

## LIKELIHOOD APPROXIMATION FUNCTIONS FOR CHARACTERIZING TRANSABDOMINAL PROSTATE ULTRASOUND IMAGES.

PantelisTheocharakis<sup>1</sup>, Dimitris Glotsos<sup>1</sup>, Nikos Dimitropoulos<sup>2</sup>, Ioannis Kalatzis<sup>3</sup>, George Nikiforidis<sup>1</sup>,  
Dionisis Cavouras<sup>3</sup>.

<sup>1</sup> Medical Image Processing and Analysis Group, Laboratory of Medical Physics,  
University of Patras  
26500 Patras, Greece  
email: [pthoach@med.upatras.gr](mailto:pthoach@med.upatras.gr), web page: <http://mipa.med.upatras.gr>

<sup>2</sup> Department of Medical Imaging  
EUROMEDICA Medical Center,  
Mesogeion Avenue 2, Athens, Greece.

<sup>3</sup> Medical Image and Signal Processing Lab., Dept. of Medical Instruments Technology, STEF,  
Technological Educational Institution of Athens,  
Ag. Spyridonos Street, Egaleo, GR-122 10, Athens, Greece.  
email: [cavouras@teiath.gr](mailto:cavouras@teiath.gr), web page: <http://medisp.bme.teiath.gr/>

**Keywords:** Bayesian Regularised Neural Network, Quadratic Bayesian, Prostate, Textural Features

### Abstract.

**Purpose:** *The Quadratic Bayesian (QB) and the Bayesian Regularized Neural Network (BRNN) classifiers, both based on the likelihood approximation of Bayes theorem, were applied to classify normal from hyperplasia and mild from medium-high hyperplasia transabdominal prostate ultrasound (US) images.*

**Material and Methods:** *Two sets of forty-five prostate pathologically confirmed US-images (mild hyperplasia and medium-high hyperplasia of prostate gland) were obtained using an HDI-3000 ATL digital ultrasound system with a wide band (5-12 MHz) probe. An equal number (45) of images with no pathological findings (normal) were also selected. From each image, a 30x30 pixel region of interest (ROI) was extracted and textural features were calculated. Classification was performed employing a two-level hierarchical tree structure. At the first level, each prostate image was classified as normal or abnormal and, if abnormal, it was characterized at the 2<sup>nd</sup> level as of mild or medium-high hyperplasia. Both classifiers were trained using at the 1<sup>st</sup> level 30 normal and 60 pathogenic and, at the 2<sup>nd</sup> level, 30 mild and 30 medium/high hyperplasia images and their performances were verified using the left out images from each set. For input into the classifiers, textural features, calculated from the selected ROIs, with highest discriminatory power ( $p < 0.001$ ) were determined by means of the t-test.*

**Results and Discussion:** *The QB's and BRNN's accuracies were 91.1% and 95.5% respectively at the first level, and 90% and 93.3% respectively, at the 2<sup>nd</sup> level of the classification tree. Best feature combinations for the QB classifier were skewness, variance and difference-variance at the 1<sup>st</sup> level and mean correlation, entropy, variance and sum of average at the 2<sup>nd</sup> level. For the BRNN classifier it was mean value and inverse difference moment at the 1<sup>st</sup> level and sum of entropy, entropy and long run emphasis at the 2<sup>nd</sup> level.*

**Conclusions:** *The proposed image analysis system may be of value to the radiologists in assessing prostate lesions on transabdominal ultrasound images.*

## 1. INTRODUCTION

The prostate gland is located between the bladder and the rectum and wraps around the urethra. It comprises smooth cells, glandular cells and stromal cells. The prostate takes its final size through puberty and remains stable until the mid-forties<sup>[1]</sup>. One of the most common non-skin malignancies in elderly men is the hyperplasia of the prostate gland. Many of the cases are small and clinically insignificant. However, some are not, and prostatic hyperplasia can lead to very uncomfortable situations.

Prostate hyperplasia may be at first assessed by transabdominal ultrasound examination<sup>[1,2]</sup>. The major risk factors, which are correlated with prostate hyperplasia, include age, family history, and genetic factors, ethnicity, infections and inflammation, dietary factors, heavy alcohol intake and lack of exercise<sup>[1]</sup>.

