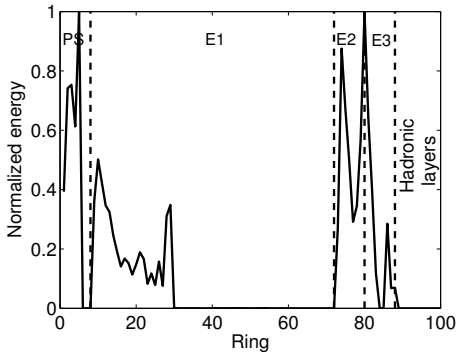


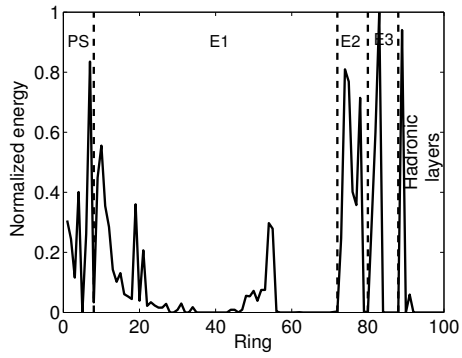
Figure 3: Ring formatting procedure.

As proposed in [7], here the RoI information is formatted into concentric energy deposition rings. In each calorimeter layer, the most energetic cell is considered as the first ring, and the following rings are sequentially formed around the first one (see Figure 3), adding the energy of the cells belonging to a ring to form ring signals. This procedure makes the signals independent from the impact point in the detector. The ring energy is normalized within each layer. As a result, a RoI is described by 100 rings.

Ring signals from typical electron and jet are illustrated in Figure 4. The calorimeter layers are limited by vertical dotted lines. One can see that the signals present similar pattern and their discrimination is not a trivial task.



(a)



(b)

Figure 4: Ring formatted signal of a typical electron (a) and jet (b).

3. SEGMENTED FEATURE EXTRACTION

Each detector layer presents different characteristics and cell granularity. Considering this, better characterization of the physical phenomena is obtained if the signal processing procedure is performed separately at each layer. In this work, a feature extraction algorithm using segmented self-organizing maps (S-SOM) is proposed for the ATLAS calorimeter based second-level trigger.

3.1 Self-Organizing Maps

The Self Organizing Map [5] is an unsupervised trained neural network that realizes a topological organization of the input data set. The SOM transforms a k-dimensional continuous input space into a discrete characteristics map (generally bidimensional). Each neuron of the map is fully connected to all inputs. SOM is widely applied on different tasks like clustering [8], data mining [9] and nonlinear independent component analysis [10].

Basically three processes are involved in the SOM training phase: competition (for each input vector there is only one activated neuron), cooperation (the winner neuron determines a neighborhood of excited neurons) and adaption (the map weights are adjusted to reenforce the winner neuron response to a given input pattern) [6].

The network outputs are obtained for each neuron i through equation 1:

$$u_i = \mathbf{x}^T \mathbf{w}_i \tag{1}$$

where $\mathbf{w}_i$ is a vector formed by the weights w_{ij} that connect the input x_j to the neuron i.

The inputs are fully connected to each neuron, and the map weights are calculated iteratively through:

$$w_j(n+1) = w_j(n) + \eta(n)h_{ij}(n)(x(n) - w_j(n)) \quad (2)$$

where $\eta(n)$ is the learning rate and $h_{ij}(n)$ the neighborhood function, which is defined as:

$$h_{ij}(n) = \exp(-d_{ij}^2/2\sigma^2(n)) \quad (3)$$

Self organizing maps belong to a class of algorithms of vector quantization, searching for a fixed number of vectors (or code-words) that better describe the input data set. The vectors $\mathbf{w}_i$ form the SOM code-book.

For a specific input vector $\mathbf{x_a}$, the highest value of u is obtained for the neuron i that minimizes the distance $|\mathbf{x_a} - \mathbf{w}_i|$ [5], in other words, $\mathbf{w}_i$ is the code-word closest to $\mathbf{x_a}$. As the map is topologically ordered, similar input patterns are mapped into neighbor regions. The neuron with highest output value is called the winner neuron.

