Purposive Evolution for Object Recognition Using an Artificial Visual Cortex

Eddie Clemente, Gustavo Olague*, and Leon Dozal

Abstract. This work presents a novel approach to synthesize an artificial visual cortex based on what we call organic genetic programming. Primate brains have several distinctive features that help in the outstanding display of perception achieved by the visual system, including binocular vision, memory, learning, and recognition, to mention but a few. These features are processed by a complex arrangement of highly interconnected and numerous cortical visual areas. This paper describes a system composed of an artificial dorsal pathway, or where stream, and an artificial ventral pathway, or what stream, that are fused to create a kind of artificial visual cortex. The idea is to show that genetic programming is able to evolve a high number of heterogeneous trees thanks to the hierarchical structure of our virtual brain. Thus, the proposal uses two key ideas: 1) the recognition of objects can be achieved by a hierarchical structure using the concept of function composition, 2) the evolved functions can be related to the tissues of an artificial organ. Experimental results provide evidence that high recognition rates could be achieved for a well-known multiclass object recognition problem.

1 Introduction

The brain is the most sophisticated organ in the human body; its fundamental task is to control and manage the activities that perform sensorial organs. The neurologists have divided the human brain into four lobes: frontal, temporal, parietal, and

Eddie Clemente
Tecnológico de Estudios Superiores de Ecatepec, Avenida Tecnológico S/N,
Esq. Av. Carlos Hank González, Valle de Anáhuac, Ecatepec de Morelos
e-mail: `eddie.clemente@gmail.com`

Gustavo Olague
CICESE, EvoVision Project, Applied Physics Division
e-mail: `olague@cicese.mx`

Leon Dozal
CICESE, EvoVision Project, Computer Science Department
e-mail: `leon.dozal@gmail.com`

* Corresponding author.

O. Schütze et al. (Eds.): EVOLVE – A Bridge between Probability, AISC 175, pp. 355–370.
springerlink.com

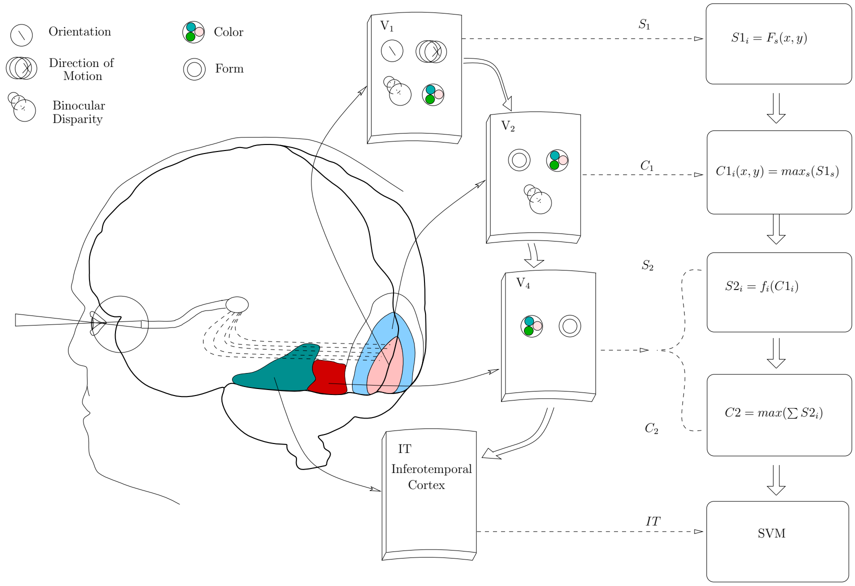


Fig. 1 Analogy between the ventral stream and the proposed computational model

occipital. The last one has special interest to the research community interested in the sense of vision, because this is the lobe where it is located the visual cortex. In fact, the primary visual cortex and secondary visual areas are specialized for image processing, object localization, and the estimation of direction, velocity and object trajectories. Figures 1 illustrates the main ideas that we are proposing to approach the problem of object recognition. We divide the approach into two key ideas. The first one is related to the identification of salient features through the application of a set of functions that should be able to identify the salience properties that characterize a given object. In general, the works that follow a modeling of the human visual system like [[8],[24], [18],[15],[11]] are based on a set of image patches that are used as a dictionary of visual words. These small images represent the most common and useful characteristics presented in all images of a database integrated by a number of visual categories. The hypothesis made in our work is that such set of image patches could be substituted with a set of mathematical functions. The second idea, that we want to outline is based on the concept of an organ. In biology, an organ is described as a collection of tissues joined in a structural unit to serve a common function. In particular, we are interested in studying the brain; specially the visual cortex and how it is explained the functionality of the main tissues involved into the process of object recognition.

The goal of this work is to outline a methodology based on organic genetic programming implemented through the modeling of the hierarchical structure of the visual cortex and the concept of function composition inspired from the idea of an organ. In this way, a functional approach is enforce in order to solve the problem of object recognition.

