HAIS'10
5th International Conference on HYBRID ARTIFICIAL INTELLIGENCE SYSTEMS
23rd - 25th June 2010 - San Sebastián, Spain

An Evolutionary Feature-Based Visual Attention Model Applied to Face Recognition

Roberto A. Vázquez, Beatriz A. Garro and Humberto Sossa

E-mail: hsossa@cic.ipn.mx and ravem@ipn.mx and bgarrol@ipn.mx

CONTENT OF THE PRESENTATION

1. Introduction.

2. Description of the proposed visual attention model:

3. Evolutionary control.

4. Experimental results.

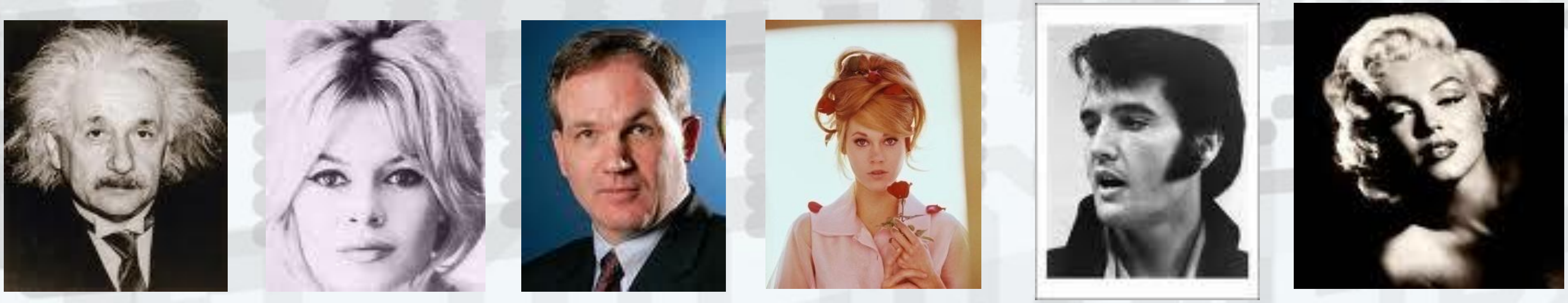
5. Conclusions.

INTRODUCTION

Visual object recognition, particularly face recognition, is a difficult computational problem.

Each face might appear as an infinite number of different 2-D images.

The face could vary relative to the viewer in position, pose, lighting, gesture, aging and background.



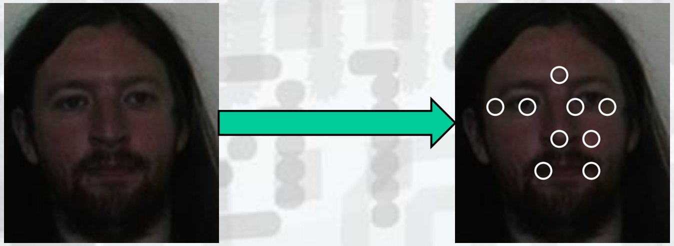
INTRODUCTION

According to: J. H. R. Maunsell and S. Treue (2006). Feature-based attention in visual cortex. Trends in Neurosciences, 29(6), 317-322:

…visual attention is a powerful mechanism that enables perception to focus on a small subset of the information picked up by our eyes.

So, according to: A. A. Salah et al. (2002). A selective attention-based method for visual pattern recognition with application to handwritten digit recognition and face recognition. IEEE T PAMI. 24(3), 420-425.

…by incorporating visual attention into a face recognition system, will allow to use only a small part of the information for efficient object recognition and classification.



INTRODUCTION

Several visual attention models have been proposed in the last years:

L. Itti and C. Koch (2001). Computational modeling of visual attention. Nature Reviews Neuroscience, 2(3), 194-203.

… describe the essential features of computational model of selective attention.

A. Chauvin et al. (2002). Natural scene perception: visual attractors and image processing. In Connectionist Models of Cognition and Perception, 7th Neural Computation and Psychology Workshop, World Scientific Press, pp. 236-245.

… propose a model inspired by the retina and the primary visual cortex cell functionalities.

L. Q. Chen et al. (2003). Image adaptation based on attention model for small form factor devices. In the 9th Inter. Conf. on Multimedia Modeling, pp. 483-490.

… propose an attention model based on visual static features for face and text detection.

INTRODUCTION

…psychophysical studies have demonstrated that feature-based attention improves detection or otherwise enhances behavioral performance across the visual field…

A. F. Rossi and M. A. Paradiso (1995). Feature-specific effects of selective visual attention. Vision Res. 35, 621–634.
M. Saenz et al. (2003). Global feature-based attention for motion and color. Vision Res. 43, 629–637.

…also, recent evidence supports color feature, in particular, showing an advantage in recognition of objects, faces and natural scenes…

J. Martinovic, T. Gruber and M. Muller (2008). Coding of visual object features and feature conjunctions in the human brain. PLoS ONE 3(11)e3781, 1-10 (2008)

GOAL OF THE INVESTIGATION:

We adopt some of these biological hypotheses an propose **an evolutionary visual attention model** that is next applied to the face recognition problem.

The proposed model is composed by three levels:

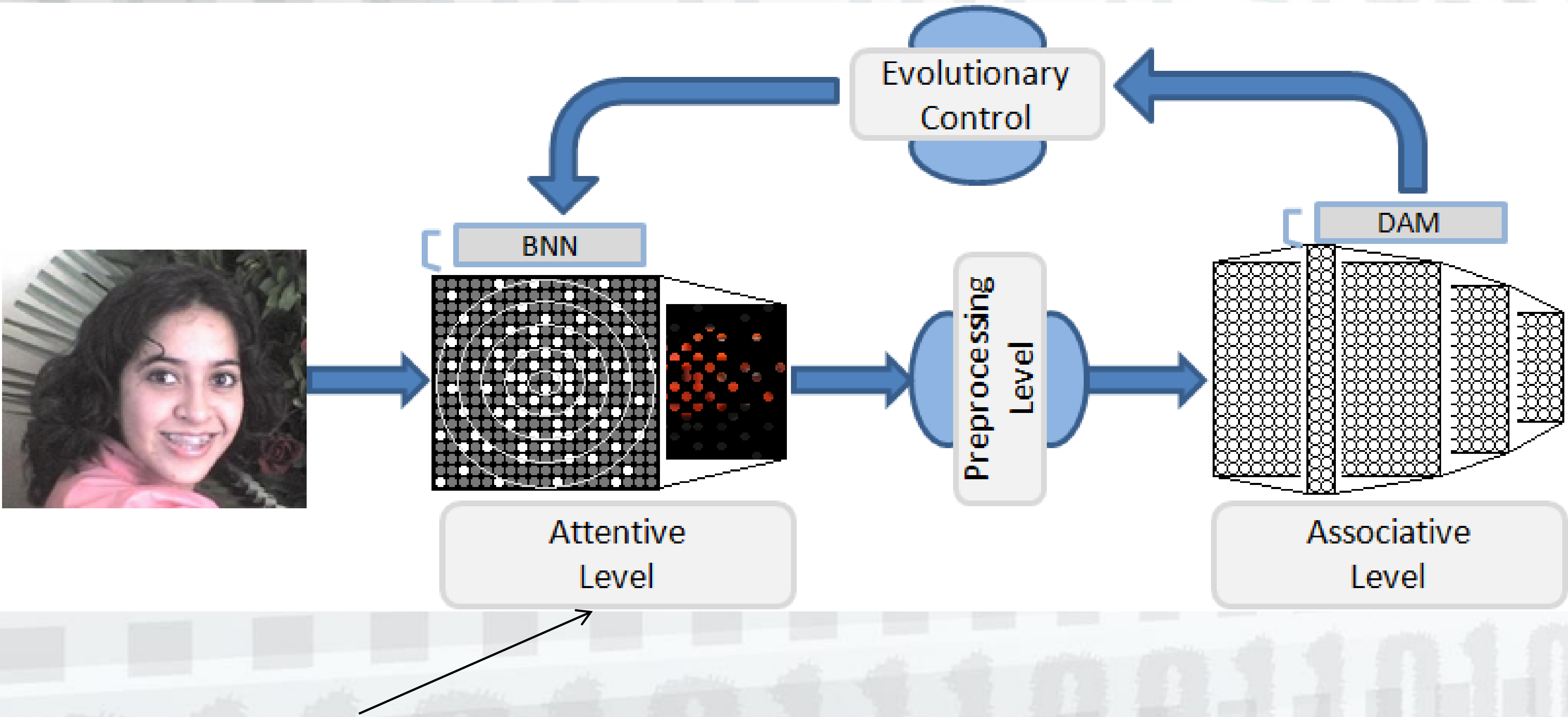
An attentive level.