When applied to classification problems, it is expected that the SOM concentrates different inputs in opposite sides of the map, facilitating the hypothesis testing procedure.

3.2 Learning Vector Quantization

Vector quantization (VQ) is a coding technique in which a input data set is mapped into a finite group of representative vectors. [6]. The k-dimensional input space is divided into a finite number of regions and the quantizer Q maps $\mathbb{R}^k$ into a finite subset Y of $\mathbb{R}^m$:

$$Q: \quad \mathbb{R}^m \rightarrow \mathbf{Y} \quad (4)$$

where $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_k\}$ is the code-book. For each code-word (m-dimensional vector) $\mathbf{y}_i$ there is a partition R_i of the input space that satisfies:

$$R_i = Q^{-1}(\mathbf{y_i}) = \{\mathbf{x} \in \mathbb{R}^k : Q(\mathbf{x}) = \mathbf{y}_i\} \quad (5)$$

$$\bigcup_{i=1}^{N} R_i = \mathbb{R}^k, \quad R_i \bigcap R_j = 0, \quad i \neq j \quad (6)$$

The code-words can be approximately estimated by the SOM through unsupervised learning procedure and the code-book is formed by the synaptic weights. For classification purposes, through a supervised procedure (using target information), the Learning Vector Quantization (LVQ) algorithm [5] slightly adjusts the code-words location (obtained through SOM) to improve the map clustering.

The LVQ algorithm selects randomly a input vector $\mathbf{x}$ and verifies whether or not its class type $\mathscr{C}_{\mathbf{x_i}}$ is the same as the one the code-word $\mathbf{w_c}$ belongs to. In case both classes are the same, $\mathbf{w_c}$ is moved towards $\mathbf{x}$ [5]:

$$\mathscr{C}_{\mathbf{w_c}} = \mathscr{C}_{\mathbf{x_i}} \rightarrow \mathbf{w_c}(n+1) = \mathbf{w_c}(n) + \alpha[\mathbf{x} - \mathbf{w_c}(n)] \quad (7)$$

where α is the learning rate ($0 < \alpha < 1$). Otherwise, $\mathbf{w_c}$ is moved away from $\mathbf{x}$ ($-1 < \alpha < 0$).

3.3 Hypothesis Testing

The map outputs, after LVQ adjustment, were used to feed Multi-layer Perceptron (MLP) classifiers [6], trained through the resilient back-propagation (RPROP) algorithm [11]. The networks comprise a single hidden layer and one output neuron. The number of hidden neurons was chosen after testing exhaustively the discrimination performance of each network.

Two distinct neural network architectures were used for hypothesis testing:

Type 1 (non-segmented): The outputs $X_k = S_k U_k$ of the S-SOM trained for calorimeter layer k (where S_k are the S-SOM weight vectors and U_k the ring formatted signals) are arranged as a single input vector $X = [X_{PS}, X_{E1}, ..., X_{H2}]^T$ to feed the neural classifier. A limitation of this procedure, considering LVL2 latency (10 ms), is that the dimensionality of the problem increases (the dimension of X may be much higher than the number of calorimeter rings U, that amounts to 100).

Type 2 (segmented): Different classifiers were trained to maximize particle discrimination at each calorimeter layer k using X_k as inputs. The input vetor of the segmented MLPs hidden neurons is defined as:

$$Z_k = W_{1,k}X_k + B_{1,k} = f_k(X_k) \quad (8)$$

where $W_{1,k}$ and $B_{1,k}$ are respectively the neuron weights and bias vector, and the MLP output is computed through: $Y_k = \sigma(W_{2,k}H_k + B_{2,k})$, where $H_k = \sigma(Z_k)$ ($\sigma(.) = \tanh(.)$ is the neurons activation function).
After training has converged, the operator f_k can be interpreted as a linear transformation of X_k that maximizes electron/jet discrimination at layer k. As the number of hidden neurons is usually smaller than the number of components of X_k, a compact representation, which preserves the most relevant information needed for particle identification, is obtained. Considering this, in the type 2 hypothesis testing, the vector $Z = [Z_{PS}, Z_{E1}, ..., Z_{H2}]^T$ was used as input for a neural classifier. The dimension of Z is usually smaller than the number of calorimeter rings, reducing the computational complexity.