2 Visual Cortex

Visual processing is performed by the brain, and the explanation about how it works is based on the idea of two visual subsystems. Today, it is widely accepted the two-streams hypothesis as a way of describing the phenomenon of visual perception. This knowledge is based on neuropsychological, neurophysiological, and psychophysical evidence regarding the existence of two visual subsystems known as ventral and dorsal streams. Thus, the explanation is founded on the idea that both systems manage the same visual information, but the difference lies in the transformations that both streams performed to the visual data. This is clearly exposed in the change of paradigm from a what/where dichotomy into a vision-for-action/vision-for-perception duality used to explain the same dorsal/ventral anatomical distinction, see [[22], [23], [29], [28], [17]]. Next, we briefly describe the ventral and dorsal streams.

The ventral stream is largely associated with object recognition and shape representation, see [19]. The ventral or what pathway starts at the retina, and it receives its main input from the parvocellular layer of the lateral geniculate nucleus of the thalamus, and it projects into $V1$, which is part of the primary visual cortex, also called striate cortex, which is located at the back of the brain. Then, the ventral pathway continues into the visual areas $V2$ and $V4$, which are part of a region known as the extrastriate visual cortex, and finally to the areas TEO and TE of the inferior temporal cortex. In computer science, the ventral stream is explained as performing a hierarchical and feedforward process that is specialized for object recognition and is biologically inspired from [9]. Most proposed models start with an image that is decomposed into a set of alternating “S” and “C” layers that are named after the discovery of Hubel and Wiesel of the simple and complex cells, see [9]. The idea was originally implemented by Fukushima in the neocognitron system, see [8]. This system was further enhanced by other authors including the convolutional networks [13], and the HMAX model [21]. In all these models the simple layers apply local filters in order to compute higher-order features, and the complex layers increase invariance by combining units of the same type, see [27].

The dorsal stream, also known as the “where” or “how” pathway, is related to the visual processing of spatial locations. Nevertheless, this part of the visual processing is still controversial; since, the dorsal stream is said to be involved in the guidance of actions, as well as, the spatial localization of objects in space. Like the ventral stream, the dorsal stream starts at the retina, and it receives its main, if not total, input from the magnocellular retinocortical layer of the lateral geniculate nucleus of the thalamus, and it projects into $V1$, but it also receives direct subcortical inputs from the superior colliculus and pulvinar structures. Then, the dorsal stream continues through $V2$, $V3$, the middle temporal area MT, and the medial superior temporal area MST, which are part of the extrastriate visual cortex; and finishing in the posterior parietal cortex and adjacent areas. In general, it is acknowledged that visual attention is performed by the dorsal stream, and the most widely accepted paradigm for visual attention is the feature integration theory, see [26]. However, there are other theories that attempts to explain the workings of visual attention in the dorsal stream, like [19] and [31]; or even a work that relate visual attention to both streams, see [6]. In computer science, the first computational approach for visual attention was introduced by Koch and Ullman in 1985, see [12]. Later, other researchers proposed several methodologies, which are based mostly in the feature integration theory, like [14], and [11]. In all these models the image is decomposed in several dimensions in order to obtain a set of conspicuity maps and then integrate them into a saliency map.

In this way, the visual system has been defined by two information processing streams organized in two broad structures subserving object and spatial vision. The classical dichotomy between object and space perception focuses on the importance of a single and general purpose representation. On the other hand, the “what” and “how” theory of Milner and Goodale [17] gives emphasis to the idea that the visual system is defined according to the requirements of the task that each stream subserves. The idea is to define multiple frames of reference giving special attention to the goal of the observer. In this way, the same object and spatial information is transformed by the visual system for different purposes. Thus, the ventral system represents the visual world in allocentric coordinates by promoting conscious perceptual awareness. On the other hand, the dorsal stream uses egocentric coordinates to transform the information about objects location, orientation and size, see [5].

3 Evolution and Teleology for Visual Processing

This section is devoted to the idea that visual processing is a product of brain evolution; and therefore it is plausible to follow an artificial evolutionary approach in the search of object recognition programs. The explanation that outlines our computational approach will be developed in two parts. The first reviews some explanations about how the brain has evolved. Next, we explain how the two stream hypothesis can be understood in teleological terms. In fact, we would like to stress that there are two main viewpoints that are used within evolutionary explanations. Today, there are two schools of knowledge, mechanistic and teleological, that attempt to provide an explanation for understanding nature. Note, that teleological explanations does not exclude mechanisms. Also, the controversy is still alive mainly because teleological explanations cannot materialise the idea of purpose and at the same time the mechanistic explanations cannot vanish the idea of purpose. Nevertheless, a purpose is not a desire; and when we refer to a purpose, we talk in terms if it is achieved or not. Hence, we claim that our personal teleological viewpoint offers the possibility of developing a newer and rich explanation about artificial evolution, which is in accordance to many theoreticians of brain evolution that use a teleological language, see [17, 5, 2]; and philosophers of science, see [1, 14, 25].

