IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 25, 2019, accepted April 3, 2019, date of publication April 22, 2019, date of current version May 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2912792

Artificial Visual Cortex and Random Search for Object Categorization

GUSTAVO OLAGUE[1], (Senior Member, IEEE), EDDIE CLEMENTE[2], DANIEL E. HERNÁNDEZ[3], AARON BARRERA[1], MARIANA CHAN-LEY[1], AND SAMBIT BAKSHI[4], (Member, IEEE)

[1]Department of Computer Science, Ensenada Center for Scientific Research and Higher Education, Ensenada 22860, México
[2]TecNM, Ensenada Institute of Technology, Ensenada 22780, México
[3]TecNM, Tijuana Institute of Technology, Tijuana 22414, México
[4]Department of Computer Science and Engineering, National Institute of Technology at Rourkela, Rourkela 769008, India

Corresponding author: Gustavo Olague (olague@cicese.mx)

This work was supported by the CONACyT through the Project 155045-"Evolución de Cerebros Artificiales en Visión por Computadora". The work of G. Olague was supported by the Seventh Framework Programme of the European Union through the Marie Curie International Research Staff Scheme under Grant FP7-PEOPLE-2013-IRSES and Grant 612689 ACoBSEC.

ABSTRACT Brain modeling is a research area within computer science devoted to the study of complex and dynamic computing algorithms that imitate brain function regarding the information processing properties of the structures that make up the nervous system. The computational and mathematical structures are composed of interacting modules, whose coordination aims to enhance their problem-solving capabilities. The computational models of the visual cortex use non-trivial interactions between a large number of components. In this paper, we propose a hierarchical structure that mimics the information flow and transformations that take place in the human brain. This paper describes a virtual system composed of an artificial dorsal pathway–or "where" stream–and an artificial ventral pathway–or "what" stream–both are fused to recreate an artificial visual cortex. In previous work, the model was refined through genetic programming to enhance its performance over challenging object recognition tasks. The system finds good solutions during the initial stage of the genetic and evolutionary search. In this paper, the goal is to show that a random search can discover numerous heterogeneous functions that are applied to a hierarchical structure of our virtual brain. Thus, the proposal presents two key ideas: 1) the concept of function composition in combination with a hierarchical structure leads to outstanding object recognition programs, and; 2) multiple random runs of the search process can discover optimal functions. The experimental results provide evidence that high recognition rates could be achieved in well-known object categorization problems; consequently, this paper corroborates the importance of the hierarchical computational structure described in the neuroscience literature.

INDEX TERMS Automatic programming, brain modeling, artificial visual cortex, brain-inspired computing, heuristic computing, deep genetic programming.

I. INTRODUCTION

Object recognition is a fundamental task for humans and all living beings endowed with the sense of sight since it allows the interaction of the organism with the surrounding environment and its understanding. In general, the human visual system can recognize and classify an object according to its category with ease. Both tasks consider that the set of attributes or features extracted from the images are general enough to classify the object as part of the class while maintaining in memory the elements that serve to identify that particular object within a given scene [1]. Although an accurate description about the processes that solve the object recognition problem remains incomplete; there is vast knowledge about the functionality of the primary brain areas involved in the performance of the visual information pathway, which leads to object categorization.

Nowadays, object recognition is said to be involved in two main tasks: the first refers to the goal of identifying an object as a single entity; while the second pertains to the categorization that consists of the arrangement of an object within a group of similar characteristics regardless of its

The associate editor coordinating the review of this manuscript and approving it for publication was Bora Onat.

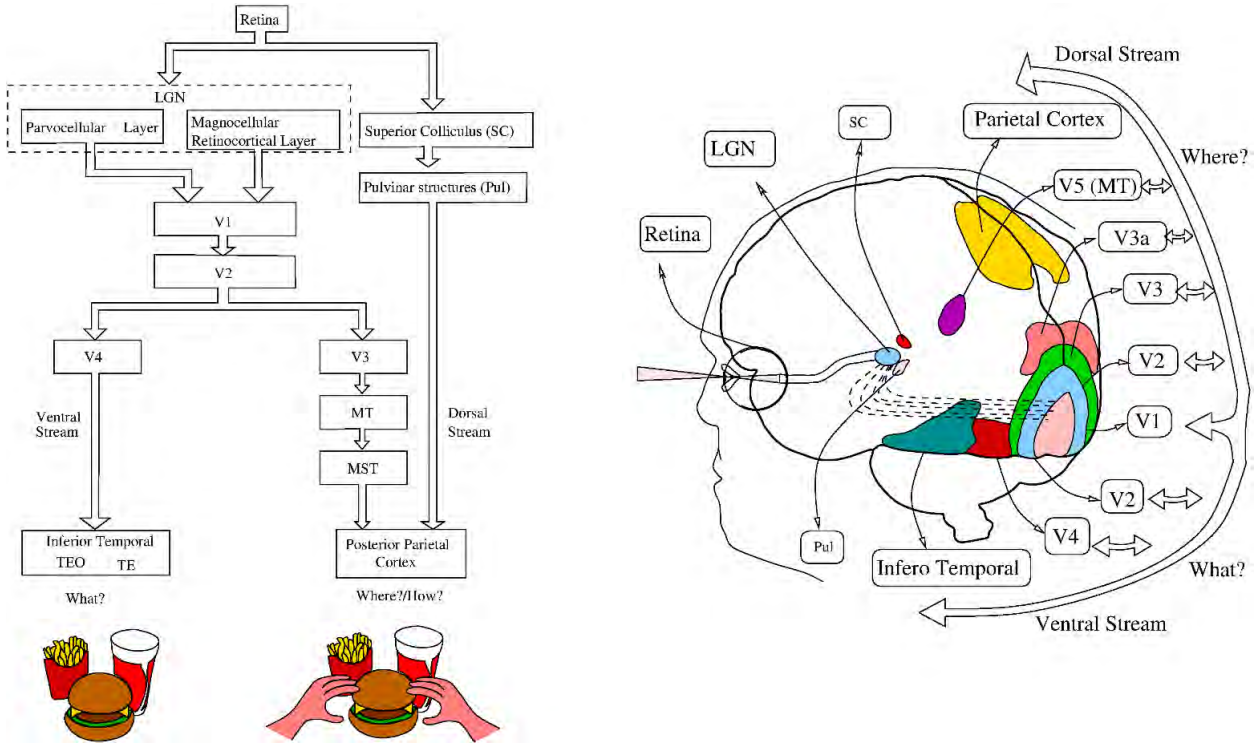

FIGURE 1. **The visual system consists of "what" and "how/where" information processing streams, that are defined as subserving different purposes that achieve highly specialized visual tasks.**

size, location, rotation, viewpoint, lighting conditions, and occlusions. Computationally, the two processes are almost identical since the input to the learning model is an image, while the output is a label describing the membership of the object – depicted on the image – to a given class. Hence, categorization involves a range of possible variations larger than identification because a recognition system must generalize not only across different viewing conditions but also across different exemplars of the class. Thus, object classification is a computationally challenging problem since an artificial vision system must be able to construct a descriptor based on a set of invariant properties of the object that could be useful to classify the image [1]–[3]. The fact that a computational model could be refined with a simple random search of a few critical functions embedded within a hierarchical structure is worthy of attention. This paper provides extensive results that corroborate the importance of such methods. This new modeling can be framed as a goal-driven approach to computer vision [4].