4. EXPERIMENTAL RESULTS

The database used in this work was obtained through a Monte Carlo simulator for proton-proton collisions that considers both detector characteristics and first-level trigger effects [2]. The available dataset, which comprises 22,581 electron and 7,509 jet signatures, was equally splitted into training and testing sets.

In order to chose the proper size of the SOM for each layer, the mean approximation error (MAE) was computed through Equation 9, varying the number of neurons in the SOM grid (P):

$$MAE = \frac{1}{N}\sum_{n=1}^{N} ||u_n - s_n^*|| \quad (9)$$

where N is the number of available signatures u_n in the training set and s_n^* is the weight vector of the S-SOM winning neuron, when the input u_n is presented.

As illustrated in Figure 5, the MAE is inversely proportional to the number of output neurons P and moreover

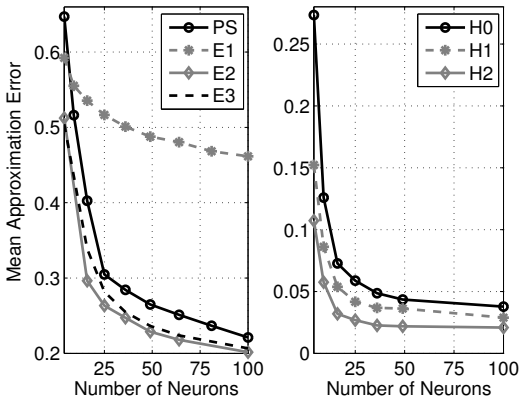


Figure 5: Mean Approximation Error (MAE) as a function of the number of output neurons.

Table 1: Optimum S-SOM dimensions (P^*) for each calorimeter layer (L).

L	**PS**	**E1**	**E2**	**E3**	**H0**	**H1**	**H2**
P^*	5×5	7×7	6×6	6×6	4×4	4×4	4×4

$|\frac{\partial(MAE)}{\partial(P)}|$ presents high values for small size maps, decreasing fast with P, until error stabilization ($|\frac{\partial(MAE)}{\partial(P)}| \cong 0$). In this work, the purpose was to achieve good description of the data (low MAE) while preventing oversized maps (high P). In order to do this and considering results in Figure 5 (especially the knee points) the optimum value P^* was chosen around the point:

$$MAE^* = MAE_{MAX} - 0.8 \times \Delta_{MAE} \tag{10}$$

where MAE_{MAX} is the maximum error and $\Delta_{MAE} = MAE_{MAX} - MAE_{MIN}$. Table 1 shows the SOM dimension chosen through this criteria for each calorimeter layer. Using these maps, the dimension of the S-SOM output vector $X = [X_{PS}, X_{E1}, ..., X_{H2}]^T$ amounts to 194.

Discrimination performance was evaluated through both ROC (Receiver Operating Characteristic) curve [12] and SP product. The ROC illustrates how both the detection (P_D) and false alarm (P_F) probabilities vary as the decision threshold

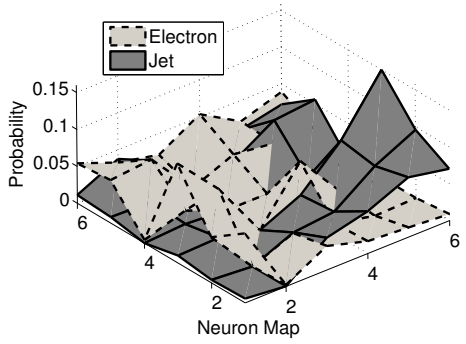


Figure 6: Neuron Activation probability computed before LVQ (E2 layer).