A preprocessing level.

An associative level.

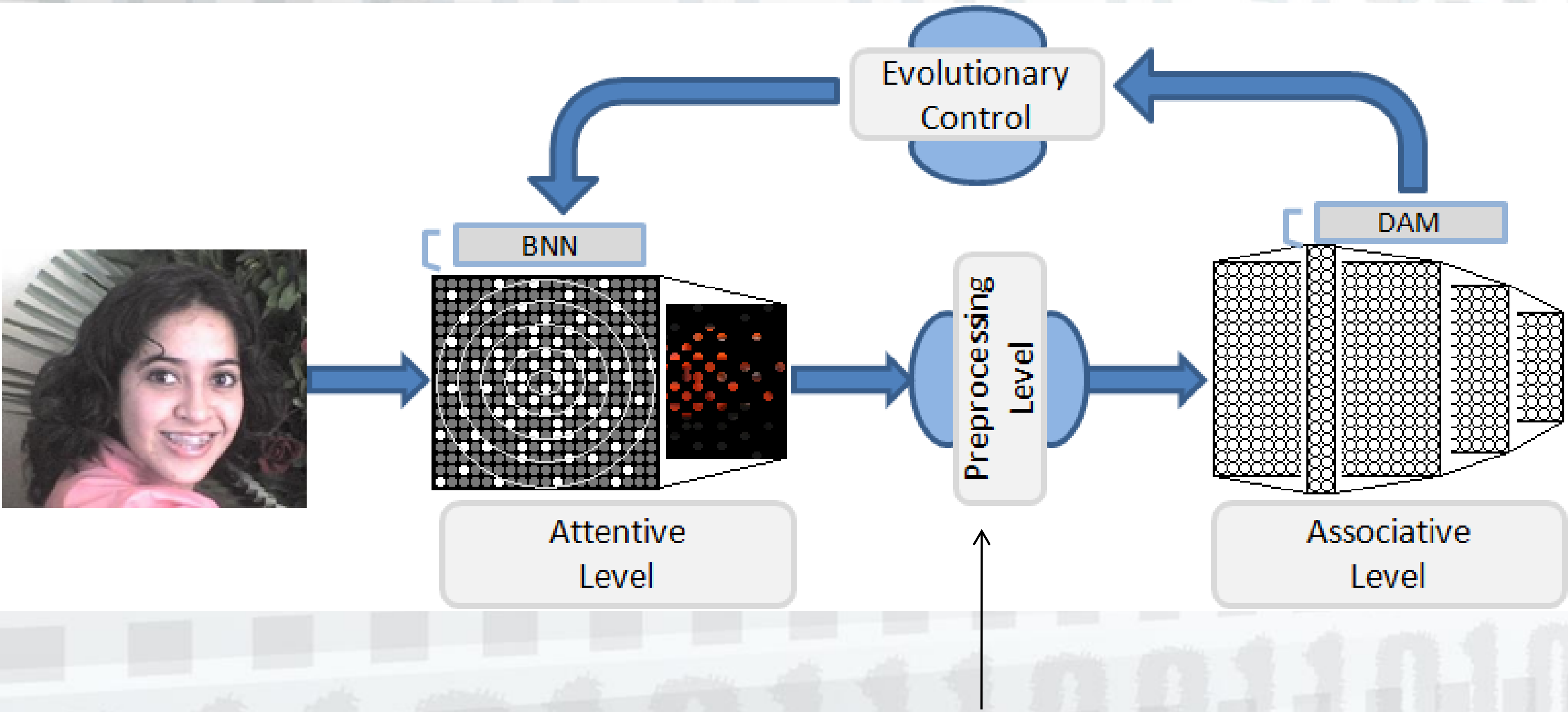
To test the accuracy of the model a benchmark of faces is used.

THE PROPOSED VISUAL ATTENTION MODEL:



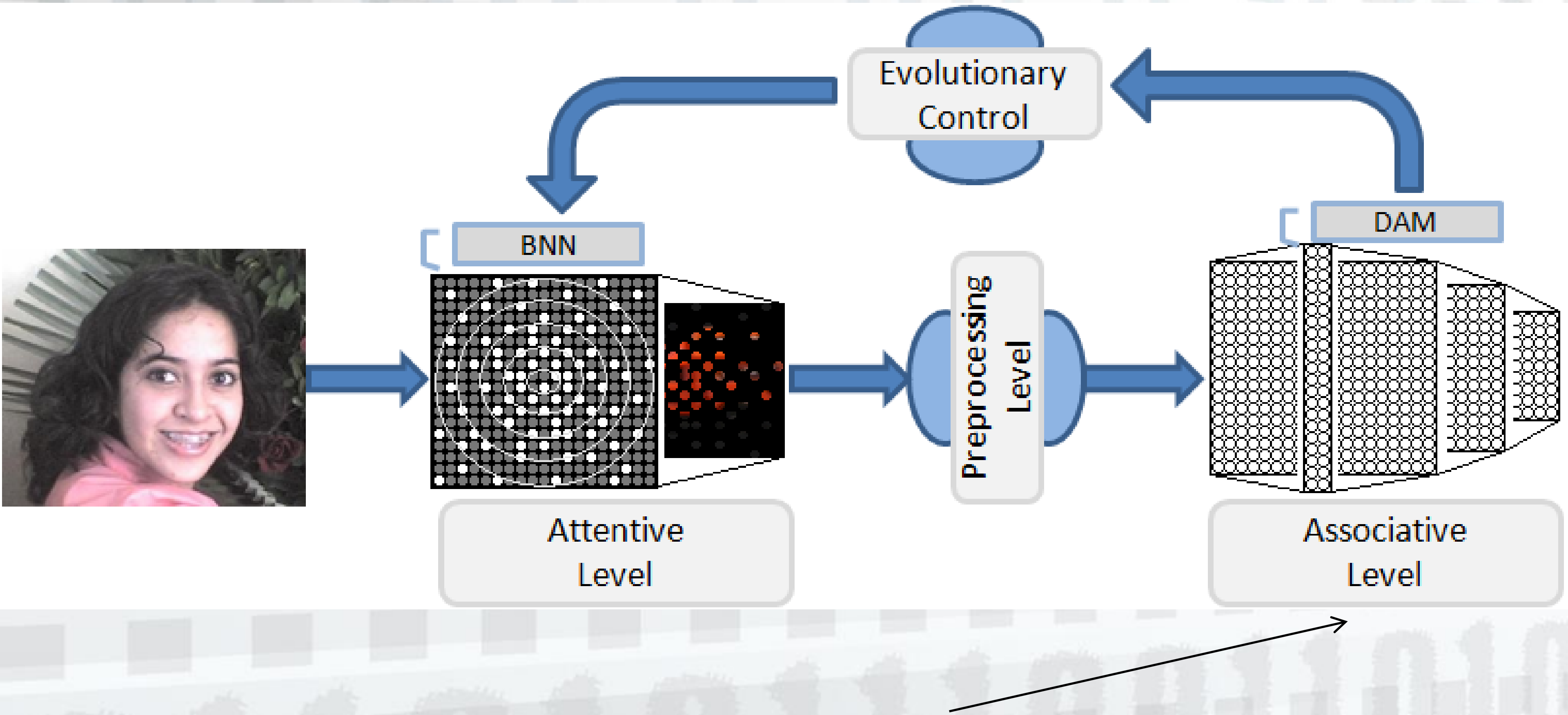
FIRST: The attentive level: determines where to look by means of a retinal ganglion network simulated as a network of bi-stable neurons and controlled by an evolutionary process.