Classical definitions of vision implicitly and explicitly assume that the purpose of the visual system is to construct some sort of internal model of the environment; in other words, a kind of visual representation of the real world, that serves as the perceptual foundation for all visually derived thought and action. The approaches to study the structure and functionality of the neocortex are based on comparative, developmental, and functional or adaptationist explanations. In particular, we hold to the tenet of many biologists that adaptation, in nature, makes the organs to suit the work they have to do; hence, developmental and functional explanations are complementary and not alternate explanations. This is clearly seen from the fact that different species from the same taxonomic group have evolved specialized visual mechanisms, which are coherent and highly correlated to particular and specific cognitive and behavioral functions, which were evolved based on the principle of natural selection. In this way, the goal in this work is to evolve a system that is based on a description of the ventral and dorsal streams and to adapt their behavior to an specific task.

4 Artificial Visual Cortex (AVC)

The proposed approach is inspired and based on the idea that an organ is a collection of tissues joint in a structural unit to serve a common function. In this way, the central nervous system is understood as the organ and the study is limited to the retina, brain, in particular the visual cortex, and how it is processed the visual information. In the literature, the computational approaches inspired from the visual cortex are always centered into the dorsal or ventral streams, with the idea of solving the visual attention or object recognition tasks, respectively. Indeed, there is not yet a significant work that attempts to model the visual cortex as a whole and unique system. In general, the only works that consider the subject use a visual attention module as input to the object recognition method to create a more complex system. Instead of this simple approach and according with the reviewed literature that states that layers $V1$ and $V2$ are part of both streams, dorsal and ventral, we propose a new modeling, see Figure 2, in order to create an artificial visual cortex using the idea of function composition. This new methodology makes necessary to understand an image like the graph of a function. The function in this case is understood like the physical, geometrical, or other properties of the scene. In this way, in order to describe the idea we define an image as the graph of a function.

Definition 1 (Image as the graph of a function) *Let f be a function $f : U \subset \mathbb{R}^2 \to \mathbb{R}$. The graph or image I of f is the subset of $\mathbb{R}^3$ that consist of the points $(x, y, f(x, y))$, in which the ordered pair (x, y) is a point in U and $f(x, y)$ is the value at that point. Symbolically, the image $I = \{(x, y, f(x, y)) \in \mathbb{R}^3 | (x, y) \in U\}$.*

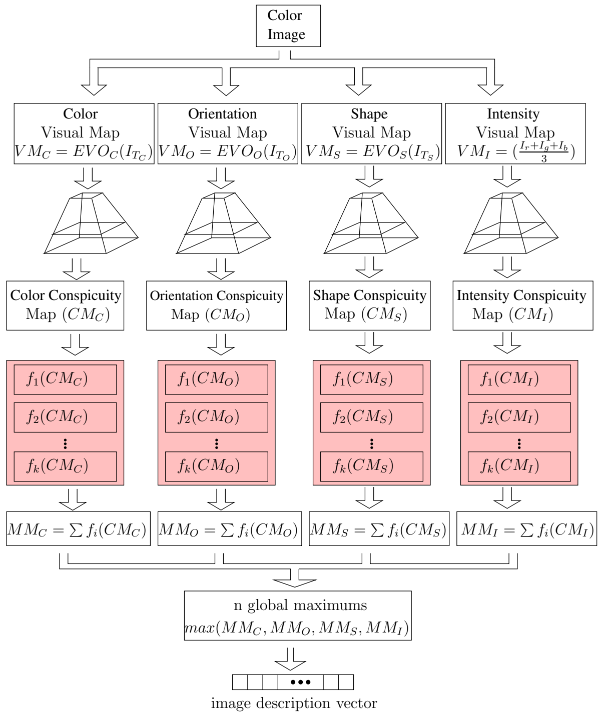


Fig. 2 Flowchart of the artificial visual cortex. Note, the similarity with the visual attention process in which the image is decomposed into several dimensions. In our approach a function driven paradigm is enforced to avoid the application of image patches.

In this way, the image is the input of a computational system that mimics the functionality of an artificial visual cortex by replicating the hierarchical structure of the natural system. Contrary to previous research devoted to object recognition, where the ventral stream is modeled through a data-driven scenario; here, the object recognition system is designed following the hierarchical structure of the dorsal and ventral streams, as well as, the idea that each layer can be modeled with a set of mathematical functions that replicate the functionality of a virtual tissue.

Genetic programming is used as the paradigm to implement the proposed approach for which a set of evolutionary visual operators (EVOs) are optimized according to the hierarchical structure being evolved in the search of an optimal object recognition program. The aim of genetic programming is to find the best set of EVOs using a number of building blocks and the whole hierarchical structure, in order to find a solution to a multi-class object recognition problem. One advantage of the functional approach compared to previous, data-driven, approaches is reflected on the lower amount of computations that brings a significant economy in the number of computer operations without sacrificing the overall quality. Next, it is described the proposed system according to the dorsal and ventral streams.