This research attempts to create complex brain models inspired by neuroscience knowledge. Computer simulations of brain-like systems based on the paradigm of genetic programming can lead to powerful new techniques in artificial intelligence [5]. This work is based on the artificial visual cortex (AVC) which shows excellent performance in solving the absence/presence problem of object recognition. The AVC is based on two models: a psychological model called the feature integration theory [6], and a neurophysiological model called the two cortical pathway [7]. This proposal has been extensively tested on different problems like object recognition [8], feature detection [9], visual attention [10], and tracking [11], [12], and it was implemented in the CUDA language [13]. In all these works good solutions were discovered by the brain programming strategy in the first iterations of the algorithm. Therefore, a question about how often those programs are discovered is relevant, since this aspect can be used in future research to devise new strategies to approach more difficult problems.

In the literature, the human visual system is studied as a model that provides insight about how to solve the object recognition problem. The natural system is understood as a rich paradigm where the notion of hierarchical processing across the visual cortex was first proposed by Hubel and Wiesel [14]–[16]. The main idea suggests a feed-forward scheme which performs a series of processes of increased complexity along the receptive fields corresponding to the observed stimuli in simple, complex, and hypercomplex cells derived from studies in the visual cortex of cats [14]. Further studies made by Ungerleider and Mishkin in 1983 proposed the existence of two routes in the visual cortex. These two pathways, called dorsal and ventral streams, have a common origin in the layers of the lateral geniculate nucleus (LGN) and the primary visual cortex [7]. The functionality of the dorsal stream focus on the location of an object within the scene, while the ventral stream is dedicated to the task of object recognition; see Figure 1. An efficient visual functionality is achieved by a great interchange of information

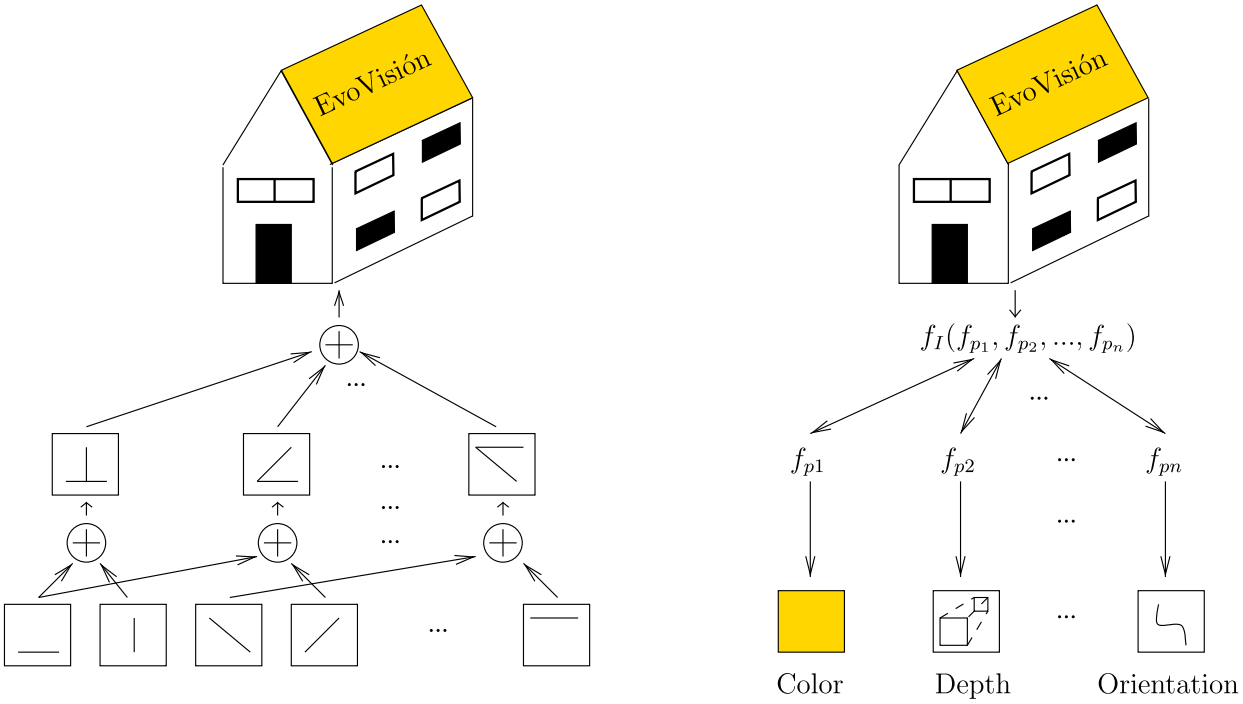

FIGURE 2. Data-driven vs. function-driven. In a function-driven process, a set of visual operators are fused by synthesis to describe the properties of the image.

between the two streams [17]. In this way, object recognition involves processes such as selectivity, defined as the ability to filter unwanted information, as well as those in charge of describing the objects. The approach proposed in this work is divided into three key steps. The first is related to the integration of salient features using four dimensions: color, shape, orientation, and intensity. The second consists of the application of selected mathematical functions commonly used within computer vision and image processing. Finally, these functions are combined into a compound of many operators by synthesis in such a way of identifying salient properties of objects that are useful in the categorization of objects. This approach differs from those of the state-of-the-art where a data-driven principle is applied using a set of patches – image regions – while creating a dictionary of visual words like in a bag-of-words approach [18]–[22]. In our work, the first hypothesis is that the dictionary of visual words can be replaced by a set of visual operators which are built with a group of mathematical functions. The second idea is based on the integration of properties in charge of the visual attention process – or selectivity – that is related to the creation of conspicuity maps and the center surround process together with description and combination of maximum responses executed by a *max* operation of the functions that select features that categorize the object. Contrary to previous biologically plausible models, where these operators rely exclusively on neuroscientific knowledge and whose implementation is based on a data-based paradigm. In this work, we propose to build these operators with a set of operations within a computational structure, in this way, the analogy will focus on the functionality of the visual operator and how is its algorithmic implementation; see Figure 2.

