2005 WSEAS Int. Conf. on DYNAMICAL SYSTEMS and CONTROL, Venice, Italy, November 2-4, 2005 (pp56-61)

Automatic Detection of Targets using Gabor Filters and Neural Networks

MIRCEA VIRGIL POPA
PAUL BECHET
ŞTEFAN DEMETER
Technical Sciences Department
Land Forces Academy "Nicolae Balcescu"
Sibiu, Revolutiei 3-5, 550170
ROMANIA

Abstract: This paper presents a method for detecting artificial objects (military targets) by combining the outputs of a set of log-Gabor filters using artificial neural networks. The system is able to automatically process frequency content of an image and to extract features –visual patterns- that have a high degree of correlation in the statistical structure across frequency bands. We present some results using a filtering technique for the automatically learned partitioning of patterns in a digital image.

Key-Words: - Image processing, Object/target detection, Gabor filter, Neural network, Filter selection, Cognitive processing

1 Introduction

Real-world images contain characteristic statistical regularities that we might consider when analyzing the information content. An image processing technique used for modeling the human vision system (HVS) is to decompose the image into a set of channels whose outputs are a function of thresholds measured in psychophysical experiments. The spatial properties of receptive fields in visual cortex suggest that it can be described adequately by a sinusoidal wave in a two-dimensional Gaussian envelope. These Gabor functions exhibits the property that it minimizes the uncertainty in both space and spatial frequency.

The Gabor functions represent an attractive mathematical tool for modeling HVS especially due to psychophysical plausibility. The technique used in many experiments is to decompose the image into a set of visual channels, each associated with a particular receptive field profile. It is likely to be a set of features, which extend across different frequency bands.

In performing detection and discrimination, we intend to maximize the response of neurons associated to a particular visual channel (bank-filter) while minimizing the response of neurons not associated to the visual channel. We use a neural network to combine the responses of log-Gabor filter banks so that to extract specific visual patterns. Log-Gabor filters were adopted as an appropriate method to construct filters of arbitrary bandwidth and a uniform coverage of spectrum.

2 Bank of Log-Gabor Filters

In images with complex backgrounds it is difficult to extract the objects and their characteristics and this fact forced us to find the algorithm so that to extract some specific characteristics of the images, allowing the detection simplification and the identification of the target-objects that are presented to some human operators. This filtering technique can be associated to the local analysis of the spatial frequencies adequate to the biological visual systems and also allow some new directions that are to be purchased.

The Gabor functions represent an attractive mathematical tool for modeling the human vision system (HVS) especially due to psychophysical plausibility. The technique used is to decompose the image into a set of visual channels, each associated with a particular receptive field profile and it is likely to find out a set of features, which extend across different frequency bands.

Log Gabor functions have a transfer function that is Gaussian when viewed on the log-frequency scale. The question is, what is the appropriate bandwidth to use with a log Gabor function? We considered two measures of 'width': first, the width required to represent 99% of the absolute area of the filter and second the 2nd Moment about the middle of the filter with respect to absolute value of the filter (both measures are minimized when the filter bandwidth is about two octaves). This matches well with psychophysical measurements of the bandwidth of cells. The spatial width at this bandwidth corresponds roughly to the width of a standard

Gabor function having a bandwidth of one octave. Thus, it seems log-Gabor functions are good at obtaining frequency information.

We use a scheme for filtering the input 2D images, for automatically learned partitioning of invariant features that have a degree of alignment in statistical structure across frequency bands. The clumps of energy in the Fourier spectrum of the image are distributed in a set of oriented spatial-frequency channels. This step implies the selection of a subset of active filters from a filter-bank of log-Gabor functions with 4 scales and 12 orientations (6 due to conjugate symmetry for on half of frequency plane).