Table 1 Set of functions and terminals used by EVO_C to create the visual map VM_C

$F_C = \{+, -, \times, \div, \lvert+\rvert, \lvert-\rvert, \sqrt{I_{T_C}}, I_{T_C}^2, log_2(I_{T_C}), Exp(I_{T_C}, Complement(I_{T_C})\}$
$T_C = \{I_r, I_g, I_b, I_c, I_m, I_y, I_k, I_h, I_s, I_v\}$

4.1 Artificial Dorsal Stream (ADS)

This first part of the system is based on the psychological model of Treisman and Gelade [26], which was successfully implemented in [[27], [11]]. The first step of the process is represented by the image acquired by the camera, whose natural counterpart is the retina. Here, the system considers digital color images that are composed of three images at different wavelengths of light that are red, green, and blue. Note, that it is possible to convert an image represented in RGB space into another color space. Thus, we say that a color image is the set of images named $I_{color} = \{I_r, I_g, I_b, I_c, I_m, I_y, I_k, I_h, I_s, I_v\}$. Next, four visual operators are applied separately to emphasize: intensity, color, orientation, and shape. In biological plausible models such as [11], some of these operators are established according to knowledge in neuroscience about how these features are obtained in the visual cortex of the brain, using a data-driven approach. Here, the operation of the dorsal stream is emulated by a set of functions that are evolved with genetic programming to obtain an optimal set of EVOs as depicted on Figure 2. Each EVO is represented as a specialized function that is evolved from a set of suitable characteristics that are used to create a set of visual maps. Note, that each visual map is processed by a number of functions used to create a pyramid that achieves invariance to position and scale. In fact, the result of this process is a conspicuity map for each considered feature. In this way, the evolution is charged of evolving the best possible function that extract color, orientation, or shape information; without focusing on the problem of achieving invariance. Thus, the hierarchical structure helps to achieve the desired result through function composition. Next, we describe the EVO features that are used within the artificial visual cortex.

4.1.1 Evolved Color Map

The color image received as input is transformed with a function, $EVO_C : I_{color} \rightarrow VM_C$, that enhance the color feature. In this way, an EVO_C is evolved with genetic programming to optimize the extraction of color information of the objects within the image. The result is an image or visual map VM_C containing the prominence in color that represents the best feature's image in color space. Thus, the evolutionary process uses the set of functions and terminals provided in Table 2. The notation is summarized as follows, I_{T_C} can be any of the terminals in T_C, as well as the output of any of the functions in F_C; the $Complement(I_{T_C})$ function symbolizes a negative image that is represented by the inversion of an image.

4.1.2 Evolved Orientation Map

The function used to compute the orientation, $EVO_O : I_{color} \rightarrow VM_O$, is evolved with genetic programming to optimize the extraction of edge information within the input image. The result of this operation is a visual map VM_O in which the pixel values represent the feature prominence, in such a way, that the higher the pixel value the greater the prominence of the feature. In this way, genetic programming applies the functions and terminals of Table 1, in order to enhance the best orientation features that are useful for the object recognition task. The notation used is as follows. I_{T_O} can be any of the terminals in T_O; as well as, the output of any of the functions in F_O; D_u symbolizes the image derivatives along direction $u \in \{x, y, xx, yy, xy\}$; G_σ are Gaussian smoothing filters with $\sigma = 1$ or 2.

Table 2 Set of functions and terminals used by EVO_O to create the visual map VM_O

$F_O = \{+, -, \times, \div, \lvert+\rvert, \lvert-\rvert, \sqrt{I_{T_O}}, I_{T_O}^2, log_2(I_{T_O}), G_{\sigma=1}, G_{\sigma=2}, \lvert I_{T_O}\rvert, \frac{I_{T_O}}{2}, D_x, D_y\}$
$T_O = \{I_r, I_g, I_b, I_c, I_m, I_y, I_k, I_h, I_s, I_v, G_{\sigma=1}(I_r), G_{\sigma=2}(I_r), D_x(I_r), D_y(I_r), D_{xx}(I_r), D_{yy}(I_r), D_{xy}(I_r), ...\}$

4.1.3 Evolved Shape Map

The function used to compute the shape features, $EVO_S : I_{color} \rightarrow VM_S$, is evolved with genetic programming to optimize the extraction of shape information in the input image. The result of this operation is a visual map VM_S that provides the form and structure of the object of interest within the image. In this way, genetic programming applies the functions and terminals of Table 3. We would like to remark that the application of this kind of morphological functions has not been applied in previous research regarding the ventral and dorsal streams. Thus, according to

Table 3 Set of functions and terminals used by EVO_S to create the visual map VM_S

$F_S = \{+, -, \times, \div, round(I_{T_S}), floor(I_{T_S}), ceil(I_{T_S}), dilation_{diamond}(I_{T_S}), dilation_{square}(I_{T_S}), dilation_{disk}(I_{T_S}), erosion_{diamond}(I_{T_S}), erosion_{square}(I_{T_S}), erosion_{disk}(I_{T_S}), skeleton(I_{T_S}), boundary(I_{T_S}), hit - miss_{diamond}(I_{T_S}), hit - miss_{square}(I_{T_S}), hit - miss_{disk}(I_{T_S}), top - hat(I_{T_S}), bottom - hat(I_{T_S}), open(I_{T_S}), close(I_{T_S})\}$
$T_S = \{I_r, I_g, I_b, I_c, I_m, I_y, I_k, I_h, I_s, I_v\}$

the literature the work reported in this paper could be considered as the first to use morphological image processing within the modeling of the visual cortex.