THE PROPOSED VISUAL ATTENTION MODEL:



SECOND: The preprocessing level: analyses and processes the information from the retinal ganglion network.

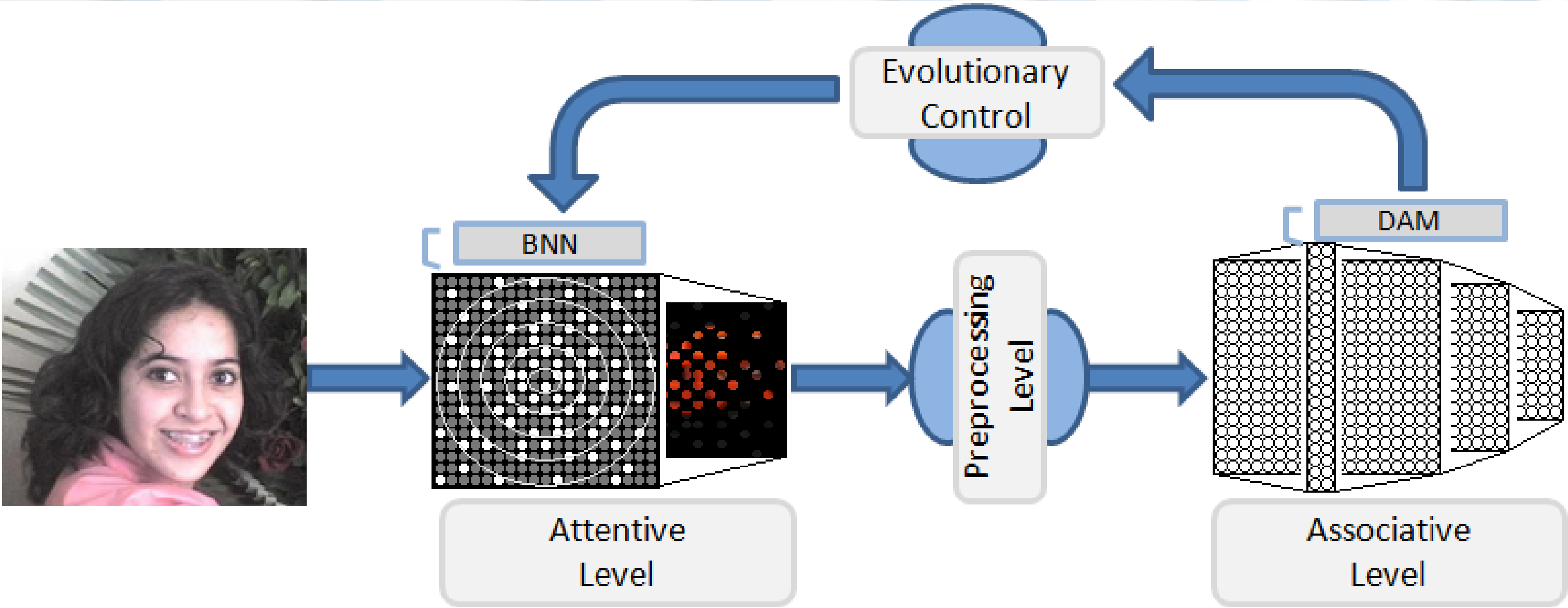
THE PROPOSED VISUAL ATTENTION MODEL:



THIRD: The associative level: uses a <u>neural network to associate</u> the visual stimuli with the face of a particular person.

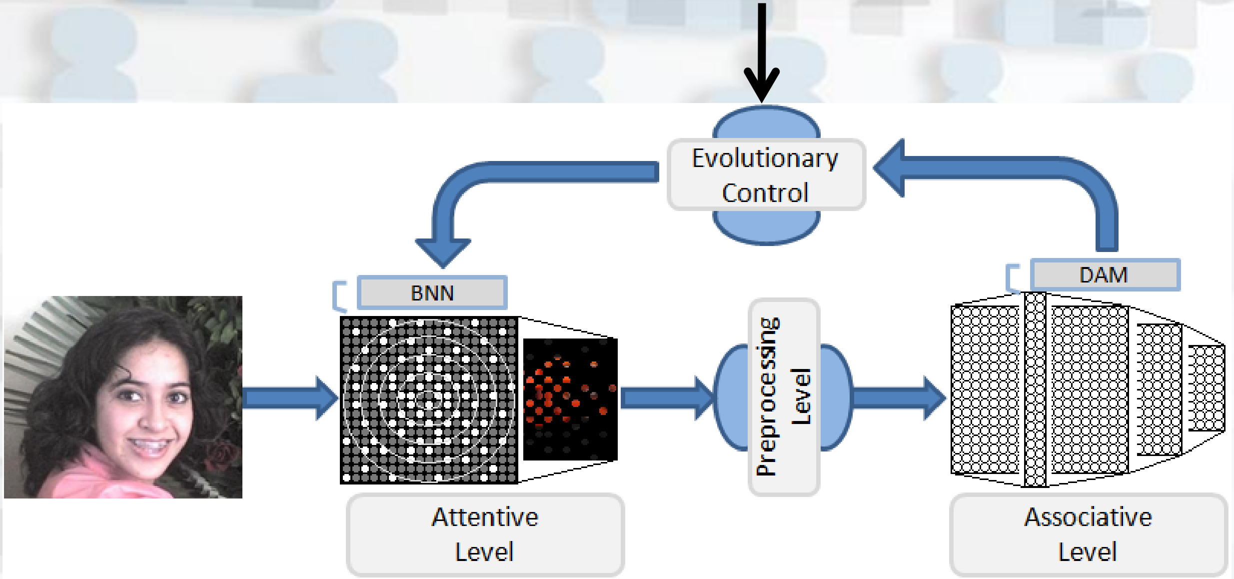
THE PROPOSED VISUAL ATTENTION MODEL:

FINALLY, a feedback between the associative level and the attentive level is used (DE algorithm) to control the attentive level.



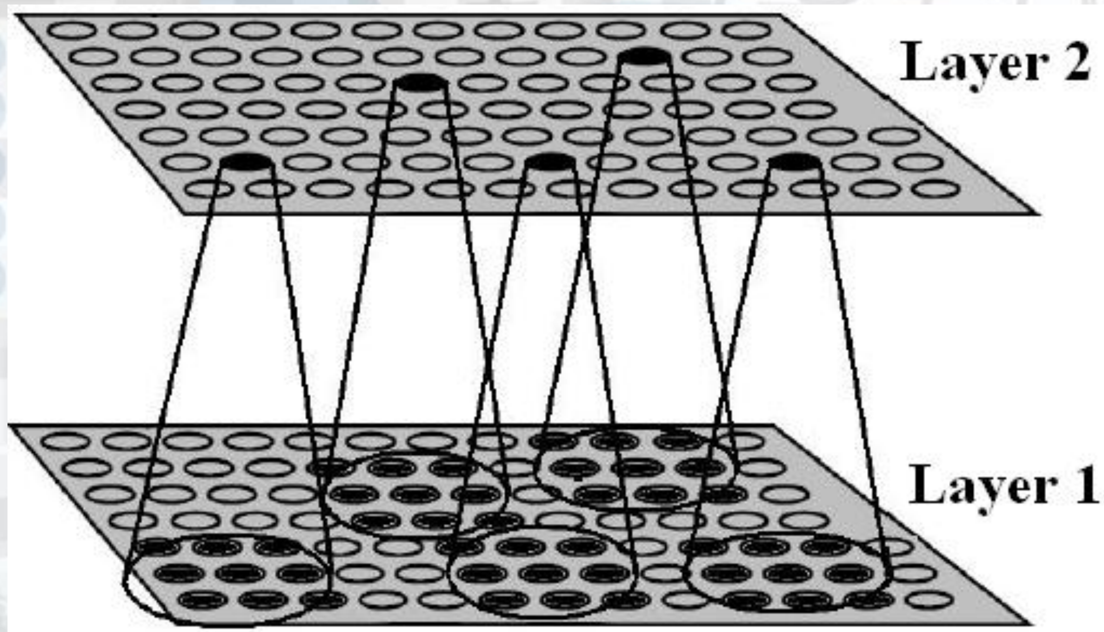
THE PROPOSED VISUAL ATTENTION MODEL:

THE FUNCTION OF THE EC IS A FOLLOWS: The evolutionary control will activate the bi-stable neurons **which maximize** the accuracy of the neural network used in the associative level during the process of recognition.