Artificial Neural Networks have been previously employed in Urology<sup>[3]</sup> for prostate cancer estimation, by using ultrasound pictorial or non-pictorial (e.g. prostatic specific antigen levels) information as input to the

Neural Networks [3,4,5]. The aim of this study was to investigate the performance of textural features and Bayesian based classifiers in discriminating ultrasound images of patients with prostate hyperplasia.

## 2. MATERIALS AND METHODS

### 2.1 Data acquisition

A set of a 135 transabdominal ultrasound images, from an equal number of patients, were acquired by a HDI-3000 ATL digital ultrasound system by the same experienced physician (N.D.) . The scanning protocol was kept fixed in order to avoid non-uniformity in data acquisition. Forty-five images were verified as having no signs of suspicious malignancy, 45 images were diagnosed as of mild hyperplasia and 45 as of medium-high hyperplasia. The images were digitized by a video frame grabber (512x512x8), which was connected to the US video output. Custom developed software was designed in MATLAB to be used for image segmentation (Fig. 1), feature extraction, and image classification.



Figure 1. The custom made computer software for segmentation, textural feature extraction and image classification .

### 2.2 Feature Extraction.

From each image, a 30x30 pixel region of interest (ROI) was extracted (see Fig. 1). For each segmented ROI, nineteen features were extracted, four from the first order statistics and fifteen from the second order statistics (10 from the co-occurrence matrix<sup>[6]</sup> and 5 from the run-length matrix<sup>[7]</sup>). Data were normalized to zero mean value and unit standard deviation<sup>[8]</sup>.

$$\tilde{x}_i = \frac{x_i - \mu}{\sigma} \quad (1)$$

where  $x_i$  is the  $i$ th feature pattern vector,  $\tilde{x}_i$  the normalized pattern vector,  $\mu$  the mean value and  $\sigma$  the standard deviation

### 2.3 Classifier evaluation and feature selection.

Both classifiers were trained using at the 1<sup>st</sup> level 30 normal and 60 pathogenic (30 mild and 30 medium/high) and, at the 2<sup>nd</sup> level, 30 mild and 30 medium/high hyperplasia images and their performances were verified using the left out images from each set. The shape of the hierarchical classification tree is shown in figure 2.

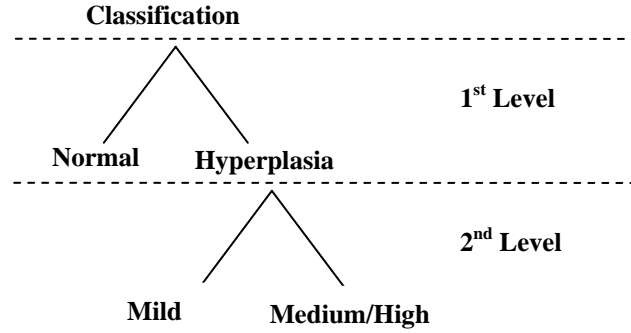


Figure 2. The hierarchical classification tree.

The statistical t-test was applied to each classification level. The p-value was determined to 0.001. The features that had  $p < 0.001$  were chosen as input to the classifiers (highest discriminatory power).

#### 2.4 Design of Quadratic Bayesian Classifier

Quadratic Bayesian (QB) classifier<sup>[9]</sup> places the pattern recognition problem into a statistical framework environment. Each of the two different classes, normal and hyperplasia, have a specific distribution of data, through stochastic and deterministic parameters, such as differences in echogenicity of the tissues and different stage of pathology, respectively.

By assuming that each class's probability density function  $p(x/\omega_i)$ , is multivariate normal (Gaussian), we designed the hyper-quadratic Bayesian decision function<sup>[9]</sup>:

$$d_i(x) = \ln p(\omega_i) - \frac{1}{2} \ln |C_i| - \frac{1}{2} \left[ (x - m_i)^T C_i^{-1} (x - m_i) \right] \quad (2)$$

where  $d_i(x)$  is the decision function of  $x$  pattern vector corresponding to the  $i$ th element of  $\omega$  class.  $P(\omega_i)$  is the probability of  $i$ th class.  $C$  is the covariance matrix.  $C^{-1}$  is the inverse of the covariance matrix and  $m_i$  is the mean value of the  $i$ th class.

$P(\omega_i)$  was assumed 0.33 and 0.77 at the 1<sup>st</sup> level and 0.5 and 0.5 at the 2<sup>nd</sup> level of the classification tree, since at the 2<sup>nd</sup> level we have the same number of images in both classes<sup>[9]</sup>.