Finally, in order to obtain the intensity of an input image I_{color}, we apply a similar process described in previous research where the red, green, and blue values of each pixel are averaged. The formula is developed as a function $VM_I : I_{color} \rightarrow I$, that is obtained with the following formulae $VM_I = \frac{I_r+I_g+I_b}{3}$.

4.1.4 Conspicuity Maps

The conspicuity maps (CMs) are obtained by means of a center-surround function, which is applied to the visual maps in order to simulate a set of center-surround receptive fields. This natural structure allows the ganglion cells to measure the differences between firing rates in center (c) and surroundings (s) of ganglion cells. First, a pyramid $VM_l(\alpha)$ of nine spatial scales $S = \{1,2,...,9\}$ is created for each of the four resulting VMs. Afterwards, an across-scale substraction $\ominus$ is performed, resulting in a center-surrond map $VM_l(\omega)$ in such a way that the value of the pixel is augmented as long as the contrast is increased within their neighbors at different scales. Finally, the $VM_l(\omega)$ maps are added using an across-scale addition $\oplus$ in order to obtain the desired conspicuity maps CM_l.

Until this stage, we have four CMs, one for each feature, as shown in Figure 2. The CMs are obtained similar to Walther and Koch model [24]. Next, instead of combining the CMs into a single saliency map, the idea here is to use the four CMs as input to an artificial ventral stream in order to derive a vector descriptor, which will be use by a classifier. In fact, the fitness function is computed from the accuracy achieved with a support vector machine (SVM).

4.2 Artificial Ventral Stream (AVS)

Now, that all regions have been highlighted; the next step is to describe such important regions. The typical approach is based on a template matching technique between the information obtained with an interest region selection process and a number of prototype patches. Traditionally, the goal is to learn a set of prototypes that are known as the universal dictionary of features and which are used to identify all object categories. Hopefully, the SVM can recognize the prototypes that correspond to a specific image of a given category. On the other hand, the proposal in this paper is to optimized the functionality of the ventral stream that is evolved with the aim of enhancing the set of prominent features that were highlighted during the interest region detection computed in previous stages. Thus, in this work the selection of interest regions is performed by the artificial dorsal stream through the transformation of the conspicuity maps. It should be noted that according to the artificial ventral stream each evolved function is a composite function that is capable of substituting several prototype features; thus, reducing significantly the total number of operations needed to define all object features that are used to describe and classify the input images. According to Figure 2, the information provided by the conspicuity maps is feedforward to k operators that emulate a set of lower order hypercomplex cells replicating the functionality of a virtual tissue. Thus, all evolved functions along each dimension are added in order to obtain a single measure that we called mental map. Hence, a mental map is obtained for each dimension: color, orientation, shape, and intensity. In this way, all mental maps are combined with a max operation that is used to highlight the necessary characteristics that recognize a specific object class. Note, that each function is an evolved visual operator (EVO) built by several GPs from the particular set of terminals and functions shown in Table 4. Note also that this second stage could be said to perform an information description operation (IDO) with the aim of discovering the best set of functions that creates the most discriminant vector of characteristics. Hence, this set of functions replaces the universal dictionary proposed in [27, 18, 24] and we claim that it corresponds to a function driven approach.

Table 4 Set of functions and terminals for the ventral stream

Functions:	$+,\ -,\ /,\ *,\ \|-\|,\ \|+\|,\ \sqrt{\cdot},\ (\cdot)^2,\ log(\cdot),$ $D_x(\cdot), D_y(\cdot), D_{xx}(\cdot), D_{xy}(\cdot), D_{yy}(\cdot), Gauss_{\sigma_1}(\cdot),$ $Gauss_{\sigma_2}(\cdot)$, 0.05$(\cdot)$
Terminals:	$C1,\ D_x(C1),\ D_{xx}(C1),\ D_y(C1),\ D_{yy}(C1),$ $D_{xy}(C1)$

5 Evolving AVCs with Organic Genetic Programming

This section describes the main aspects for the evolution of AVCs through the application of what we called organic genetic programming (OGP). All elements introduced in the OGP embody an organic motivation, in a sense of describing an organ composed of tissues, which could be part of an artificial living organism. Figure 3 illustrates the complexity of the proposed system using a kind of heterogeneous and hierarchical genetic programming. In our model the genotype is built from several trees that can be seen as the genes and which are arranged into a complex chromosome. The phenotype is decoded according to Figure 2. Thus, the algorithmic process that mimics the visual information processing of an AVC should be seen as a single entity. In other words, the functional representation of the artificial organ is represented by the whole hierarchical and heterogeneous structure. The representation that is proposed has the aim of ensuring the development of complex functions, while freely increasing the number of programs according to the task at hand; in this case, the classification of several object classes. In this way, the structure can grow in the number and size of its elements. Hence, it is important to note that each individual