Log-Gabor filters determine a Gaussian in the spatial frequency domain around a specific center frequency (r_0, θ_0). This can be represented in frequency domain as a sum of the even-symmetric and i-times the odd-symmetric log-Gabor filter [1]:

$$\phi(r_0,\theta_0) = \exp\left(- \frac{\left(\log\left(\frac{r}{r_0}\right)\right)^2}{2\left(\log\left(\frac{\sigma_r}{r_0}\right)\right)^2} \right) \exp\left(- \frac{(\theta-\theta_0)^2}{2\sigma_\theta^2} \right) \tag{1}$$

where: θ_0 - is the orientation angle of the filter; r_0 - is the central radial frequency; σ_θ , σ_r - are the angular and radial sigmas of Gaussian.

In the frequency domain the even symmetric filter is represented by two real-valued log-Gaussian 'bumps' symmetrically placed on each side of the origin. The odd-symmetric filter is represented by two imaginary valued log-Gaussian 'bumps' anti-symmetrically placed on each side of the origin. Combining the convolution of the even and odd symmetric filters into the one operation and due to the linearity of the Fourier Transform it can be performed the following algorithm: multiply the FFT of the odd-symmetric filter by i (to make it real valued) and add it to the FFT of the even symmetric filter. The anti-symmetric 'bump' from the odd-symmetric filter will cancel out the corresponding symmetric bump from the even-symmetric filter. This leaves a single 'bump' (multiplied by 2) on the positive side of the frequency spectrum. Thus if we construct a filter in the frequency domain with a single log-Gabor 'bump' on the positive side of the frequency spectrum we can consider this filter to be the sum of the FFTs of the even and odd symmetric filters (with the odd symmetric filter multiplied by i). If we perform the convolution by multiplying this frequency domain filter by the FFT of the image and take the inverse FFT we end up with the even-symmetric convolution residing in the real part of the result and the odd-symmetric convolution residing in the imaginary part. Results of applying this method [2] in Matlab, are shown in figure 1.

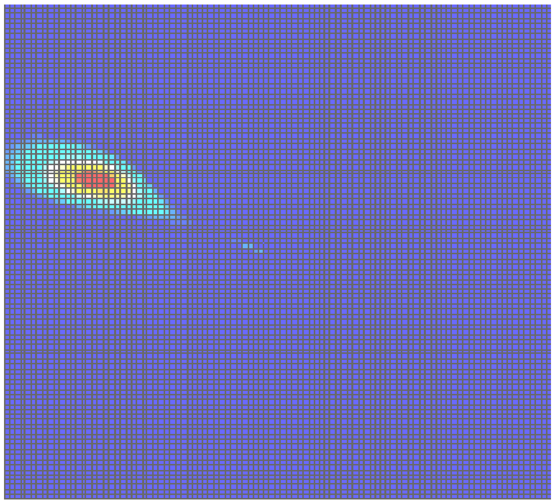


Fig. 1 Representation of the log-Gabor filter characteristics

The complex values of the convolution result simultaneously encoding the magnitude and phase response of the quadrature filters as a complex image. The local energy of the analyzed image using a log-Gabor filter is [3]:

$$E(x,y) = \left(O_e^2(x,y) + O_o^2(x,y)\right) \tag{2}$$

where $O_e(x,y)$ and $O_o(x,y)$ are the results of image convolution with even and odd-symmetric log-Gabor filters. The real part of (1) multiplied by frequency transformed image and after transforming back into spatial domain, the results of applying the filter pair are extracted as the real part for $O_e(x,y)$ and imaginary part of $O_e(x,y)$ [4].

The results of applying the bank of 24 log-Gabor filters for 4 scales and 6 orientations (12 due to conjugate symmetry for on half of frequency plane) are shown in figure 2.

3 The selection of activated filters

In performing detection and discrimination, we intend to maximize the response of neurons associated to a particular visual channel (bank-filter) while minimizing the response of neurons not associated to the visual channel. Log-Gabor filters were adopted as an appropriate method to construct filters of arbitrary bandwidth and a uniform coverage of spectrum.