Therefore, the face recognition problem can be seen as a maximization problem where by means of a DE algorithm it is determined where and what to see in order to maximize the accuracy of a neural network during a face recognition task.

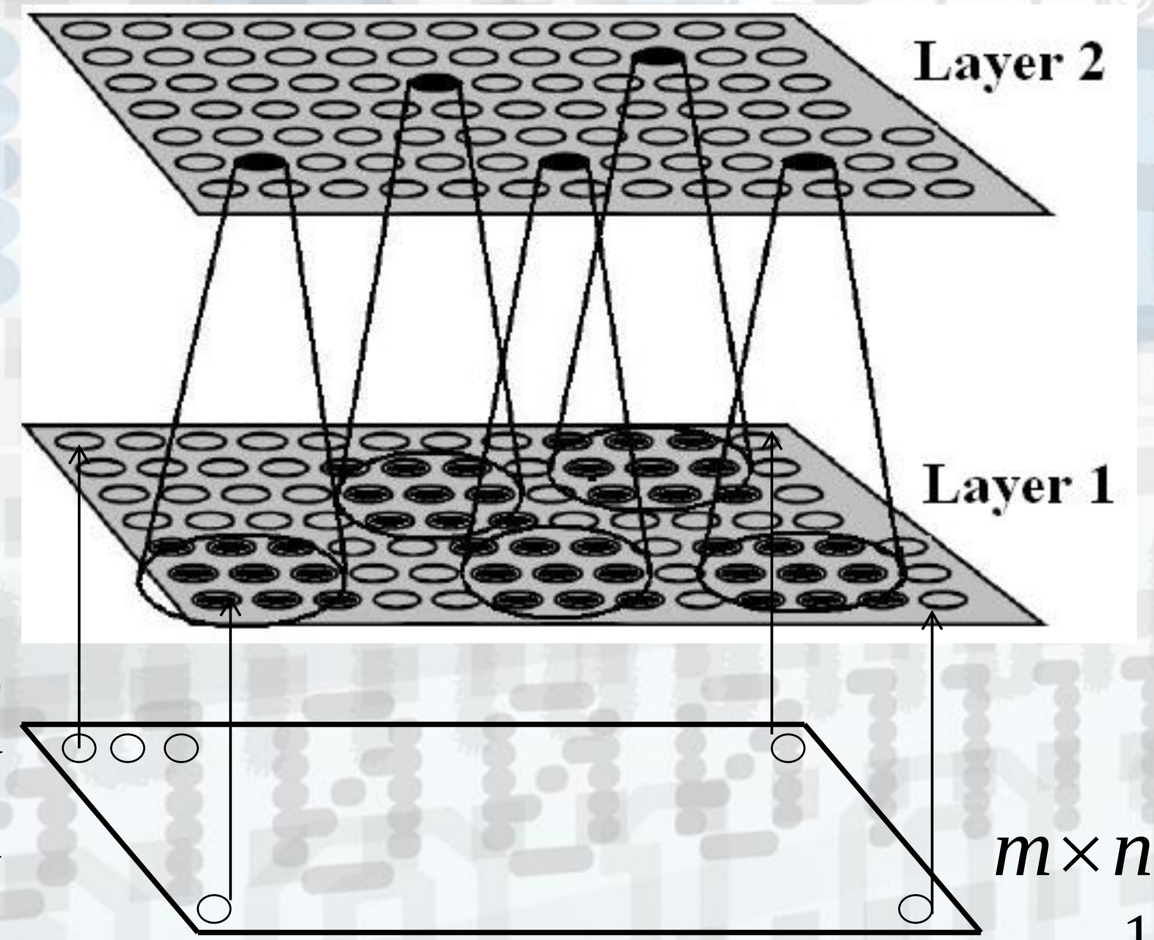
THE ATTENTION LEVEL:



A visual stimulus is represented as a 2D image: $f\left(i, j\right)$

$m \times n$

THE ATTENTION LEVEL:

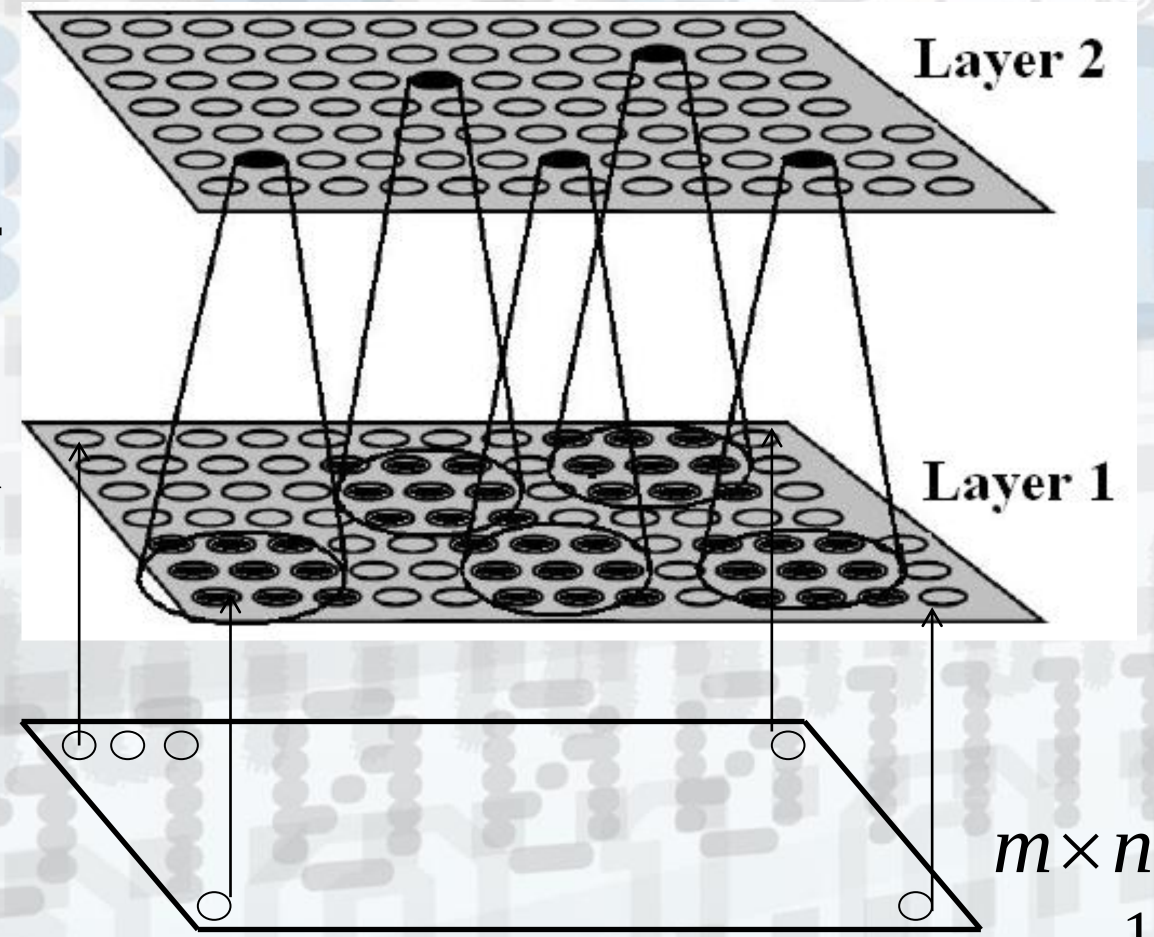


Each pixel from the visual stimulus directly converges to the retinal ganglion (Layer 1):

THE ATTENTION LEVEL:

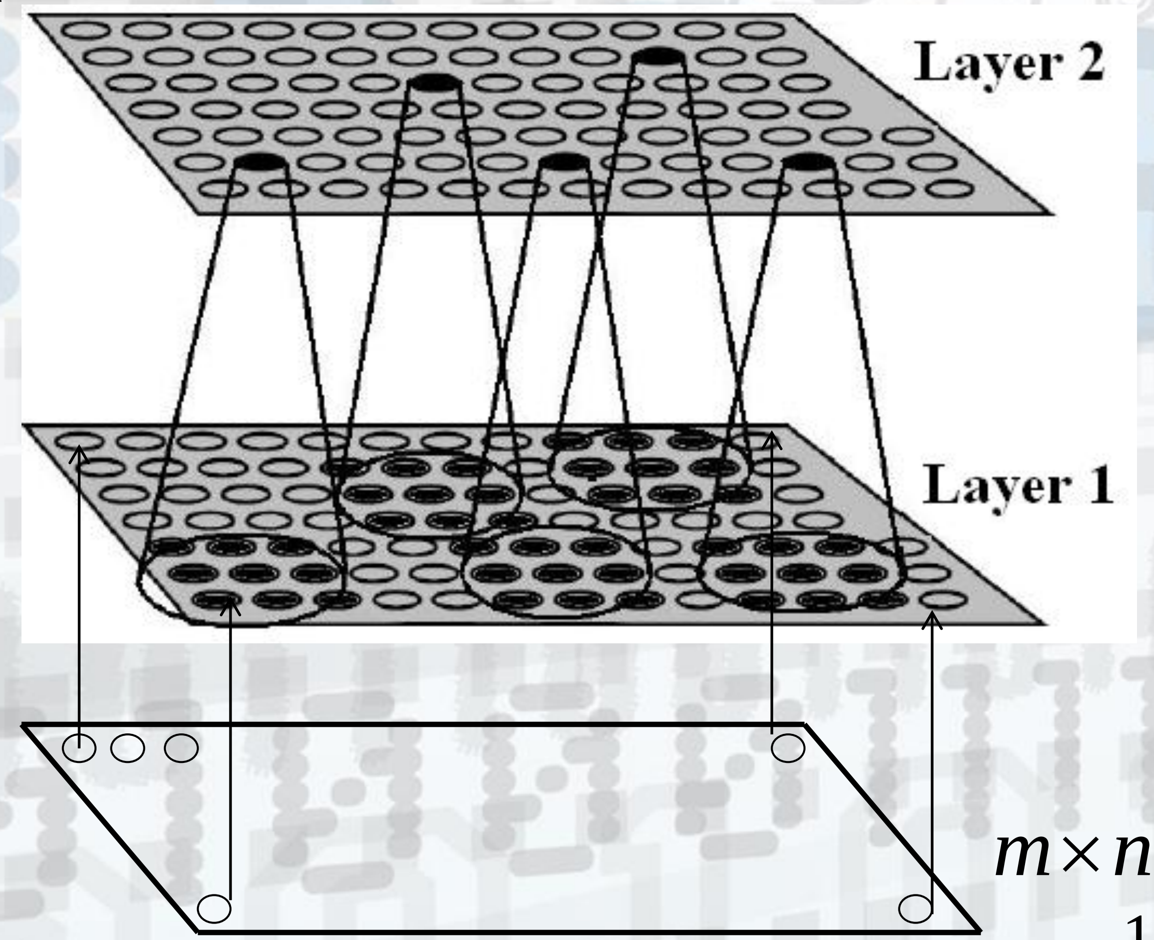
The retinal ganglion is simulated by 2 layer network.

Each layer is composed by $m \times n$ bi-stable neurons.



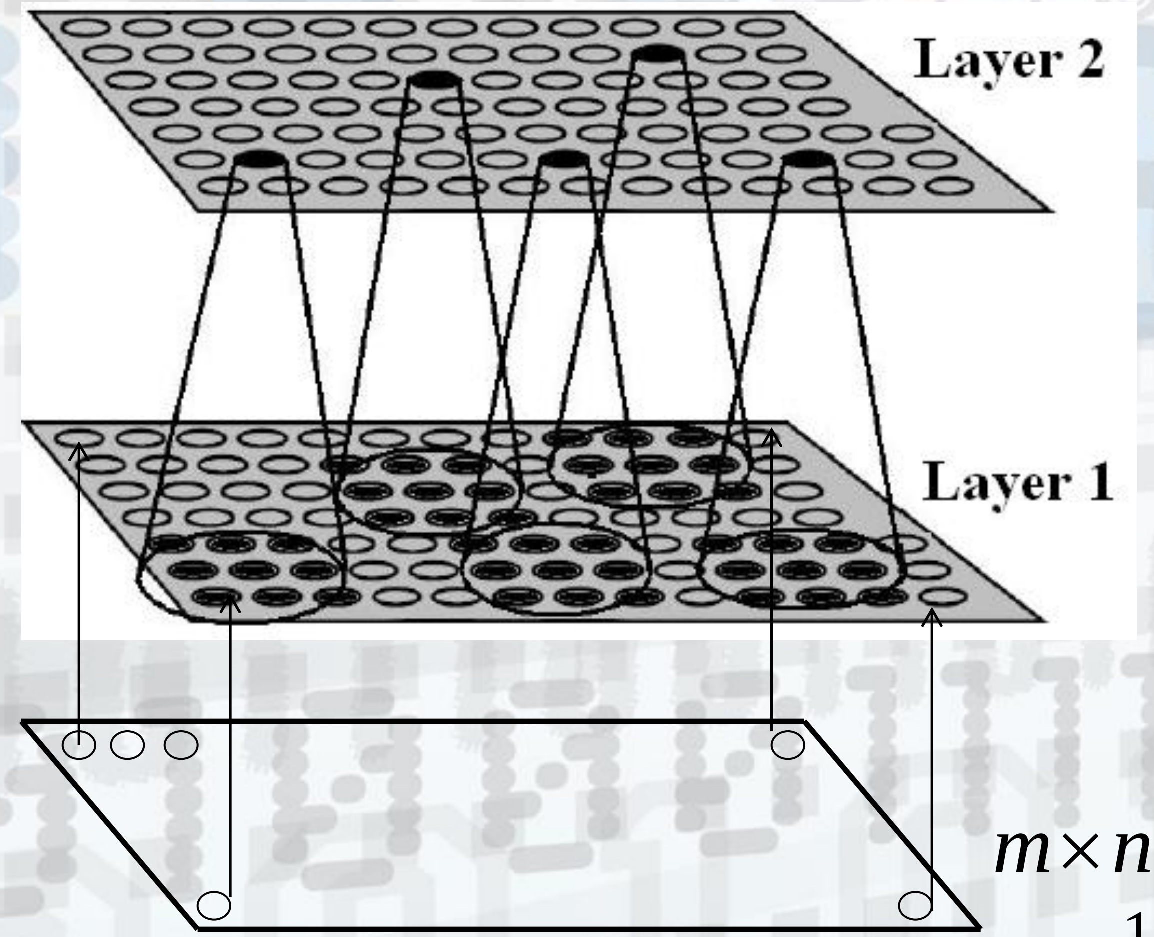
THE ATTENTION LEVEL:

A neuron from the second layer receives the information from a neighborhood of neurons W, composed by $a \times b$ bi-stable neurons.