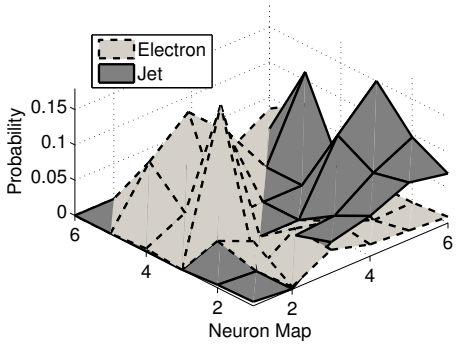


Figure 7: Neuron Activation probability computed after LVQ (E2 layer).

changes. The SP is defined as [7]:

$$\frac{Ef_e + Ef_j}{2} \times \sqrt{(Ef_e \times Ef_j)} \tag{11}$$

where $Ef_e = P_D$ is the detection efficiency for electrons and $Ef_j = (1 - P_F)$ is the corresponding efficiency for jets. The threshold value that maximizes SP provides both high P_D and low P_F.

As illustrated in Figure 6, for E2 layer, the S-SOM (considering a winner-takes-all operation) was able to concentrate the particle types at opposite sides of the map. Similar results were obtained for the other calorimeter layers.

Using the LVQ algorithm to further adjust the map weights, by means of supervised training, the SOM outputs were shifted towards the map edges (see Figure 7). Therefore the border between the mapped regions assigned to electrons and jets was reduced as a result of LVQ.

Considering the proposed hypothesis testing procedures, through the non-segmented (Type 1) routine, the dimensionality of the problem increases from 100 calorimeter rings to 194 nodes, obtained from the S-SOM, used to feed the MLP classifier. In the segmented (Type 2) hypothesis testing procedure, the calorimeter information was compacted into 24 segmented discriminating components. The SP products computed, respectively for Type 1 and Type 2 classifiers, were 0.950 (P_D=97.4%, P_F=2.4%) and 0.948 (P_D=97.9%, P_F=3.2%). Considering this, Type 2 presented slightly poorer performance, but with the advantage of minimizing the computational complexity of the discrimination routine, which is highly desired.

Different e^-/j discriminators were compared using this simulated second-level trigger database (testing set). The baseline algorithm for e^-/j discrimination used at ATLAS (T2Calo) [2] extracts, directly from calorimeter measurements, parameters that estimate the shape of the energy deposition profile. Thresholds on these parameters perform the particle discrimination. The Neural_Ringer [7] is another particle discrimination procedure that is implemented in the ATLAS software platform. Using a MLP neural classifier operating over the ring formatted signals, Neural_Ringer algorithm achieved better discrimination performance and similar computational cost when compared to T2Calo. In a previous work [13], Segmented Independent Component Analysis (SICA) was applied for feature extraction and a MLP classifier performed the hypothesis testing.

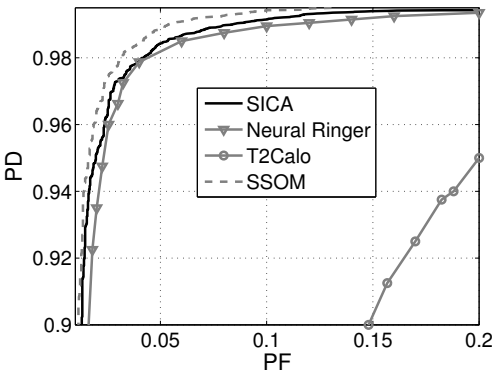


Figure 8: ROC curves.

Table 2: False electron signature rate (jets/second) for 95% efficiency considering different techniques.

SSOM	SICA	NeuralRinger	T2Calo
410	490	580	5,000

Analyzing the ROC curves (see Figure 8), the feature extraction through S-SOM improves the discrimination efficiency when compared to SICA and Neural_Ringer and T2Calo. For the proposed S-SOM hypothesis testing procedures, Type 2 was considered in this plot.

Considering LHC operating conditions at high luminosity, 25,000 jets/second are expected to reach ATLAS LVL2 [2]. It means that, for a fixed P_D, 1% increase of P_F implies on recording more 250 false electrons per second. As shown in Table 2, for P_D=95%, the proposed technique reduces the false alarm (less 80, 170 and 4590 jets/sec when compared respectively to SICA, Neural_Ringer and T2Calo), providing cleaner data for offline analysis.