#### 2.5 Design of the Neural Network with Bayesian regularization.

The implemented supervised neural network<sup>[10]</sup> consisted of three layers. The input layer with inputs the set of features with the highest discriminatory power at each hierarchical tree level, two hidden layers with ten and five neurons respectively, with the sigmoid transfer function, and the output layer with the linear transfer function. The output layer had only one single output with ranged targets from  $-1$  to  $1$ . The Levenberg – Marquardt<sup>[10]</sup> backpropagation algorithm with a Gauss-Newton approximation to the Hessian matrix<sup>[10,11]</sup> was used in order to reevaluate the weight and bias.

By presenting all inputs to the network the corresponding network outputs are computed by equation:

$$a^{m+1} = f^{m+1}(W^{m+1} \cdot a^m + b^{m+1}) \quad (3)$$

for  $m=0,2,\dots,M-1$ , where  $a$  is the vector with the output of each  $m$  layer,  $W$  is the weight vector,  $f$  is the transfer function,  $a$  is the pattern vector and  $b$  the bias vector. If each target occurs with equal probability, the mean squared error is proportional to the sum of squared errors over the  $Q$  targets in the training set<sup>[10]</sup>:

$$F(x) = \sum_{q=1}^Q (t_q - a_q)^T (t_q - a_q) = \sum_{q=1}^Q e_q^T e_q = \sum_{q=1}^Q \sum_{j=1}^{S^M} (e_{j,q})^2 = \sum_{i=1}^N (v_i)^2 \quad (4)$$

where  $e_{j,q}$  is the  $j$ th element of the error for the  $q$ <sup>th</sup> input/target pair. The error vector is:

$$V^J = \begin{bmatrix} v_1 & v_2 & \dots & v_N \end{bmatrix} = \begin{bmatrix} e_{1,1} & e_{2,1} & \dots & e_{S^M,1} & e_{1,2} & \dots & e_{S^M,Q} \end{bmatrix} \quad (5)$$

the parameter vector is

$$x^T = \begin{bmatrix} x_1 & x_2 & \dots & x_N \end{bmatrix} = \begin{bmatrix} w_{1,1}^1 & w_{1,2}^1 & \dots & w_{S^M,R}^1 & b_1^1 & \dots & b_{S^M}^1 & w_{1,1}^M & \dots & w_{S^M}^M \end{bmatrix} \quad (6)$$

and  $N = Q \times S^M$ , where S is the number of neuron of the  $m$ th hidden layer. The whole idea is to update the  $x$  vector by :

$$\Delta x_k = - \left[ J^T(x_k) J(x_k) + \mu_k I \right]^{-1} J^T(x_k) v(x_k) \quad (7)$$

$$J(x) = \begin{bmatrix} \frac{\partial e_{1,1}}{\partial w_{1,1}^1} & \frac{\partial e_{1,1}}{\partial w_{1,2}^1} & \dots & \frac{\partial e_{1,1}}{\partial w_{S^M,R}^1} & \frac{\partial e_{1,1}}{\partial b_1^1} & \dots \\ \frac{\partial e_{2,1}}{\partial w_{1,1}^1} & \frac{\partial e_{2,1}}{\partial w_{1,2}^1} & \dots & \frac{\partial e_{2,1}}{\partial w_{S^M,R}^1} & \frac{\partial e_{2,1}}{\partial b_1^1} & \dots \\ \vdots & \vdots & & \vdots & \vdots & \\ \frac{\partial e_{S^M,1}}{\partial w_{1,1}^1} & \frac{\partial e_{S^M,1}}{\partial w_{1,2}^1} & \dots & \frac{\partial e_{S^M,1}}{\partial w_{S^M,R}^1} & \frac{\partial e_{S^M,1}}{\partial b_1^1} & \dots \\ \frac{\partial e_{1,2}}{\partial w_{1,1}^1} & \frac{\partial e_{1,2}}{\partial w_{1,2}^1} & \dots & \frac{\partial e_{1,2}}{\partial w_{S^M,R}^1} & \frac{\partial e_{1,2}}{\partial b_1^1} & \dots \\ \vdots & \vdots & & \vdots & \vdots & \end{bmatrix} \quad (8)$$