THE ATTENTION LEVEL:

Each neuron has two states: **active and inactive.**

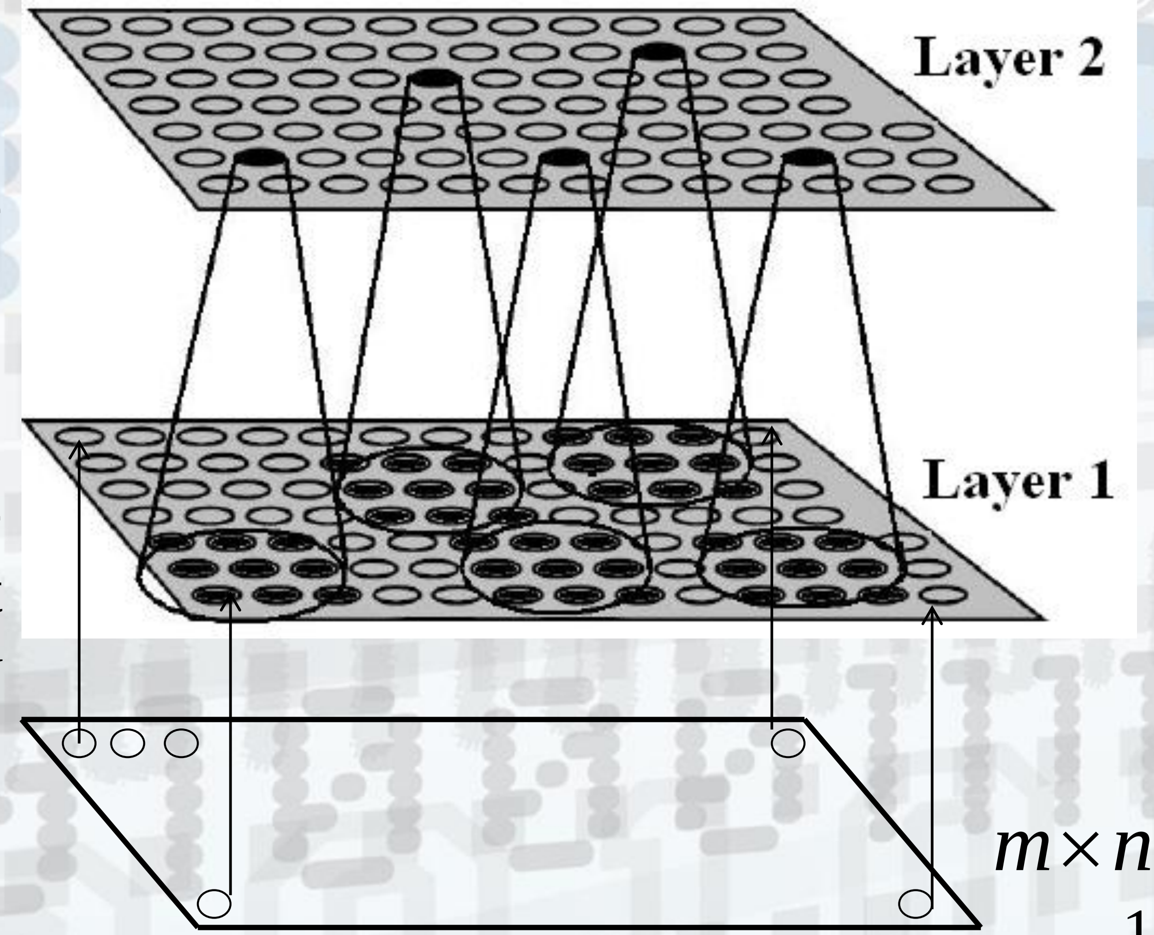


THE ATTENTION LEVEL:

A neuron is **active** when it receives an excitatory signal from the evolutionary control.

The information from the visual stimulus **passes** to the next level.

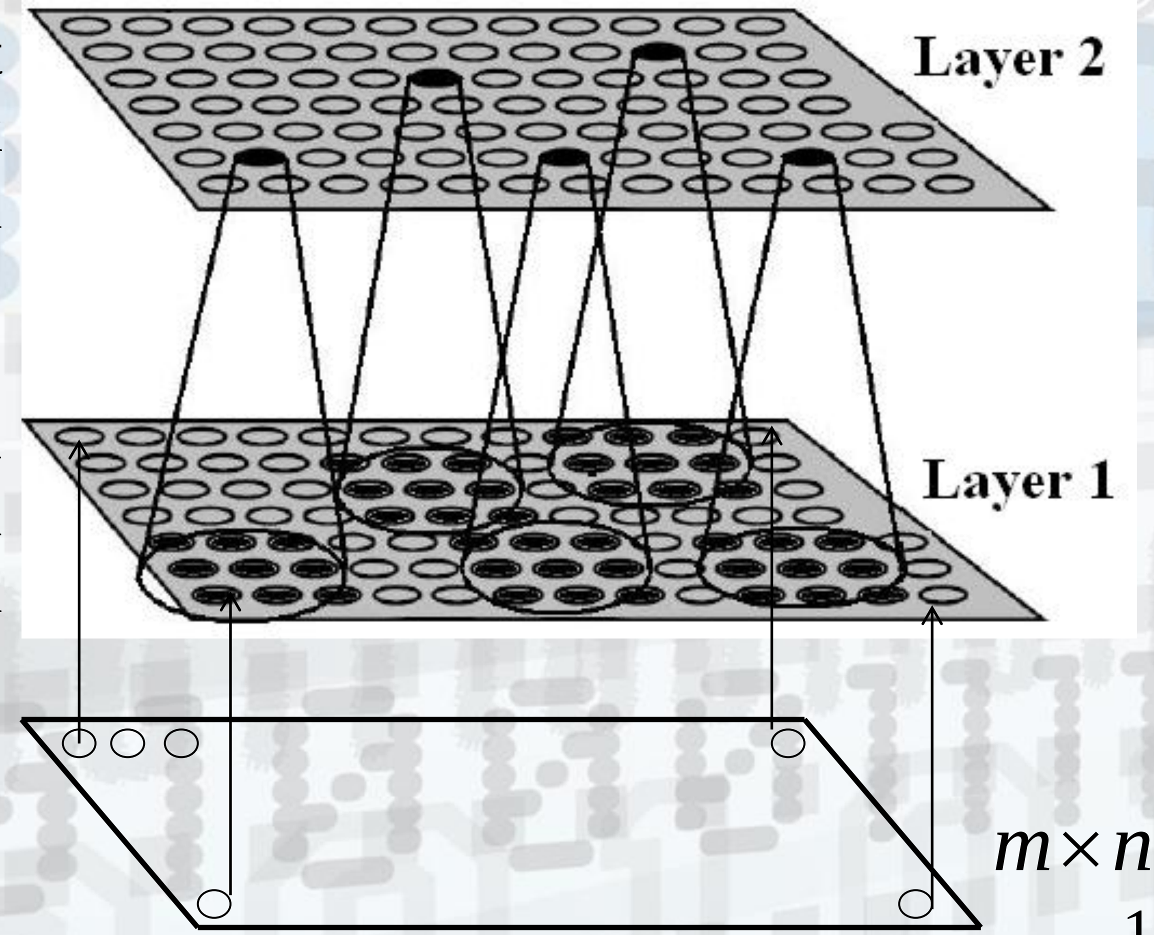
On the contrary, the neuron is **inactive** if it receives an inhibitory signal.



THE ATTENTION LEVEL:

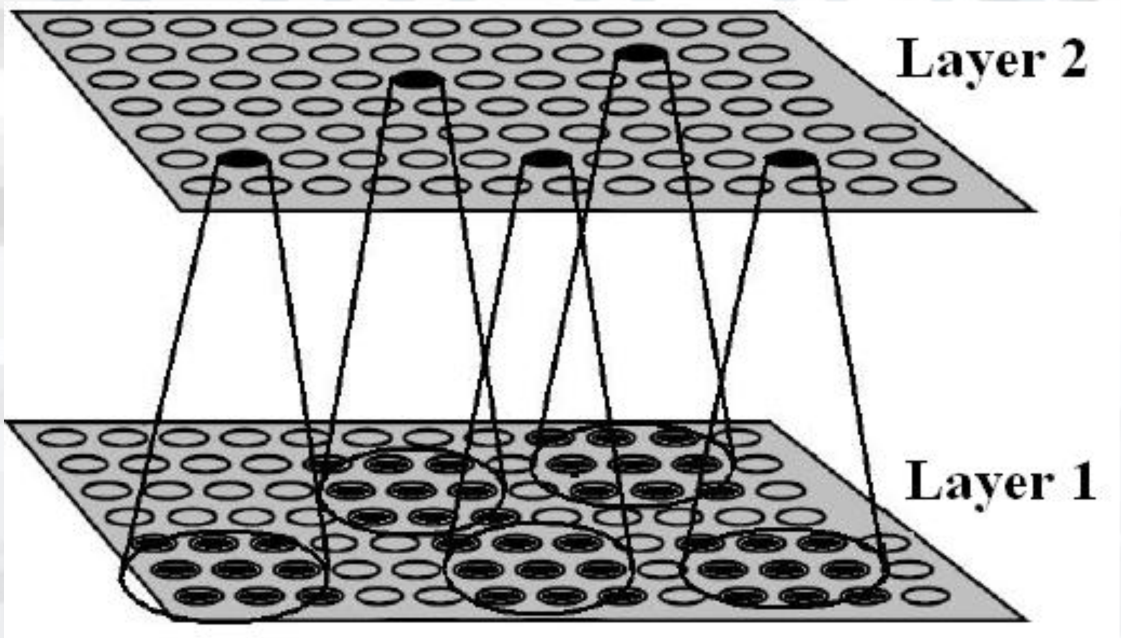
This mechanism **selects** the information that will be taken into account in the next levels; i.e., it determines where to look inside of the visual stimulus.