A. RESEARCH CONTRIBUTIONS

This paper provides a thorough insight into the random search and the motivation of applying it to the artificial visual cortex (AVC) in the problem of object recognition. Therefore, extending the first results published at the EvoStar conference [23]. We remark four contributions.

- First, a computational model of the visual cortex is proposed based on the fusion of previous visual attention and object recognition proposals using a hierarchical structure that is similar to previous models.
- Second, a functional approach is enforced through a set of mathematical functions that are specialized in the processes of visual information extraction and description.
- Third, in the proposed method the total number of visual operators made of mathematical functions and embedded within the hierarchical structure can be discovered through a few random trials while achieving outstanding results on two of the standard testbeds.
- Fourth, as a byproduct of the approach, the total number of computational operations that are used to classify an image is relatively small in comparison with similar strategies that are based on the application of image patches.

This paper is organized as follows. Section II is devoted to the task of reviewing the state-of-the-art. Then, Section III provides a general description of our approach. Later, Section IV presents experimental results. Finally, Section V

gives our conclusions while offering some suggestions for possible future work.

II. RELATED WORK

In the twentieth century with the arrival of digital computers, several works attempted to emulate the functionality of the human visual cortex to perform tasks like object recognition, visual attention and object detection. This section provides a list of relevant works to outline the state-of-the-art of these biologically-inspired computational approaches.

From a computational standpoint, the first work dealing with object recognition was developed in 1980 by Fukushima, who proposed a neural model called Neocognitron to solve this task [18]. In their work, the computational structure was inspired by the visual nervous system and the hierarchical model that was first discovered by Hubel and Wiesel [14]. This model was capable of recognizing letters and numbers considering a shift in position. Later, Biederman in 1987 suggested the recognition-by-components (RBC) theory, where an object can be recognized by the combination of elements called geons: blocks, cylinders, and funnels or truncated cones that are similar to phonemes in a human language [24]. These approaches were extended by Perrett and Oram through a model based on ''patterns'' made of simple conjunctions of 2-D elements, that increase their complexity along a series of stages, by mimicking the cell properties of the ventral cortical stream [25]. It is noteworthy that this hierarchical model is invariant to rotation and size transformations of an object. Then, Ullman and Soloviev in 1999 proposed the conjunction of multiple overlapping image fragments – or visual patterns – to achieve shift invariance for complex shapes [26]. This work was later extended to classification [27]. In the same year, Riesenhuber and Poggio introduced a hierarchical feedforward architecture with similar matching and pooling stages as the Neocognitron, but with the incorporation of the *max* operation as a better model of the complex cell in contrast to a linear summation [28]. This method provided a robust response to position invariance while arguing that its functionality is biologically plausible. Later, the model was tested on the recognition of artificial paper-clip images and was improved in [19], [20] to achieve a robust object-recognition performance through a universal dictionary of features. Along with this line of research, several works proposed to optimize the number of patches or elements, as well as to improve the description of the object following the hierarchical model [29].

During the same period, many computational visual attention systems arose based on similar hierarchical structures that were adapted from the psychological theory of feature-integration proposed by Treisman and Gelade [6], which suggests that attention must be processed at two successive stages. The first called preattentive stage that is computed in parallel along several feature dimensions of the scene such as shape, color, orientation, spatial frequency, brightness, and direction of movement. Then, a second stage called focal attention provides the integration of the initially separable features into unitary objects. Afterward, Koch and Ullman proposed the construction of a *saliency map* using a neuronal network process called winner-take-all, which combines the information of the feature maps and provides as output the most conspicuous locations of the scene [30]. Later, Milanese proposed a visual attention system based on the models of [30] and [31], which uses filters as operations to compute two color opponencies: red-green and blue-yellow; with 16 different orientations and local curvature information [32]. These operations define the feature maps that are later transformed by the application of a *conspicuity operator*, which is motivated by the on-off cells in the cortex. This operator is usually referred to as the center-surround mechanism that is applied to define the so-called *conspicuity map*; a term that is frequently used to denote the feature-dependent prominence. Finally, the conspicuity maps are integrated into a *saliency map* by a relaxation process that identifies a small number of convex regions. Along with this line of research, Itti *et al.* proposed a model that is widely used since it encapsulates the ideas of [30] and [32]. The main contribution is the implementation of theoretical concepts of the visual attention process and its application to artificial and real-world scenes [22]. This technique enables the detection of feature dimensions at different scales followed by the center-surround mechanism.

Today few works attempted to integrate the two approaches. Fukushima in 1987 implemented a hierarchical neural network that serves as a model for selective attention and objects recognition [33]. When several patterns are presented simultaneously, the model performs discriminatory attention to each one, segmenting it from the rest while recognizing it separately. Afterward, Olshausen *et al.* in 1993 defined a biologically plausible model that combines attentional mechanism and object recognition processes to form position and scale invariant representations of the visual world [34]. Then, Walther *et al.* suggested a combined model for spatial attention and object recognition [35]. In their work, visual attention follows the computational model proposed by Itti and Koch [22] and object recognition is achieved through the HMAX model of Riesenhuber and Poggio [28]. This model was applied to the problem of recognizing artificial paperclips. Later, Walther and Koch in 2007 suggested, with a computational model, that features learned by the HMAX model used for the recognition of a particular object category may also serve for top-down attention tasks [36]. Finally, Heinke and Humphreys applied a model called SAIM for the visual search involving simple lines and letters [37]. This model, in a first stage, selects the object within the image and subsequently performs an object identification step using a template matching technique.

In our work, we propose a new hierarchical model following the preattentive stage of visual attention described in [6], [30] to locate conspicuity regions within the image. Then, a description process is performed using the max operator in combination with a series of functions that emulate the functionality of the V4 area in the visual cortex.