A proper study of the processing time is needed for the proposed hypothesis testing algorithms, but a simple comparison with the Neural_Ringer discriminator indicates that the HLT requirements should be satisfied. It has been demonstrated that the ring formatting and neural classification algorithms, implemented at the actual HLT software platform, are faster than T2Calo [7]. Considering S-SOM Type 2 classifier, the segmented discriminating components can be expressed as: $Z_k = W_{1,k}X_k + B_{1,k}$, where $X_k = S_k U_k$ are the S-SOM outputs. After some simple manipulation one can verify that $Z_k = \Theta_k U_k + B_{1,k}$, where $\Theta_k = W_{1,k}S_k$, and so, only a matrix multiplication (Θ_k) is added to Neural_Ringer algorithm. Type 1 discriminator also adds only a matrix multiplication to Neural_Ringer, but requires more detailed study, as problem dimensionality increases from 100 to 194.

5. CONCLUSIONS

A novel feature extraction procedure was proposed for the ATLAS detector second-level trigger. Segmented Self-Organizing Maps (S-SOM) were applied for feature extraction over ring formatted calorimeter signals.

Neural classifiers perform electron/jet discrimination using as inputs the S-SOM information. The method presents better performance when compared to the baseline electron/jet discriminator in use at ATLAS, achieving 97.4% electron efficiency for 2.4% of jet misclassification.

A proper study of the processing time must be performed with the proposed algorithm implemented at the ATLAS trigger software platform, although a preliminary analysis indicates that the LVL2 time requirements should be satisfied.

6. ACKNOWLEDGEMENTS

The authors are thankful for the financial support provided to this work by CNPq and FAPERJ (Brazil) and the European Union (through the HELEN project). We also thank the ATLAS Trigger/DAQ collaboration for providing the simulation data and for fruitful discussions concerning this work.

REFERENCES

[1] M.J. Price, "The LHC Project", *Nuclear Instruments and Methods in Physics Research*, 2002.

[2] A. Gesualdi Mello, et al., "Overview of the High-Level Trigger Electron and Photon Selection for the ATLAS Experiment at the LHC", *IEEE Transactions on Nuclear Science*, vol. 53, n. 5, pp. 2839-2843, October, 2006.

[3] R. Wigmans, "Calorimetry: Energy Measurement in Particle Physics", Clarendon Press, 2000.

[4] P. Adragna, et al., "The ATLAS Hadronic Tile Calorimeter: From Construction Toward Physics", *IEEE Transactions on Nuclear Science*, vol. 53, n. 5, pp. 1275-1281, June, 2006.

[5] T. Kohonen, *Self Organizing Maps*, Springer-Verlag, 3rd Ed., 2001.

[6] S. Haykin, *Neural Networks, Principles and Practice*, Bookman, 2001.

[7] A. Anjos, R.C. Torres, J.M. Seixas, B.C. Ferreira, T.C. Xavier, "Neural Triggering System Operating on High Resolution Calorimetry Information", *Nuclear Instruments and Methods in Physics Research*, No. 559, pp. 134-138, 2005.

[8] E. F. Simas Filho, J. M. Seixas and L. P. Caloba, "Local Independent Component Analysis Applied to Highly Segmented Detectors ", in *Proceedings of the IEEE International Symposium on Circuits and Systems*.

[9] T. Babnik, R. Aggarwal and P. Moore, "Data mining on a transformer partial discharge data using the self-organizing map", *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 14, pp. 444–452, April 2007.

[10] M. Haritopoulos, H. Yin, N. M. Allinson, *Image Denoising using Self-organizing Map-based Nonlinear Independent Component Analysis*, Neural Networks Journal, Vol.15, pp. 1085-1098, 2002.

[11] M. Riedmiller, H. Braun, *A Direct Adaptive Method for Faster Backpropagation Learning, the RPROP algorithm*, Proceedings of the IEEE International Conference on Neural Networks, pp. 586-591, 1993.

[12] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, John Wiley, 2001.

[13] E. F. Simas Filho, J. M. Seixas and L. P. Caloba, "Segmented Independent Component Analysis for Online Filtering Using Highly Segmented Detectors", *Proceedings of the 7th International Conference on Intelligent Systems Design and Applications - ISDA07*, pp. 659-664, October, 2007.