The visual attention model **will focus only** in those pixels connected with the active neurons from first layer.



THE ATTENTION LEVEL:

Let $f(i,j)$ a visual stimulus and: $y^l(i,j)$ the output of neuron which belongs to l-th layer.



The output of the neurons belonging to the second layer is given by:

$$y^2(i,j) = s(i,j)\cdot\left[\frac{1}{a\times b}\sum_{(i,j)\in W} y^1(i,j)\right]$$

where

$$y^1(i,j) = f(i,j)$$

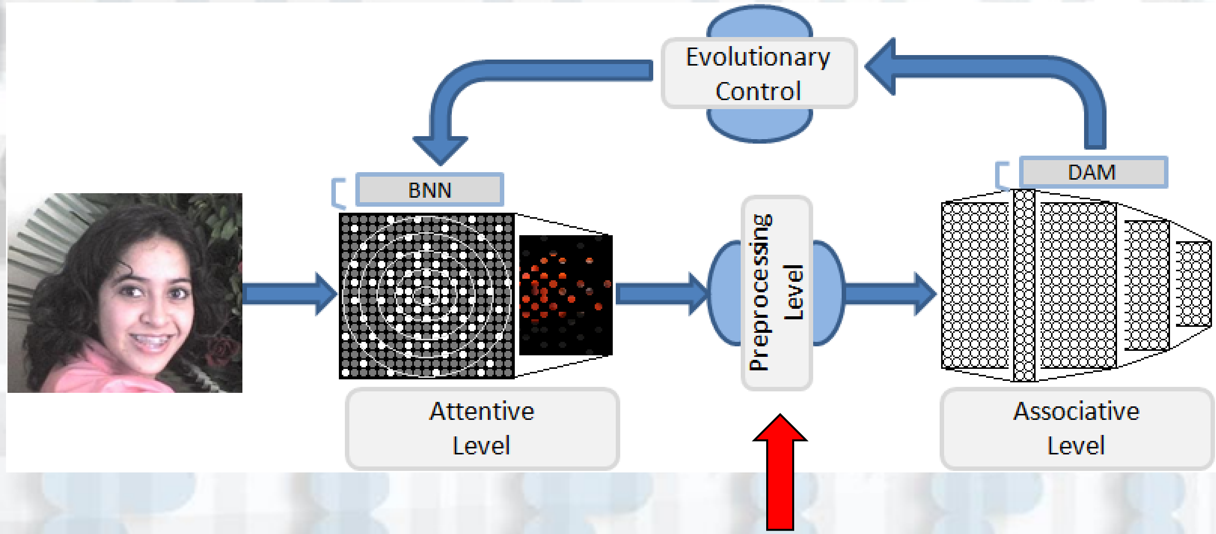
THE ATTENTION LEVEL:

$$y^2(i,j) = s(i,j) \cdot \left[\frac{1}{a \times b} \sum_{(i,j)\in W} y^1(i,j) \right]$$

is the <u>number of neurons</u> belonging to the neighborhood

is the <u>inhibitory/excitatory</u> signal sent by the evolutionary control.

THE PREPROCESSING LEVEL:



This level allows to process the information from the attentive level that will be used by the associative level.

This processing level can be used to modify the behavior of the bi-stable neurons, for example by allowing the neurons to pass only low frequencies or high frequencies or other kind of information.

In this paper we concentrate on the functioning of the attention level.

THE ASSOCIATIVE LEVEL:

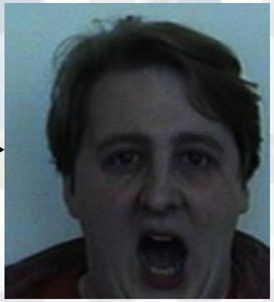
At this level we can associate the selected information form the attentive level with the information we want to recall.

In the content of face recognition, we can, for example, to associate a photo of a person with his/her name or a photo of the person with another of his/her photos.

→ Bryan

or

→



In the first case if a photo of the person is presented to the system it will respond with his/her name. This can be done with **hetero-associative memory.**

In the second case, the system will respond with the stoked photo of the person. This can be done with an **auto-associative memory**.

THE ASSOCIATIVE LEVEL:

We could use any associative memory model.

In this work we have decided to use the Associative Model (AM) recently developed in:

R. A. Vazquez and H. Sossa (2008). A new associative model with dynamical synapses. Neural Processing Letters, 28(3), 189-207.

THE EVOLUTIONARY CONTROL:

As we already said, the evolutionary control combined with the attentive level allows us determine which information from the visual stimulus is used to learn and recognize a face.

Based on the classical DE algorithm (see paper), the best et of active neurons to be used is iteratively determined until we get the best performance of the AM.

K. Price, R. M. Storn and J. A. Lampinen (2005). Diffentential evolution: a practical approach to global optimization. Springer-Verlag.

THE EVOLUTIONARY CONTROL:

1. Select an initial set of active neurons

2. Train the AM using the set of training faces and the selected information from the active neurons.

3. Evaluate the efficiency of the AM by using the validating set of faces.

Iterative Process

GOAL: To get the best combination of a AM and set of active neurons.

For this we use an evaluation function the evaluates the efficiency of the AM and the selected set of active neurons.

THE EVOLUTIONARY CONTROL:

What do we expect?

Through each generation to maximize the accuracy of the associative model in terms of the selected active neurons.

At the end of this evolutionary learning process, we expect to obtain the set of active neurons that maximizes the accuracy of the Associative Memory.

EXPERIMENTAL RESULTS:

To test the efficiency of the proposal we have used the database of faces reported in: L. Spacek. 1996.

Collection of facial images: Grimace. Available from http://cswww.essex.ac.uk/mv/ allfaces/grimace.html

This DB contains 20 photos of different people. Each one is in colour of 180 x 200 pixels.

EXPERIMENTAL RESULTS:

The database was divided into three sets of images.

First photo of each person (15 in total) was used to train the AMs.

The remaining 285 photos (19 for each person) were divided in two sets: **validation set and testing set.**

Validation set was used to find the set of active neurons.

Testing set was used to test the efficiency of the proposed method.

EXPERIMENTAL RESULTS:

We performed eight experiments.

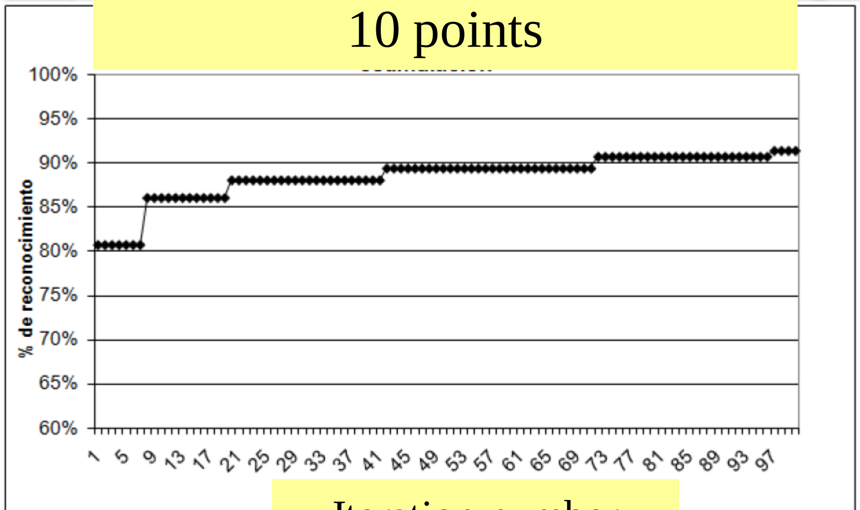
In each experiment the number of desired active neurons was varied from 5 to 40 in steps of 5.