We use a scheme for filtering the input 2D images, for automatically learned partitioning of

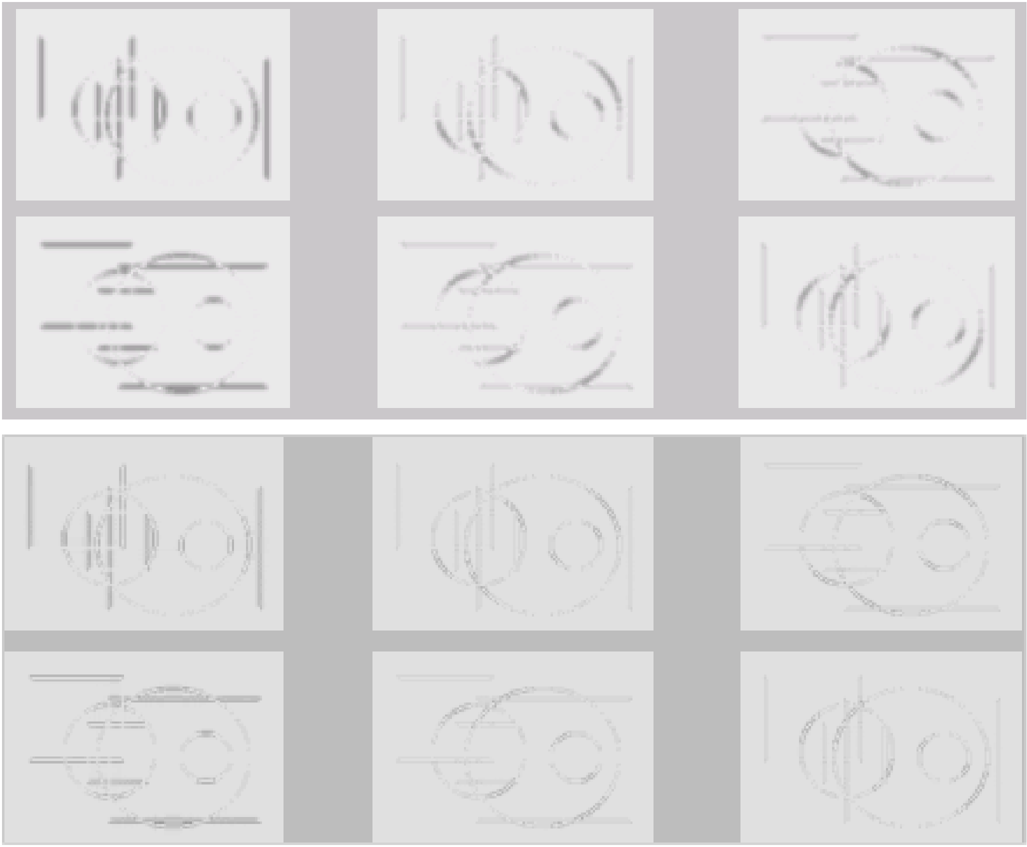


Fig. 2 Outputs of log-Gabor filters (selection)

invariant features that have a degree of alignment in statistical structure across frequency bands. The clumps of energy in the Fourier spectrum of the image are distributed in a set of oriented spatial-frequency channels. This step implies the selection of a subset of active filters from a filter-bank of log-Gabor functions with 4 scales and 12 orientations (6 due to conjugate symmetry for on half of frequency plane).

In order to represent the image into its most important components, strongly responding filters must be selected for the input image. This set of filters must be sensitive to visual patterns in the image, in fact to the clumps of energy of Fourier spectrum.Having the set of 24 images representing the outputs of filter bank, it is necessary to select and combine this set so that to approximate the visual pattern in the input image. In [1], first for any two activated filters, their responses are compared based on the distance – a β norm between their statistical structure, computed over those pixels which form local energy peaks of the filtered response. Second, in next stage, is performed clustering on the basis of the activated filters to highlight scale and orientation of responses.