$$x_k^* = x_k + \Delta x_k \quad (9)$$

where  $k$  is the iteration number,  $J$  is the Jacobian matrix by Gauss-Newton approximation to the Hessian matrix,  $x_k^*$  the updated networks parameters by  $\Delta x_k^*$ . This algorithm has the very useful feature that, as  $\mu_k$  is increased, it approaches the steepest descent algorithm with small learning rate, while as  $\mu_k$  is decreased to zero, it becomes Gauss-Newton. The algorithm begins with  $\mu_k$  set to some small value (e.g. ,  $\mu_k=0.01$ ). If step does not yield a smaller value for  $F(x)$ , then the step is repeated with  $\mu_k$  multiplied by some factor  $\theta > 1$  (e.g.  $\theta=10$ ). Eventually  $F(x)$  should decrease, since we would be taking a small step in the direction of steepest descent<sup>[10]</sup>. If a step does not produce a smaller value for  $F(x)$ , then  $\mu_k$  is divided by  $\theta$  for the next step, so that the algorithm approaches Gauss-Newton, which would provide faster convergence.

The network's training stops if either (i) the maximum number of iterations (epochs) is reached, (ii) the performance goal (the desired  $F(x)$  minimum value) is achieved, (iii)  $\mu$  reaches a maximum prior arranged value, or (iv) the maximum allowance of time, in which the network should be trained, is met.

### 3. RESULTS AND DISCUSSION

Clinician's decision-making has always been based on facts and predictions for individual patients. Patients are assigned into groups depending on his knowledge and experience. His decision is based on the observation and interpretation of different facts, such as the value of prostate specific antigen (PSA), size of the prostate, age,

heredity etc. All of these parameters have lead to the establishment of many statistical models in urology<sup>[12]</sup>, such as the usage of ROC analysis, logistic regression<sup>[12,13]</sup> etc.. What we provide in this study is a new approach to this particular prostate pathology classification problem, by combining textural image characteristics and image analysis methods. Other studies have introduced artificial intelligence techniques but with the usage of medical and non-medical indexes<sup>[13,14]</sup>. Our perspective is the appliance of image processing and analysis methods. Useful diagnostic information can be derived by modeling and interpreting each of these textural features for this specific prostate pathology.

Table 1 shows the classification accuracy achieved by the Quadratic Bayesian classifier at the first classification level (normal VS hyperplasia) of the hierarchical classification tree. Difference variance from the co-occurrence matrix, skewness and variance both from the first order statistics comprised the best feature combination vector. The classifier attained 93.3% precision in classifying correctly 14 out of 15 normal samples and 90% accuracy in characterizing correctly 27 out of 30 hyperplasia ultrasound ROIs, thus providing an overall accuracy of 91.1%, at the 1<sup>st</sup> level of the hierarchical classification tree.

QB Classifier	Normal	Hyperplasia	Accuracy (%)
Normal	14	1	93.3
Hyperplasia	3	27	90
Overall Accuracy			91.1

Table 1. Classification accuracies achieved by the QB for discriminating normal from hyperplasia images.

Figure 3 shows a 3-dimensional scatter diagram and decision boundary of the Bayesian Quadratic Classifier.

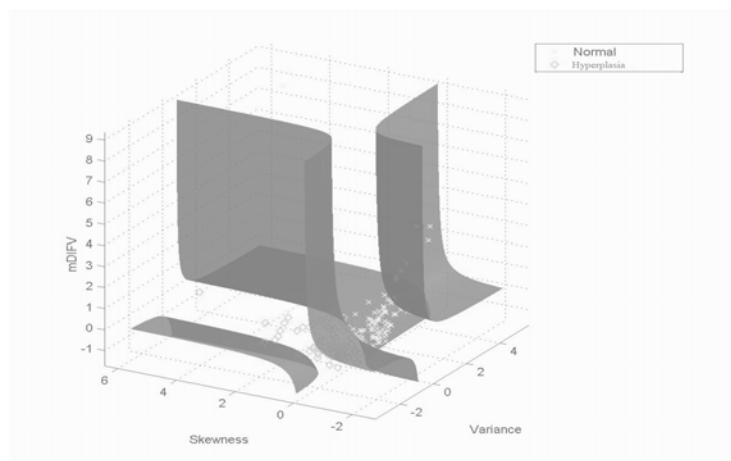


Figure 3. QB scatter diagram and decision boundary at the 1<sup>st</sup> level of the hierarchical tree structure.