The parameters for the DE algorithm were set to NP=10, $MAXGEN$=100, CR=0.5 and F=rand(0.3,0.9).

EXPERIMENTAL RESULTS:

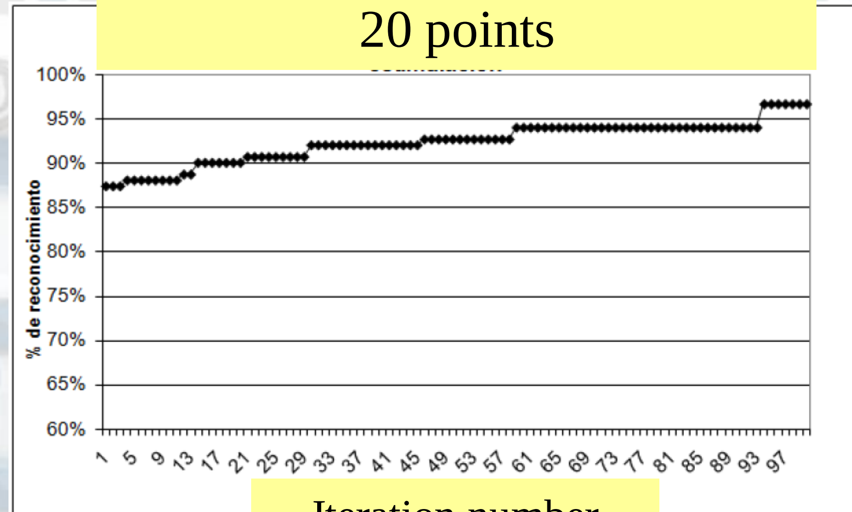
Performance of the proposal in terms of the selected number of neurons.

10 points



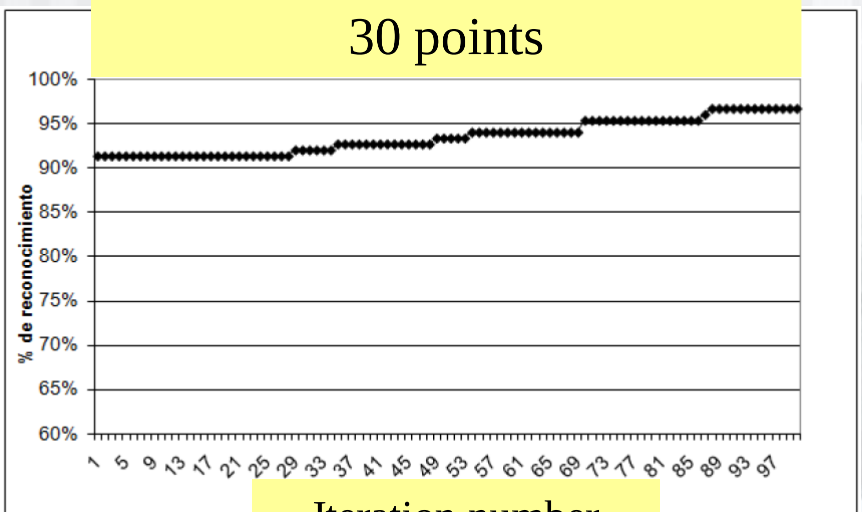
Iteration number

20 points



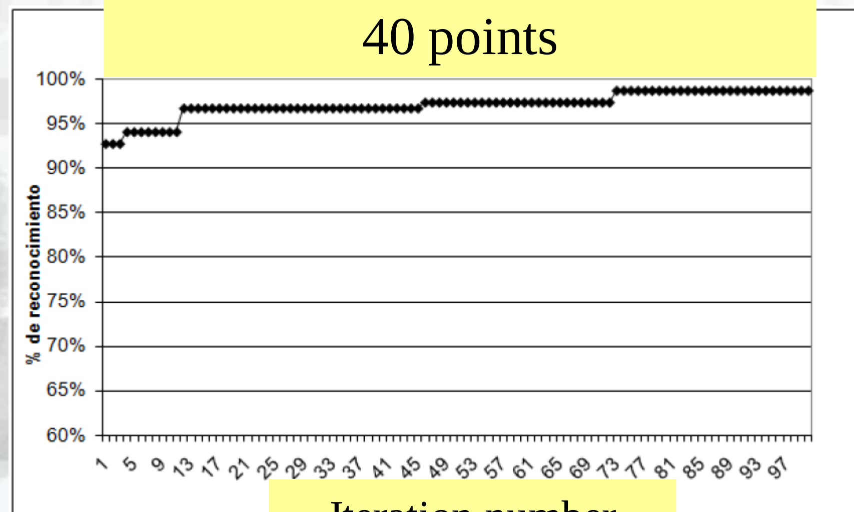
Iteration number

30 points



Iteration number

40 points



Iteration number

EXPERIMENTAL RESULTS:

These results were compared against the results provided by the infant vision model (RIVSM) described in:

R. A. Vazquez and H. Sossa (2007). A computational approach for modeling the infant vision system in object and face recognition. J. BMC Neurosci. 8(suppl 2), P204.

R. A. Vazquez, H. Sossa, B. A. Garro (2007). Low frequency responses and random feature selection applied to face recognition. LNCS 4633:818–830.

R. A. Vazquez, H. Sossa and B. A. Garro (2007). 3D Object recognition based on low frequency responses and random feature selection. LNAI 4827:694–704.

R. A. Vazquez, H. Sossa and B. A. Garro (2009). The role of the infant vision system in 3D object recognition. LNCS 5507:800–807.

R. A. Vazquez, H. Sossa and B. A. Garro (2010). 3D object recognition based on some aspects of the infant vision system and associative memory. Cognitive Computation, 2(2):86–96.

EXPERIMENTAL RESULTS:

Table 1. Comparison of the evolutionary feature-based visual attention model (EFVAM) against the infant vision system model (RIVSM).

	RIVSM		EFVAM		
Active Neurons from BNN	%Tr. Er.	%Te. Er.	%Tr. Er.	%Te. Er.	Gen.
5	0	47.4	11.3	**20**	100
10	0	32.2	8.6	**16.6**	100
15	0	22.3	5.3	**12**	100
20	0	17	3.3	**11.3**	100
25	0	17.3	2.6	**10**	100
30	0	14.2	3.3	**10**	100
35	0	12.6	1.3	**6**	100
40	0	12.1	1.3	**2.6**	100

Tr. Er = Training Error, Te. Er. = Testing Error

CONCLUSIONS:

Only 5 active neurons were required to reach an error of 20% during the recognition process.

…The model was able to find the most five representative pixels from the visual stimulus.

With only 40 active neurons the model was capable to correctly recognizing almost the 98% of the testing set faces.

…although this model is perhaps not biologically plausible, it is capable to recognize faces even if they appear in different facial expressions.

Through several experiments, we have shown how the accuracy of the proposed method is increased through the evolutionary learning process.

…**less that the 1% of the information provided by the visual stimulus was used.**

FONDO DE COOPERACIÓN INTERNACIONAL EN CIENCIA Y TECNOLOGÍA UNIÓN EUROPEA-MÉXICO

Authors thank the COTEPABE-IPN, European Union, the European Commission and CONACYT for the economical support. This paper has been prepared by economical support of the European Commission under grant FONCICYT 93829. The content of this paper is an exclusive responsibility of the CIC-IPN and it cannot be considered that it reflects the position of the European Union.

¡ THANKS !

Juan Humberto Sossa Azuela

E-mail: hsossa@cic.ipn.mx and humbertosossa@gmail.com

Center for Computing Research – National Polytechnic Institute

Tel. 55 (55) 5729 6000 ext. 56512