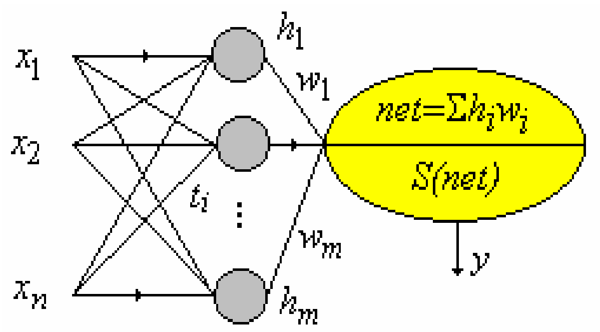


Fig. 3 Network used for solving equation (4)

For solving this problem, given the 24 images $G_{i,j}$ (i =1…4, for scales and j =1…6 for orientations) it is necessary to solve the equation:

$$\sum_{k=1}^{24} \alpha_k \cdot G_{i,j} = T \tag{3}$$

where T is the desired visual pattern (target) and $\alpha_k \in R^+$ are the weights associated in order to select the active set of filters. Implementation have been made using Levenberg-Marquant algorithm for minimization the objective function F:

$$F = T - \sum_{k=1}^{24} \alpha_k \cdot G_{i,j} \tag{4}$$

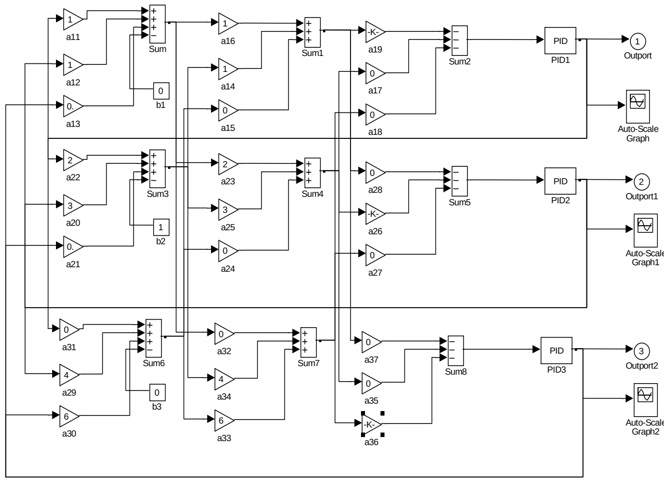


Fig. 4 Example of network solving (6)

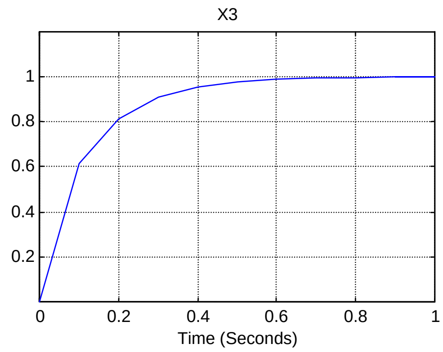


Fig. 5 Output of network

and by using a neural network, so that to calculate the necessary weights α_k (figure 3).

Let us assume (3) as a set of algebric equations written in in scalar form as:

$$\sum_i \sum_j g_{ij} \cdot \alpha_j = t_i \qquad (5)$$

where α is the k-dimensional unknown vector (k=24), ti is the (m x n) dimensional target martix and g = [gij]k is the [m x n x k] matrix coefficients. The problem can be formulated as folows:

find the vector $x \in \Re^k$

which minimizes the objective function (6)

$$E(\mathbf{x}) = \|\mathbf{g} \cdot \boldsymbol{\alpha} - \mathbf{t}\|_p = \|\mathbf{r}(x)\|_p, \quad \mathrm{p} \geq 1$$

where r is the residual error vector for a given actual x. For p=1 the problem is referred to as an L_1 (least absolute deviations) problem and for p=2 the case of L_2 (linear least-squares) problem.