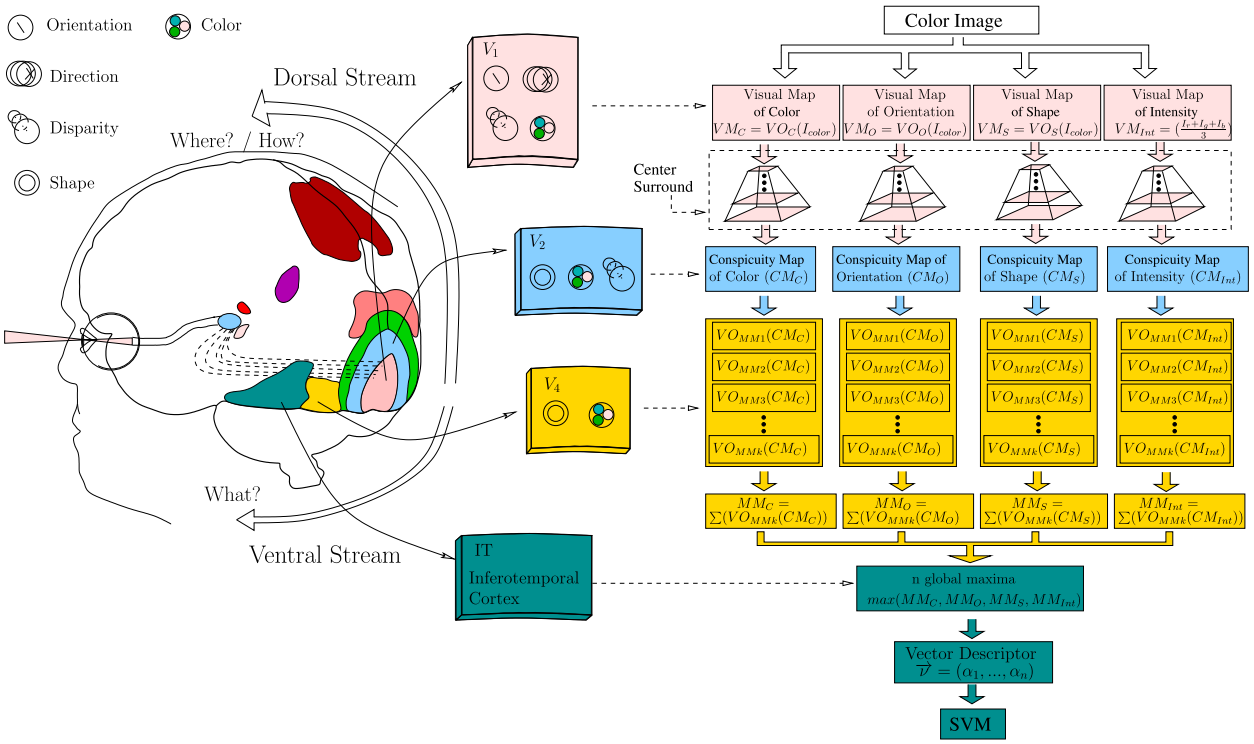

FIGURE 3. Conceptual model of the artificial visual cortex. The color image is decomposed into four dimensions (color, orientation, shape, and intensity). Then, a hierarchical structure is charged with solving the object classification problem through a function-driven paradigm.

This approach differs from traditional models to object recognition [1], [19], [20], [28], [36], [37] where a set of patches – or visual words – are used to identify the object. In our proposed approach the discovered functions provide the functionality of multiple patches; hence, helping in the creation of a straight-forward process as will be shown in the experimental results.

III. THE AVC ALGORITHM

This section aims to describe our artificial visual cortex approach tested on three object recognition tasks with increasing difficulty. The idea follows the analogy of the hierarchical processing of visual information performed by the brain to classify an object. The section is organized as follows. The purposeful approaches are briefly outlined in order to understand the two principal pathways of the natural visual system. Then, the AVC algorithm is detailed to give an account of our proposal.

A. AVC AS A GOAL-DRIVEN PROCESS

The natural visual system is composed of two main pathways defined by the dorsal and ventral streams [7], [38]–[41]. These two pathways share the first stages of the visual information processing located at layers $V1$ and $V2$. Later, the streams diverge intending to subserve two different tasks that are specially contrived to achieve the object and spatial vision; see Figure 3. The classical dichotomy between object and spatial perception focuses on the importance of the purposeful representation that serves a single or general task. Furthermore, the "what" and "where/how" theory of Milner and Goodale [38] gives also an emphasis on the hypothesis that the visual system is defined according to the requirements of the task that each stream subserves. Thus, the idea is to define multiple frames of reference giving special attention to the goal of the observer. The object, as well as the spatial information, are transformed by the visual system for different purposes. Thus, the ventral system along the pathway: V1, V2, V4, and IT areas, represents the visual world in allocentric coordinates by promoting conscious perceptual awareness; while, the dorsal stream along the visual route: V1, V2, V3, V3a, V5, and Parietal Cortex areas use egocentric coordinates to transform information about the object's location, orientation, and size [42]. Consequently, the problem of object recognition suggests an integrative action of the dorsal and ventral streams [38], [43]. In this manner, the goal of this work is to emulate the functionality for acquisition and transformation of features at several dimensions, as well as the description of regions within an image by imitating the stages performed by the dorsal and ventral streams.

These tasks are emulated with a set of mathematical functions specialized in obtaining visual information from images in four different dimensions. We execute a random search of the set of functions called visual operators (*VOs*). The AVC is divided into two main parts. In the first stage, the proposed system executes the acquisition and transformation of features. Then, in a second stage, the AVC performs the description and classification of objects.

B. ACQUISITION AND TRANSFORMATION OF FEATURES

The early stage of the system follows the psychological model of visual attention proposed by Treisman and Gelade [6], which was successfully implemented in [22]. The image acquired with the camera represents the first step of our algorithm. The system considers digital color images in the RGB color model, which are also transformed into the CMYK and HSV color models. The idea is to build the set $I_{color} = \{I_r, I_g, I_b, I_c, I_m, I_y, I_k, I_h, I_s, I_v\}$, which corresponds to the *red, green, blue, cyan, magenta, yellow, black, hue, saturation, and value* components of their respective color models and which are used to provide the initial representation of the scene.

1) FEATURE DIMENSIONS

Each VO is defined as a mapping $VO_d : I_{color} \rightarrow VM_d$; where the set I_{color} gives the input to the visual operator and the output corresponds to a visual map (VM_d) for a particular feature dimension. The transformations are performed to recreate the feature extraction process of the brain; resulting into a visual map (VM) per dimension. The VOs define specific image features along with several dimensions: color, shape, orientation, and intensity, $d \in \{C, S, O, Int\}$. Next, we explain these features.