Table 2 shows the accuracy for BRNN classifier at the 1<sup>st</sup> level of the hierarchical tree structure. Mean value and inverse difference moment from the co-occurrence matrix were the input features to the network. The classifier attained 93.3% precision in classifying correctly 14 out of 15 normal samples and 96.6% accuracy in characterizing correctly 29 out of 30 hyperplasia ultrasound ROIs, thus providing an overall accuracy of 95.5%, at the 1<sup>st</sup> level of the hierarchical classification tree. This accuracy is higher than the corresponding of the QB classifier, which however is easier to train and faster to implement.

BRNN Classifier	Normal	Hyperplasia	Accuracy (%)
Normal	14	1	93.3
Hyperplasia	1	29	96.6
Overall Accuracy			95.5

Table 2. Classification accuracies achieved by the BRNN for discriminating normal from hyperplasia images.

Figure 4 shows a 2-dimensional scatter diagram and decision boundary of the BRNN. We can see that the decision boundary is of quadratic form.

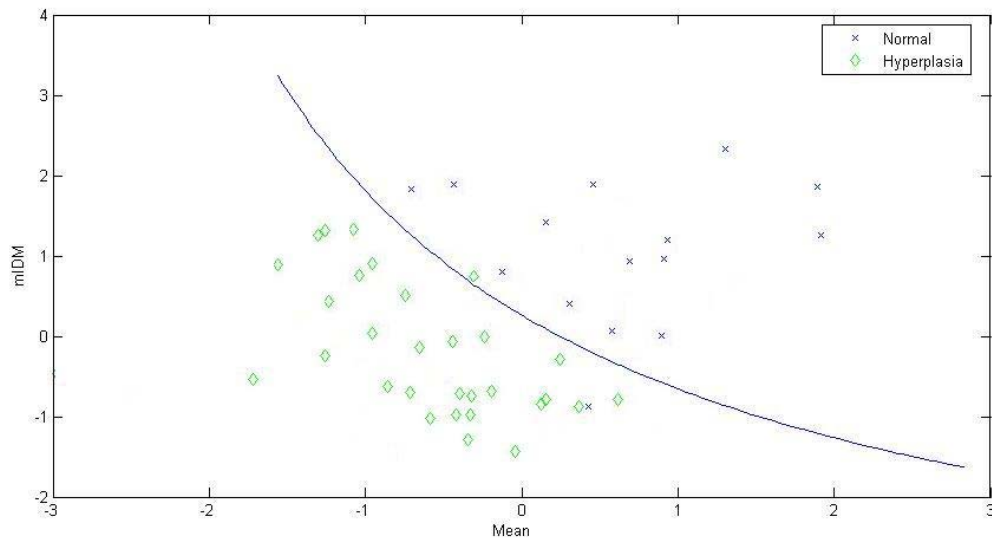


Figure 4. BRNN scatter diagram and decision boundary at the 1<sup>st</sup> level of the hierarchical tree structure.

Table 3 shows the accuracies for the Bayesian Quadratic at the 2<sup>nd</sup> classification level, for best feature vector mean correlation, entropy, variance and sum of average all calculated from the co-occurrence matrix. The classifier discriminated correctly 14/15 (93.3%) Mild and 13/15 (86.6) Medium-High hyperplasia samples, thus providing an overall accuracy of 90%, at the 2<sup>nd</sup> level of the hierarchical classification tree.

QB Classifier	Mild H/a	Medium-High H/a	Accuracy (%)
Mild H/a	14	1	93.3
Medium-High H/a	2	13	86.6
Overall Accuracy			90

Table 3. Classification accuracies achieved by the QB for discriminating mild from medium/high hyperplasia images.

Table 4 shows the results for the BRNN with best feature vector sum of entropy, entropy from the co-occurrence matrix and long run emphasis from the run-length matrix. The classifier discriminated correctly 14/15 (93.3%) Mild and 14/15 (93.3) Medium-High hyperplasia samples, thus providing an overall accuracy of 93.3%, at the 2<sup>nd</sup> level of the hierarchical classification tree. Again, the BRNN classifier achieved higher accuracies than the QB classifier.

BRNN Classifier	Mild H/a	Medium-High H/a	Accuracy (%)
Mild H/a	14	1	93.3
Medium-High H/a	1	14	93.3
Overall Accuracy			93.3

Table 4. Best accuracies achieved by the BRNN classifier for discriminating Mild from Medium-High hyperplasia samples.

Figure 5 shows a 3-dimensional scatter diagram and decision boundary of the BRNN.