The use of L_1 norm can provide a useful alternative to least squares or Chebyshev norm solution of the system of linear/nonlinear equations because the least absolute deviation solutions have certain properties:

- a L_1 solution of an overdeterminated system of linear equations always exists although the solution is not necessarily unique;
- L_1 solutions are robust (the solution is resistant to some large data changes (or noise);
- L_1 problems are equivalent to linear programming problems.

The architecture resembles a version of perceptron and the circuit consists of three layers to neurons connected in a feedforward mode (figure 4, 5).

In the learning stage, combining the input images according with the appropriate selection of filters (weights) results a filter bank personalized for a specific visual pattern. For example, having a synthetic input image containing the target visual pattern, it is decomposed by the log-Gabor filters, then the network calculates the associated weights and thus the active filter bank for that specific target and the reconstructed image of object (figure 6, 7).

After training the network with the set of different visual patterns and building the database of associated filter banks, the processing scheme can be applied for obtaining visual distinctness in digital images.

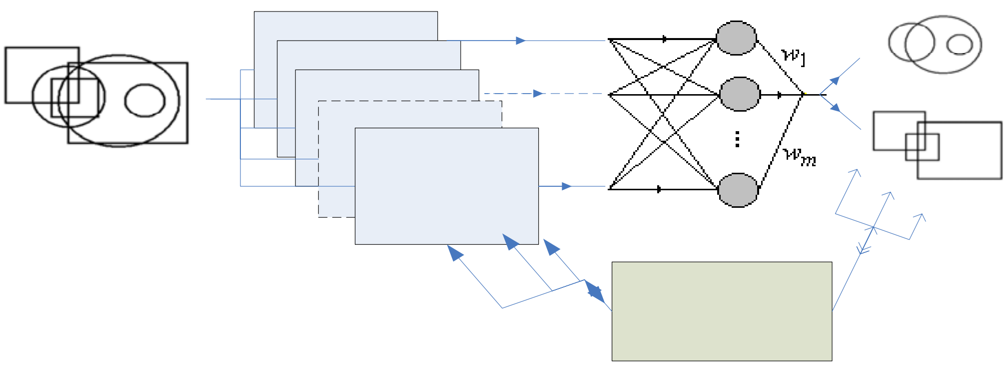
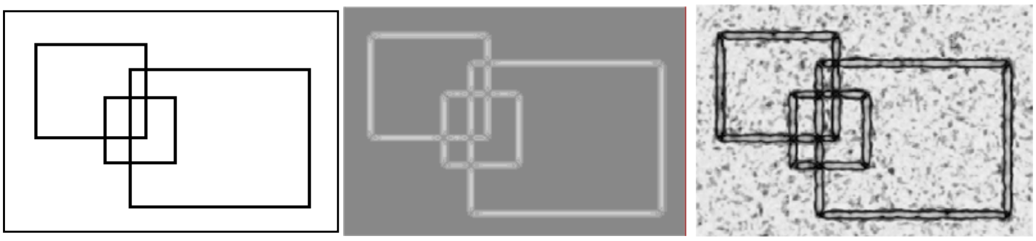


Fig. 6 Automated segregation of visual patterns



a b c

Fig. 7 "Target" image (a), reconstructed image (b) and results for the same target with 30% impulsive noise (c)

This algorithm has been applied to solve different tasks: extracting visual information regarding targets (artificial objects) from images with complex natural backgrounds (figures 8, 9); extracting target (complex irregular objects) from images with complex artificial backgrounds (figure 10).