- Color dimension. The goal of this process is to highlight prominent regions associated with color properties on the image. In the natural system, this operation is carried out by the retina, where color opponencies are estimated and further processed in the V1, V2, and V4 brain areas [44]. Input color images are transformed through the color visual operator VO_C to find prominent regions based on the color dimension. The mapping in the computational model is represented as follows.

$$VO_C : I_{color} \rightarrow VM_C, \tag{1}$$

  where VM_C is the visual color map representing the prominence of pixels in color. In this work, color feature extraction is performed through function composition of multiple operators. Note that we are considering two special functions to compute color opponency values according to the proposal explained in [45]. Moreover, a third function known as the image complement is applied in such a way that each pixel value is subtracted from the maximum and the difference is used as the final result in the output image.
- Shape dimension. The method that extracts visual information on the shape of the object uses morphological information of the object. In nature, such functionality is carried out in areas of the brain such as V2 and the temporal cortex [46]–[50]. The goal of extracting shape information is to highlight morphological information that can be used for object recognition. The mapping in the computational model is represented as follows.

$$VO_S : I_{color} \rightarrow VM_S, \tag{2}$$

  Note that the application of this mathematical tool can be considered as the first implementation of such concepts within the analogy of the artificial visual cortex. We propose to create compound operators by the composition of four basic morphological operations known as erosion, dilation, opening, and closing. Indeed, more complex operators can be created, from these simple ones, like the hit-or-miss transform, skeleton, perimeter, top-hat, bottom-hat, and others [51].
- Orientation dimension. The composition of orientation characteristics determined the edge and corner operators applied to an image. These operators emulate the functionality of the simple and complex cells present in the primary visual cortex. The orientation features are highlighted to detect borders and junctions on the image similar to the evolution of interest point detectors and descriptors [52], [53]. The mapping performed by VO_O is defined as follows.

$$VO_O : I_{color} \rightarrow VM_O, \tag{3}$$

  where VM_O corresponds to the visual map for the orientation attribute. Numerous functions were applied together with the Gaussian derivative function proposed in [54] and Gaussian smoothing filters with $\sigma = \{1, 2\}$.
- Intensity dimension. Finally, the intensity measure corresponds to the amount of light perceived by a photosensitive device. In humans, the intensity is measured by specialized ganglion cells in the retina [6], [17]. In order to compute the intensity, the following formula is applied.

$$VM_{Int} = \frac{I_r + I_g + I_b}{3},$$

  where I_r, I_g, and I_b are the color bands of the image, while VM_{Int} is the intensity of the visual map [32], [36].

2) CENTER SURROUND PROCESS

The center-surround method is based on the functionality of the ganglion cells, located in the retina and lateral geniculate nucleus, that measures the difference between the firing rates at the center and surrounding areas of their receptive fields. The goal of this process is to generate a conspicuity map (CM) per dimension according to the model proposed in [45]. The algorithm consists of a two-step process where the information is built to emulate its natural counterpart as follows. First, the computation of the CMs is modeled as the difference between fine and coarse scales, which are computed through a pyramid of nine levels $P_d^\sigma = \{P_d^{\sigma=0}, P_d^{\sigma=1}, P_d^{\sigma=2}, P_d^{\sigma=3}, \ldots, P_d^{\sigma=8}\}$. Each pyramid is calculated from its corresponding VM_d using a Gaussian smoothing filter resulting in an image that is half of the input map size, and the process is repeated recursively eight times to complete the nine-level pyramid. Second, the pyramid P_d^σ is used as input to a center surround procedure to derive six new maps that result from

the difference between some of the pyramid levels calculated as follows.

$$Q_d^j = P_d^{\sigma=\lfloor\frac{j+9}{2}\rfloor+1} - P_d^{\sigma=\lfloor\frac{j+2}{2}\rfloor+1},$$

where $j = \{1, 2, \ldots, 6\}$. Note that the levels of P_d^σ have different size and are scaled down to the size of the top level to calculate their difference. Next, each of these six maps is normalized and combined into a unique map through the summation operation, which is then normalized and scaled up to the VM_d maps' original size using a polynomial interpolation to define the final CM_d.

C. DESCRIPTION AND CLASSIFICATION STAGE

After the construction of the *CMs*, the next stage along the AVC is to define a descriptor vector used as input to an SVM model for classification purposes. This stage is analogous to the functionality of the $V4$ layer, as well as the Inferotemporal Cortex (IT) since it is said that these two regions perform the classification stage.

1) COMPUTATION OF THE MENTAL MAPS

In the natural system, the V4 area of the visual cortex is distinguished by responding to complex stimuli of orientation, spatial frequency, as well as to forms such as spirals and complex patterns [55], [56]. With this, our analogy consists of building a map that discriminates the unwanted information from the conspicuous maps and only focuses on the object of the image that is to be classified, enhancing the characteristics of that object. This map is called a mental map. In this stage of the process a single set of visual operators is used to produce a mental map (MM_d) per dimension. After the computation of the conspicuity maps, a set of visual operators VO_{MM} is applied to describe the image content. Note that the proposed visual operators are homogeneous and independently applied to each feature dimension. This operation is defined as follows:

$$MM_d = \sum_{i=1}^{k} (VO_{MM_i}(CM_d)), \tag{4}$$

where d is the dimension index, and k represents the cardinality of the set VO_{MM}. Each summation is applied to integrate the output of all operations VO_{MM_k} to produce a MM_d per dimension. After that, the four Mental Maps are concatenated into a single array, and the n highest values are selected to define the vector $\overrightarrow{v}$ that describes the image. The input to these operators is the corresponding conspicuity map per dimension.

In contrast to our proposal, well-known methodologies [18]–[22] are based on a template matching paradigm to learn a set of prototype image patches. Hence, our approach substitutes the set of templates with the set of visual operators to characterize one object class with excellent results as we will show in the experiments. Note also that the proposed brain modeling is very different from current proposals like deep learning, where the models correspond to networks of artificial cells grouped in multiple layers [18], [28], [91].

2) LABEL ASSIGNMENT

In the natural visual system, the response of the V4 area is connected to the inferotemporal of the brain (IT) whose response is selectively activated to the observed object, showing invariance to transformations such as scale, position, and orientation. That is, the IT area exhibits the ability to carry out the task of recognizing objects from the visual stimulus it receives [57]–[61]. In the present work, the computational analogy is a classifier, implemented with a support vector machine (SVM). Therefore, an SVM is trained to learn a mapping $f(\mathbf{x})$ that associates descriptors $\mathbf{x}_i$ to labels y_i. Our problem is formulated in terms of a binary classification task, whose main aim is to find a decision surface that best separates the elements of the class. In this work, we use a non-linear SVM working with the discriminate hyperplane defined by:

$$f(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b, \tag{5}$$

where the given training data is $(\mathbf{x}_i, y_i)$, $i = 1, \ldots, l$, $y_i \in \{-1, 1\}$, $\mathbf{x}_i \in \mathbf{R}^p$ and $K(\mathbf{x}_i, \mathbf{x})$ is the kernel function. The sign of the output indicates the class membership of $\mathbf{x}$. Thus, finding the best hyperplane is performed through an optimization process that uses the margin between the class and non-class as the search criteria.