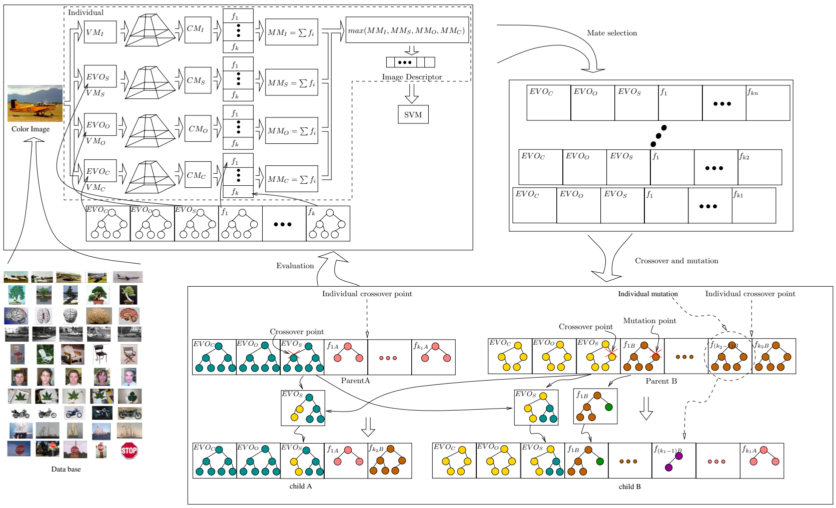


Fig. 3 General flowchart of the methodology to synthesize an artificial visual cortex

within a population should be understood as the whole AVC and it is therefore not only a list of tree-based programs, but the whole information processing depicted in Figure 2.

6 Experimental Results

This section provides details about the experiments in order to explain the system that was implemented to learn an artificial visual cortex. All experiments were performed in a Dell Precision T7500 Workstation, Intel Xeon 8 Core, NVIDIA Quadro FX 3800 and Linux OpenSUSE 11.1 operating system. The system was tested using 10 classes and 15 images per class of the Caltech 101 database, see [7]. The classifier used in the experiments was the SVM implementation developed by Chan and Lin, see [3], in order to compare with the HMAX model [24]. Table 5 presents a summary of the best results and a comparison with the HMAX model, an implementation HMAX-CUDA, and a previous proposal called the artificial ventral stream (AVS), see [4]. Note, that the total number of convolutions is much lower than the HMAX and HMAX-CUDA. This aspect is important since the factor of improvement is on the order of hundred of operations. However, the performance of the AVC is lower than the HMAX model but its level is worse in testing, while the effectiveness of our approach remains constant. Figure 4 shows the run where the best program was obtained, see Figure 4. Also, the Figure 6 illustrates the range of

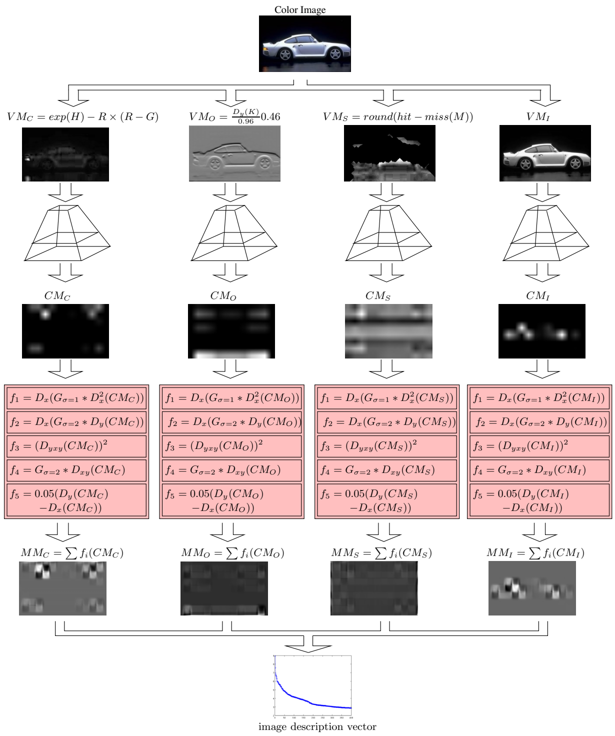
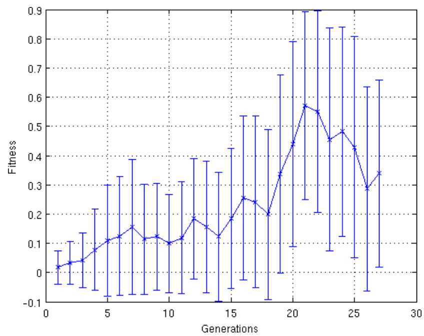


Fig. 4 Flowchart of the best individual achieved with the methodology to synthesize an artificial visual cortex