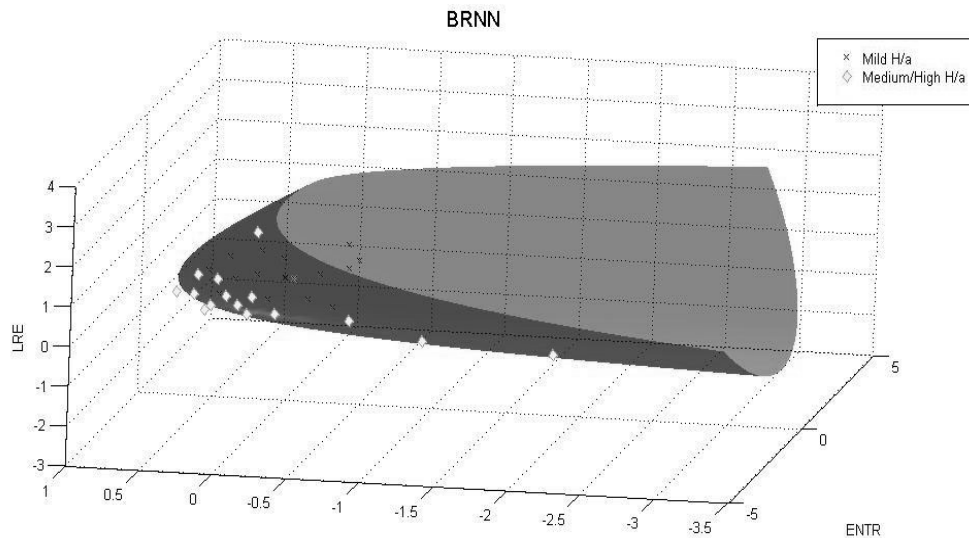


Figure 5. BRNN scatter diagram and decision boundary at the 1<sup>st</sup> level of the hierarchical tree structure.

For both the BRNN networks, at the two levels of the hierarchical tree structure, two layers were used with ten neurons in the first layer and five neurons in the second layer. As we can see from the tables the Bayesian Quadratic classifier achieved lower overall accuracies from the BRNN. The robustness of BRNN is due to its mathematical background. Due to the lack of a large database of images, the BRNN tries to generalize well and avoid over-fitting and this is done by keeping both the sum squares of errors (SSE) and sum square of weight (SSW)<sup>[10]</sup> values as low as possible. This leads to a quite high overall accuracy and also to a better response to novel inputs.

## REFERENCES

- [1] National Cancer Institute – [www.nc.nih.org](http://www.nc.nih.org)
- [2] American Cancer Society – [www.cancer.org](http://www.cancer.org)
- [3] Anagnostou T., Remzi M., Lykourinas M., Djavan B. (2003), “Artificial Neural Networks for decision-making in urologic oncology”, *European Urology*, Vol. 43, pp 593-603.
- [4] Batuello J.T., Gamito E.J., Crawford E.D., Han M., Partin A.W., McLeod D.G., O’Donnel C. (2001), “Artificial Neural Network model for the assessment of lymph node spread in patients with clinically localized prostate cancer”, *Adult Urology*, Vol. 57(3) ,pp.481-485.
- [5] Finne P., Finne R., Auvinen A., Juusela H., Aro J., Maatanen L., Hakama M., Rannikko S., Tammela T.L.J., Stenman U. (2000), “Predicting the outcome of prostate biopsy in screen-positive men by multilayer perceptron network”, *Adult Urology*, Vol 56, pp. 418-422.
- [6] Haralick, R.M., Shanmugam, K., Dinstein, I. (1973), “Textural features for image analysis”, *IEEE Trans. Syst. Man. Cybern.* Vol. 3, pp. 610-621.
- [7] Galloway, M.M, (1975), “Texture analysis using gray level run lengths”, *Comput. Graphics Image Process.* Vol. 4, pp. 172-179.
- [8] Theodoridis, S., Koutroumbas, K., *Pattern Recognition. Academic Press*, New York, 1999.
- [9] Tou, J.T., Gonzalez, R.C. (1974), *Pattern Recognition Principles*. Addison-Wesley Publishing Company, Massachusetts.
- [10] Hagan M.T, Demuth H.B. and Beale M. (1996), *Neural Network Design* , Boston: PWS Publishing Co.
- [11] MacKay D. J. C. (1992), “Bayesian Interpolation”, *Neural Computation*, Vol. 4, pp. 415-447.