IV. EXPERIMENTS AND RESULTS

Experimentation was carried out to provide evidence to support the claim that efficient and reliable solutions, for not trivial recognition problems are discovered through random search. Note that despite using random search the hierarchical structure is not random, but it follows well-designed models in principle way that give coherence to the proposed algorithm. This section is organized as follows. Firstly, the experimental results are given using the simplest database. Secondly, a comparison is made with other methodologies. Finally, conclusions and future work are drawn about the work.

A. EXPERIMENTAL DESIGN

This section presents the experimental settings designed to show the advantage of applying a random search instead of an evolutionary search. It describes the databases used during training and testing. Several experimental tests are presented together with the best results achieved by our approach. Note that we use the CalTech 5 and CalTech 101 image databases, despite serious concerns raised about them [2], [62]. Nevertheless, that test is still widely used in the object recognition community, and many state-of-the-art algorithms report their classification results with it [63]–[66]. Nonetheless, in order to show the effectiveness of the proposed model, the experiments also include

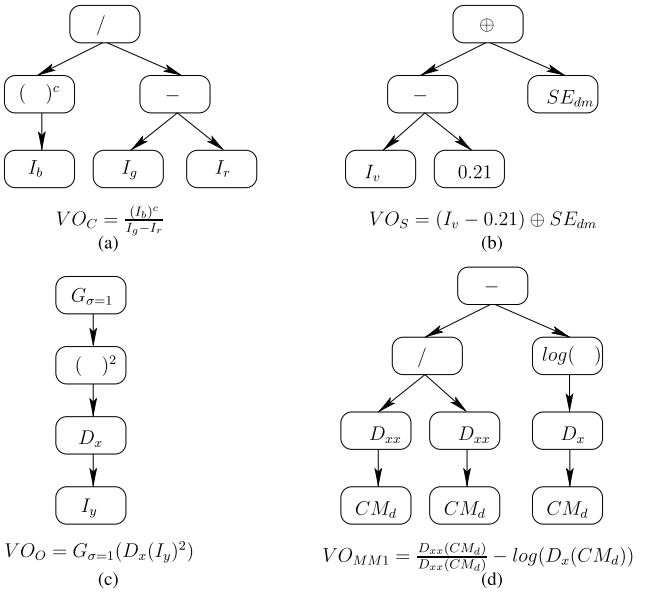

FIGURE 4. **These diagrams depict a set of syntax trees that were used within the AVC_{M1} solution reported later.**

a more challenging dataset GRAZ, similar to the following works: [19], [20], [67]–[80].

1) METHODOLOGY TO OBTAIN AN AVC SOLUTION

The methodology used to generate AVC programs followed the algorithm of Section III, where an important step is the construction of *VOs*. Note that such mathematical and computational functions are susceptible to being discovered through some optimization approach [5]. During the development of the proposal, we remark that brain programming unexpectedly found good solutions in the first iterations of the algorithm. Hence, a study about the frequency and quality of solutions applying a random search can provide valuable information for future research using this kind of models. The operators consist of syntax trees made of internal and leaf nodes, which are defined by a set of primitive elements called function and terminal sets defined before the initial run. In our work, each tree has its own sets of functions and terminals carefully chosen according to the desired functionality that we attempt to emulate within the AVC. All *VOs* were generated through a random procedure with a maximum depth of 5 levels, where half of the trees were balance trees and the other half were arbitrary trees adding nodes until the maximum depth is reached. As a result, the approach generates random tree structures with some branches longer than others. Figure 4 shows the solution AVC_{M1} from Table 5, which is provided here as an example of the *VOs* using syntax tree representation. The implementation was programmed in MATLAB running on a Dell Precision T7500 workstation, Intel Xeon eight-core, CPU E5506 at 2.13GHz, NVIDIA Quadro FX3800 and Linux OpenSuse 11.1 operating system.

The methodology to study the absent/present classification problem is divided into three steps. The first two steps define the training stage while the last is devoted to the testing stage.

TABLE 1. **Functions for the visual operators (*VOs*).**

Function	Description
$A+B, A-B, A\times B, A/B$	Arithmetic functions between two images A and B
$log(A), exp(A)$	Transcendental functions over the image A
$(A)^2$	Square function over the image A
$\sqrt{A}$	Square root function over the image A
$(A)^c$	Image complement over the image A
$Op_{r-g}(I)$, $Op_{b-y}(I)$	Color opponencies Red - Green and Blue - Yellow
$thr(A)$	Dynamic threshold function over the image A
$k+A, k-A, k\times A, A/k$	Arithmetic functions between an image A and a constant k
$round(A)$, $half$, $\lfloor A\rfloor$, $\lceil A\rceil$	Round, half, floor and ceil functions over the image A
$A\oplus SE_d$, $A\oplus SE_s$, $A\oplus SE_{dm}$	Dilation operator with disk, square, and diamond structure element (SE)
$A\ominus SE_d$, $A\ominus SE_s$, $A\ominus SE_{dm}$	Erosion operator with disk, square, and diamond structure element (SE)
$Sk(A)$	Skeleton operator over the image A
$Perim(A)$	Find perimeter of objects in the image A
$A\circledast SE_d$, $A\circledast SE_s$, $A\circledast SE_{dm}$	Hit or miss transformation with disk, square, and diamond structures
$T_{hat}(A)$, $B_{hat}(A)$	Performs morphological top-hat and bottom-hat filtering over the image A
$A\circledcirc SE_s$, $A\odot SE_s$	Opening and closing morphological operators on A
$\lvert A\rvert, \lvert A+B\rvert, \lvert A-B\rvert$	Absolute value applied to A, and the addition and subtraction operators
$inf(A,B)$, $sup(A,B)$	Infimum and supremum functions between the images A and B
$G_{\sigma=1}(A), G_{\sigma=2}(A)$	Convolution of the image A and a Gaussian filter with $\sigma=1$ or 2
$D_x(A), D_y(A)$	Derivative of the image A along direction x and y

In this way, all image databases were randomly divided into three subsets for each class, in such a way of applying each subset to each step. This process is detailed next.