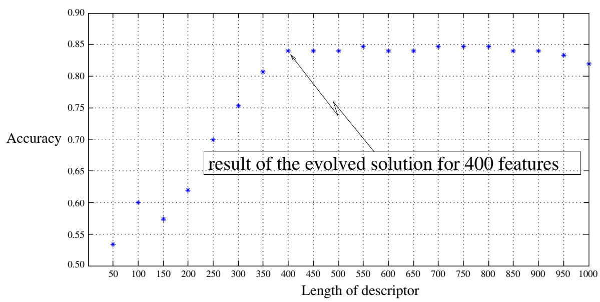
descriptor values of the best solution for each class. We provide also the overall results of the best AVC through the confusion matrix, see Table 7. Due to the level of accuracy being achieved by the AVC we decide to make a simpler test. We evaluated the performance of the proposed model in the object present/abscent experiment using several object classes from the same CalTech data set. In this experiment, each data set was randomly divided in two sets for training and testing using 50 images for each set out of 800 images. We remark that for this case the algorithm scores a perfect solution during the initial random population. Therefore, it was not necessary to evolve the AVC to find a solution to the problem. Table 6 shows a comparison with the HMAX model using boost and SVM as classifiers for the following classes: airplanes, cars, faces, leaves, and motorbikes.

Table 5 This table shows the comparison of performances between HMAX, HMAX-CUDA, AVS and AVC

	Image size	HMAX MATLAB	HMAX CUDA	Artificial V. S.	Artificial V. C.
Running time over different image size	896 × 592	34s	3.5s	2.6s	9.91s
	601 × 401	24s	2.7s	1.25s	5.32s
	180 × 113	9s	1s	0.23s	0.49s
Performance over 15 training images per 10 classes		94%	94%	78%	85.3%
Performance over 15 testing images per 10 classes		73%	73%	80%	84%
Number of convolutions		4848	4848	216	95



(a) Average fitness with standard deviation



(b) Behavior of accuracy with respect to the descriptor length

Fig. 5 Figure (a) shows the average fitness and standard deviation of the run that produces the best individual. Figure (b) depicts the performance after changing the descriptor length.

Table 6 This table shows the performance comparison between HMAX, HMAX-CUDA, and AVC. Note, that in the case of the HMAX model a learning process was necessary to identify the best patches, while for the AVC only a random sampling to discover the best solution.

Data sets	Performance of HMAX		Artificial V. C.
	boost	SVM	
Airplanes	96.7 %	94.9 %	100 %
Cars	95.1 %	93.3 %	100 %
Faces	98.2 %	98.1 %	100 %
Leaves	97.0 %	95.9 %	100 %
Motorbikes	98.0 %	97.4 %	100 %

Table 7 This table shows the results of the best solution in the form of a confusion matrix obtained during the AVC testing. The final accuracy $acc = 84\%$ classifies correctly (126/150) images.

	Airplanes	Bonsai	Brains	Cars	Chairs	Faces	Leaves	Motorcycle	Schooner	Stop Signal
Airplanes	15	0	0	0	0	0	0	0	0	0
Bonsai	0	9	2	0	3	0	0	0	0	1
Brains	0	2	11	0	1	0	0	0	1	0
Cars	0	0	0	14	0	0	0	0	1	0
Chairs	0	0	2	1	11	0	0	0	0	1
Faces	0	0	0	0	0	14	0	0	1	0
Leaves	0	0	0	0	0	0	15	0	0	0
Motorcycle	0	0	0	0	0	0	0	15	0	0
Schooner	0	1	0	0	0	0	0	0	12	2
Stop Signal	0	2	1	0	0	1	0	0	1	10

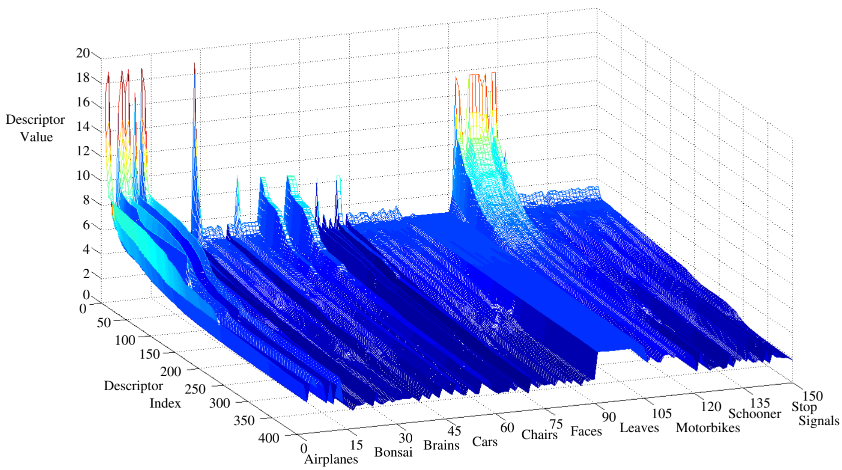


Fig. 6 This plot shows the descriptors of the best individual that are used as input to the SVM.

7 Conclusions

This work shows that a complex program, mimicking an artificial visual cortex, with numerous trees can be evolved to approach successfully a multi-object recognition problem.