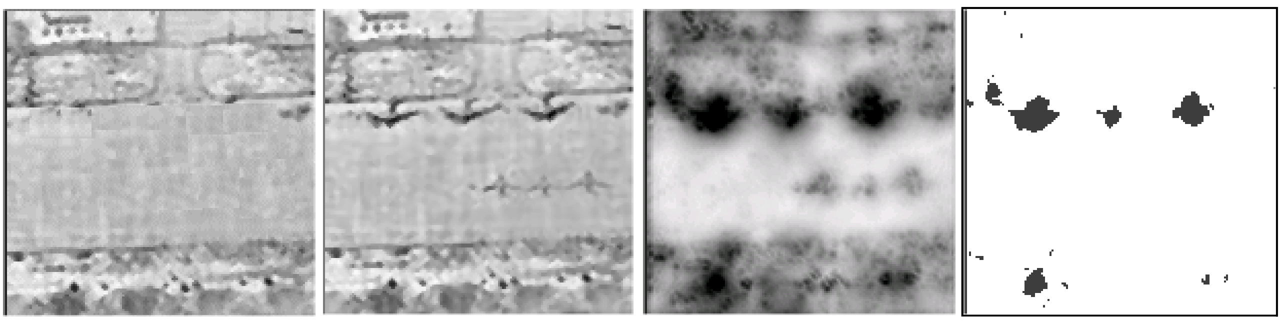
4 Conclusions and further work

Spectral decomposition is a way to describe texture in image processing. The spectral decomposition using Gabor filtering has often been justified by the fact that it provides a good approximation of the natural processes in the primary visual cortex [8]. The texture feature images are the magnitudes of the responses of the Gabor filters. The latter encodes the energy content and is independent of the position within the texture. Each vector is further normalized to be a unit vector to emphasize the texture structure information and reduces the dependence of the responses on lighting.

The Fourier approach has some strengths and weaknesses. First it has a good deal of plausibility from low-level image processing considerations; second, is its formal mathematical status. A more important advantage is that power spectra appear to solve the problem of shape equivalence over the similarity transformations offers invariance over differences of luminance contrast and potentially can represent texture information. Main drawback is that Fourier analysis represents usually an entire image and not individual objects (parts) and also it does not seem to be good predictors of perceived shape but it may provide an alternative for encoding positional information.

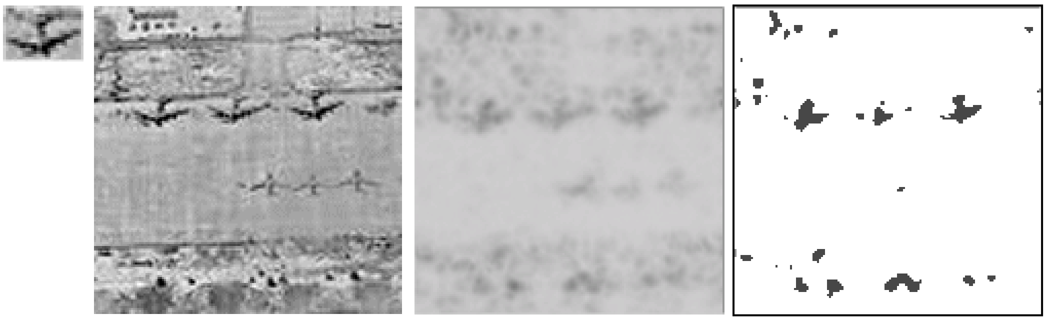
It is useful to think of features in terms of their Fourier components, rather than in terms of intensity gradients. This allows us to describe a wide range of feature types within the framework of a single model. Features are assumed to lie at points of high phase congruency, and the angle at which the congruency occurs describes the feature type. Experiments indicate that images contain feature types of all phase angles, with a broad distribution. Accordingly it can be concluded that gradient based operators, which look for points of maximum intensity gradient, will fail to correctly detect and localize a large proportion of features within images. The automatic detection scheme presented has been successfully applied to various synthetic images. The necessary image data and information about the objects (targets) can be recurrently used to train the network and to keep up-to-date the information in database of target-objects.