1) The process starts by randomly generating a set of *VOs* to be used inside the AVC structure. Table 1 provides a set of functions that are used to create the *VOs*. Then, it proceeds to the training stage of the SVM using images from the first subset, called training-A. If the SVM achieves a given threshold in classification accuracy, the process continues to step 2; otherwise, the *VOs* together with the SVM are discarded, and the process is restarted.
2) Next, the system uses the set of *VOs* found in step 1 while training a new SVM with the second image subset, called training-B. Once again, if the SVM scores above the given threshold in accuracy the process continues to step 3 and the AVC structure is the solution; on the other hand, both *VOs* and SVM are discarded, and the search continues at step 1.
3) In the last step, the best AVC structure is tested by classifying the third image subset. The testing is performed with the estimated SVM from step 2 and the *VOs* from step 1. The whole process is repeated until the best set of solutions is discovered.

2) IMAGE DATABASES

The performance of the AVC was evaluated through a binary test using five classes from the Caltech-5 database in

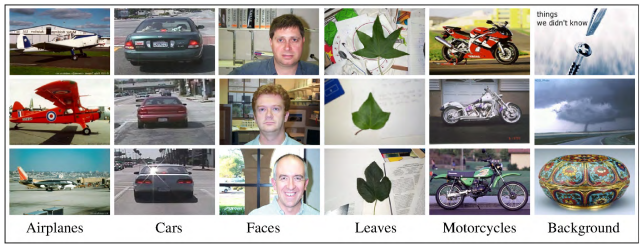

FIGURE 5. Sample images from CalTech-5 database, and the category background from CalTech-101 database.

FIGURE 6. Sample images from background category in CalTech-5 database.

combination with the Google background of Caltech-101, see [81], [82]. We select as the positive classes from CalTech-5: airplanes, cars, faces, leaves, and motorcycles, along with two different background image classes from CalTech-5 and CalTech-101 to use them as the negative class for two experiments described below. Figures 5 and 6 show some sample images of the databases that are commonly used to evaluate state-of-the-art systems. The reason to select these databases is that the objects are on the foreground and centered in the image making it an excellent test environment for our approach. All experiments were carried out with images of 140 pixels in height. In the case of images with different size all were rescaled to the proper height preserving the aspect ratio.

3) EXPERIMENTAL EVALUATION OF THE AVC FOR CLASSIFICATION OF COLOR IMAGES

The goal in this experiment is to analyze the effect on the recognition performance by using training sets of different sizes. Thus, the AVC model was trained with randomly selected subsets (positive images) of size: 1, 10, 20, 30, 40, 50, 60, and 70; while using a constant subset of 50 negative images. In the case that the AVC solution never passes the test, after 7500 random runs, the solution was discarded from further tests. Then, the numbers of images selected for training-B were set to 50 positive images and 50 negative images. Table 2 provides the number of random runs that were necessary to discover 100 solutions giving a total of 700 solutions with 100% accuracy. All solutions were tested, and the mean and standard deviation are reported in the following section.

Testing the Performance of the Random Search

This experiment aimed to evaluate the AVC performance from the standpoint of a random search. Figure 7 shows the results of 3500 solutions of five classes, where the *x-axis* indicates the number of training images per class, while the *y-axis* shows the number of solutions, and finally, the *z-axis* provides the accuracy achieved during the testing stage. The size of subsets during testing were 50 positive images and 50 negative images. The best solutions were obtained with airplanes, faces and leaves classes scoring 95%, 99%, and 97% respectively; while for cars and motorcycles, the best solutions scored a classification accuracy of 77% and 75% respectively. Note that these final scores are similar regardless of the subset size that is applied during the training stage. These solutions are provided with their corresponding formulae in Table 5. The summary of this experiment is given in Table 3.

Looking for the Optimal Size of the Training Set

In this section, we address the question of what proportion of samples should be used in the training set. Also, how this

TABLE 2. Total number of random runs needed to discover one hundred solutions per class for all subset sizes.

Class	Size of the training set						
	10	20	30	40	50	60	70
Airplanes	2501	3219	1916	2088	1993	1971	3253
Cars	7294	11032	6652	10674	4447	4845	22037
Faces	1811	3041	1489	2148	1556	1462	1940
Leaves	1781	1960	1392	1843	1893	1355	1763
Motorcycles	10419	23131	12871	20214	9386	6662	39470

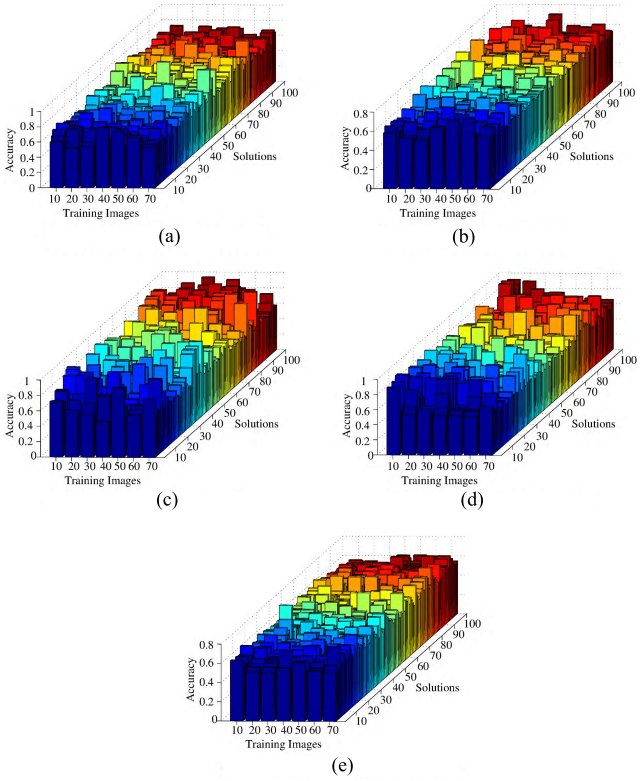

FIGURE 7. These figures show the solutions' performance for different sizes of the training set. Each bar corresponds to a solution depicted in the form of a bar chart making a total of seven hundred solutions per figure.

TABLE 3. This table shows a summary of the results of the AVC testing which was obtained with a random search.