Acknowledgements. This research was founded by CONACyT through the Project 155045 - "Evolución de Cerebros Artificiales en Visión por Computadora". First author supported by scholarship 164884 from CONACyT. This research was also supported by TESE through the project DIMI-MCIM-004/08.

References

1. Ayala, F.J.: Teleological explanations in evolutionary biology. Philosophy of Science 37(1), 1–15 (1970)
2. Barton, R.A.: Visual specialization and brain evolution in primates. Proceedings of the Royal Society of London Series B-Biological Sciences 265(1409), 1933–1937 (1998)
3. Chang, C.-C., Lin, C.-J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27 (2011), Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`
4. Clemente, E., Olague, G., Dozal, L., Mancilla, M.: Object Recognition with an Optimized Ventral Stream Model Using Genetic Programming. In: Di Chio, C., Agapitos, A., Cagnoni, S., Cotta, C., de Vega, F.F., Di Caro, G.A., Drechsler, R., Ekárt, A., Esparcia-Alcázar, A.I., Farooq, M., Langdon, W.B., Merelo-Guervós, J.J., Preuss, M., Richter, H., Silva, S., Simões, A., Squillero, G., Tarantino, E., Tettamanzi, A.G.B., Togelius, J., Urquhart, N., Uyar, A.Ş., Yannakakis, G.N. (eds.) EvoApplications 2012. LNCS, vol. 7248, pp. 315–325. Springer, Heidelberg (2012)
5. Creem, S.H., Proffitt, D.R.: Defining the cortical visual systems: “what”, “where”, and “how”. Acta Psychologica 107, 43–68 (2001)
6. Desimone, R., Duncan, J.: Neural mechanisms of selective visual attention. Annu. Rev. Neurosci., 193–222
7. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories, p. 178 (2004)
8. Fukushima, K.: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics 36, 193–202 (1980)
9. Hubel, D.H.: Exploration of the primary visual cortex. Nature, 515–524 (1982)
10. Hubel, D.H., Wiesel, T.N.: Receptive fields of single neurones in the cat’s striate cortex. J. Physiol. 148(3), 574–591 (1953)
11. Itti, L., Koch, C.: Computational modelling of visual attention. Nature Review Neuroscience 2(3), 194–203 (2001)
12. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Hum. Neurobiol. 4(4), 219–227 (1985)
13. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11), 2278–2324 (1998)
14. Lennox, J.G.: Darwin was a teleologist. Biology and Philosophy 8(4), 409–421 (1993)
15. Mel, B.W.: Seemore: Combining color, shape, and texture histogramming in a neurally-inspired approach to visual object recognition. Neural Computation 9, 777–804 (1997)
16. Milanese, R.: Detecting salient regions in an image: from biological evidence to computer implementation. PhD thesis, Department of Computer Science, University of Genova, Switzerland (December 1993)
17. Milner, A.D., Goodale, M.A.: The Visual Brain in Action, 2nd edn. Oxford University Press, Oxford (2006)
18. Mutch, J., Lowe, D.G.: Object class recognition and localization using sparse features with limited receptive fields. Int. J. Comput. Vision 80, 45–57 (2008)
19. Oram, M.W., Perrett, D.I.: Modeling visual recognition from neurobiological constraints. Neural Networks 7(6), 945–972 (1994)
20. Rensink, R.A.: The dynamic representation of scenes. Visual Cognition 7(1-3), 17–42 (2000)

21. Riesenhuber, M., Poggio, T.: Hierarchical models of object recognition in cortex. Nature Neuroscience (11), 1019–1025
22. Schneider, G.E.: Contrasting visuomotor functions of tectum and cortex in the golden hamster. Psychologische Forschung 31(1), 52–62
23. Schneider, G.E.: Two visual systems. Science 163(3870), 895–902 (1969)
24. Serre, T., Kouh, C., Cadieu, M., Knoblich, G., Kreiman, U., Poggio, T.: A theory of object recognition: Computations and circuits in the feedforward path of the ventral stream in primate visual cortex. Technical report, Massachusetts Institute of Technology Computer Science and Artificial Intelligence Laboratory, CBCL-259 (2005)
25. Short, T.: Darwin's concept of final cause: neither new nor trivial. Biology and Philosophy 17, 323–340 (2002), doi:10.1023/A:1020173708395
26. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. Cognitive Psychology 12(1), 97–136 (1980)
27. Ullman, S., Vidal-Naquet, M., Sali, E.: Visual features of intermediate complexity and their use in classification. Nature Neuroscience 5(7), 682–687 (2002)
28. Ungerleider, L.G., Haxby, J.V.: "What" and "where" in the human brain. Current Opinion in Neurobiology 4(2), 157–165 (1994)
29. Ungerleider, L.G., Mishkin, M.: Two Cortical Visual Systems, ch. 18, pp. 549–586 (1982)
30. Walther, D., Koch, C.: Modeling attention to salient proto-objects. Neural Networks 19(9), 1395–1407 (2006)
31. Wolfe, J.M.: Visual Attention, 2nd edn., ch. 8, pp. 335–386. Academic Press (2000)