The automatic detection scheme presented has been successfully applied to various synthetic images. The necessary image data and information about the objects (targets) can be recurrently used to train the network and to keep up-to-date the information in database of target-objects. The



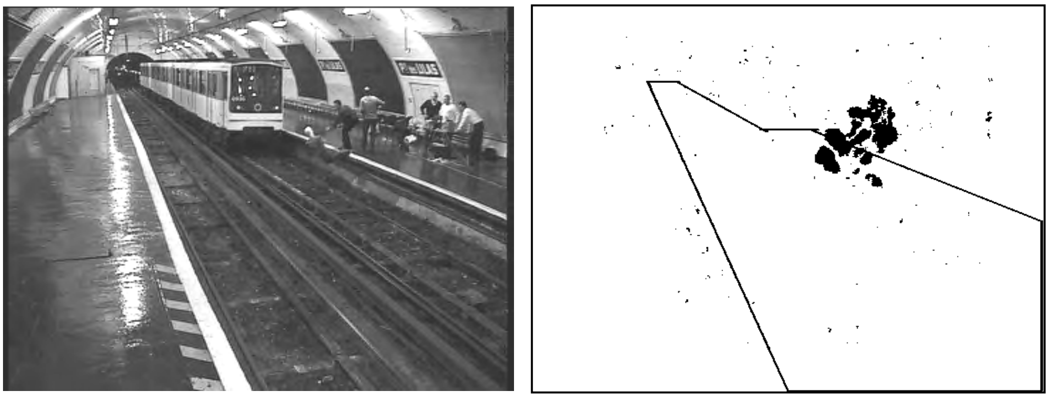
a b c d

Fig. 8 Image used for selection of active filters (a), input image (b), output image (c) and regions of interest for decision process (d)



a b c d

Fig. 9 Image used for selection of active filters (a), input image (b), output image (c) and regions of interest for decision process (d)



a b

Fig. 10 Extracting abnormal human behavior from context (with courtesy of Visionwave, AVSS 2005 Challenge)

objectives set have been met:

- selecting the appropriate filter bank the computational costs can be reduced;
- detection-recognition time can be reduced.

The results are close to those related by Garcia and Vidal [1]. Because of the differences between the statistical structure across scales and orientations of targets and backgrounds, the distinctness of an artificial object can be determined. This paper presents only a part of a model for cognitive processing of visual information. Further work can be carried out to integrate this scheme into an invariant system for detection and recognition of targets in action and for flexibility and adaptability to the presence of noise and a variety of backgrounds.

References:

[1] Field, D.J., *Relations between the Statistics of Natural Images and the Response Properties of Cortical Cells*, J. Optical Soc. A, vol.4, no. 12, pp.2379-2394. 1987.

[2] Morrone M. C., Owens, R. A., *Feature detection from local energy*, Pattern Recognition Letters, 6:303–313, 1987.

[3] Rasmunssen, J., *Skills, rules and Knowledge, signals, signs and symbols and other distinctions in human performance models*, IEEE Trans. Sys. Man. Cybernetics, 1983.

[4] Hoffma, N. J., Skrzypek J., "*Visual Recognition of Script Characters and Neural Network Architectures", Neural Networks: Advances and Applications*, Elsevier, 1991.

[5] Zlate M., *Psihologia mecanismelor cognitive*, Polirom Iaşi, 1999.

[6] Miclea M., *Psihologie cognitivă*, Ed. Polirom Iaşi, Romania, 1999

[7] Kovesi P. D., *Invariant Measures of Image Features From Phase Information*. PhD thesis, The University of Western Australia, May 1996, pk@csse.uwa.edu.au

[8] Palm C., D. Keysers, *Gabor filtering of complex hue/saturation images for color texture classification*, Int. Conf. on Computer Vision, Patt. Recogn. and Im. Proc. vol. 2, pp. 45–49, 2000.

[9] J. A. Garcia, X. R. Fdez-Vidal, R. Rodriguez-Sanchez, *Visual Distinctness Determined by Partially Invariant Features*, RTO MP-45, 1999

[11] B. Robbins, "The detection of 2D image features using local energy", Ph.D. thesis, The University of Western Australia, 1996