Class	No. Images on Training	Mean	Std	Max	Min
Airplane vs Google background	10	0.6409	0.0854	0.8700	0.5000
	20	0.6122	0.0800	0.8300	0.5000
	30	0.6284	0.0938	0.8800	0.4500
	40	0.6171	0.0927	**0.9500**	0.5000
	50	0.6104	0.0802	0.8800	0.4900
	60	0.5774	0.0715	0.8300	0.4900
	70	0.6141	0.0841	0.9200	0.5000
Cars vs Google background	10	0.5751	0.0643	0.7600	0.4600
	20	0.5523	0.0597	**0.7700**	0.4500
	30	0.5626	0.0583	0.7100	0.4700
	40	0.5703	0.0607	0.7500	0.4500
	50	0.5575	0.0561	0.7400	0.4600
	60	0.5433	0.0510	0.7400	0.4600
	70	0.5540	0.0495	0.7300	0.4700
Faces vs Google background	10	0.6696	0.1501	0.9800	0.4800
	20	0.6728	0.1413	0.9900	0.4500
	30	0.6572	0.1350	0.9800	0.5000
	40	0.6453	0.1288	0.9800	0.4500
	50	0.6494	0.1491	0.9800	0.5000
	60	0.5870	0.1072	**0.9900**	0.4600
	70	0.6201	0.1200	0.9800	0.5000
Leaves vs Google background	10	0.6887	0.1433	0.9700	0.4800
	20	0.6705	0.1290	0.9500	0.4800
	30	0.6446	0.1210	0.9500	0.5000
	40	0.6585	0.1311	**0.9700**	0.4900
	50	0.5920	0.0939	0.9300	0.4800
	60	0.6225	0.1198	0.9700	0.4800
	70	0.6181	0.0990	0.9400	0.5000
Motorcycles vs Google background	10	0.5812	0.0507	0.7500	0.4600
	20	0.5813	0.0690	**0.7500**	0.4600
	30	0.5630	0.0519	0.7000	0.4400
	40	0.5572	0.0530	0.7100	0.4700
	50	0.5496	0.0489	0.7400	0.4700
	60	0.5448	0.0477	0.7000	0.4600
	70	0.5764	0.0620	0.7400	0.4700

proportion impacts the classification accuracy of the SVM process. In the search for an optimal method, we need to devise a strategy where the number of training images can be adjusted to the optimal sample size. Hence, the strategy to look for an optimal sample size involves two interrelated aspects known as frequency and period. Frequency is equal to the number of perfect solutions, in this case, 100 divided by the total number of random trials, while the period corresponds to the average interval between each perfect solution. Indeed, the period is reciprocal of the frequency. The definition of optimal sample size can be computed throughout the variation on the number of training images concerning the occurrence of perfect solutions for each class, which as a consequence produces a different rate on the frequency or period. Figure 8 provides the mean period and standard deviation as well as the frequency of solutions using several training image sets for each class. Note that all charts were scaled along the vertical axis for readability purposes. The frequency of solutions during training was generally higher for faces and leaves classes. These two classes present the lowest uncertainty. Moreover, in the case of 60 training images, the average period was the lowest of all classes, and we can say that this is the optimal training size. The motorcycle class represents the hardest problem since it scores the most significant uncertainty.

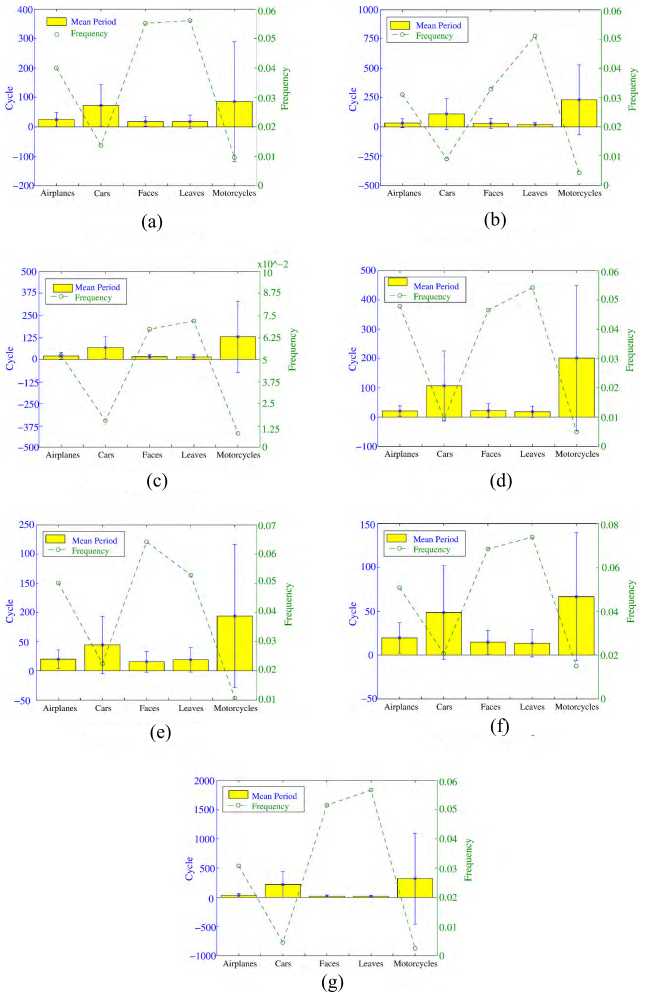

FIGURE 8. These graphics provide the computational effort required to find a solution for each class. The bar chart provides the mean period and standard deviation in terms of cycles or random runs that were required to find an AVC with 100% accuracy during training, while the dashed line represents the frequency required to discover a perfect solution.

TABLE 4. Average number of mental maps over the one hundred solutions per class for all subset sizes.

Size of training set	Class				
	Airplanes	Cars	Faces	Leaves	Motorcycles
10	6.4 ± 3.2	6.3 ± 3.4	5.8 ± 3.1	6.1 ± 3.0	5.4 ± 3.9
20	7.1 ± 3.2	6.3 ± 3.2	6.1 ± 3.2	6.7 ± 3.2	6.4 ± 3.2
30	6.5 ± 3.4	6.5 ± 3.2	6.7 ± 3.3	6.6 ± 3.5	6.5 ± 3.3
40	6.8 ± 3.0	6.7 ± 3.5	6.5 ± 3.1	5.7 ± 3.0	7.0 ± 3.5
50	6.6 ± 3.1	6.3 ± 3.6	6.7 ± 3.1	6.6 ± 3.4	6.4 ± 3.2
60	6.7 ± 3.4	6.5 ± 3.5	6.4 ± 3.2	7.1 ± 3.2	6.1 ± 3.2
70	6.7 ± 3.3	6.4 ± 3.1	5.9 ± 3.1	6.5 ± 3.4	5.8 ± 3.5

Description of AVC Solutions from a Structural Standpoint

This section describes the functionality of the AVC regarding the structure. It brings an analysis of the utility of functions that comprise the *VOs*. The aim is to discover the most useful functions that provide improved performance during classification for each image category. An analysis is provided to describe the complexity and diversity of solutions.

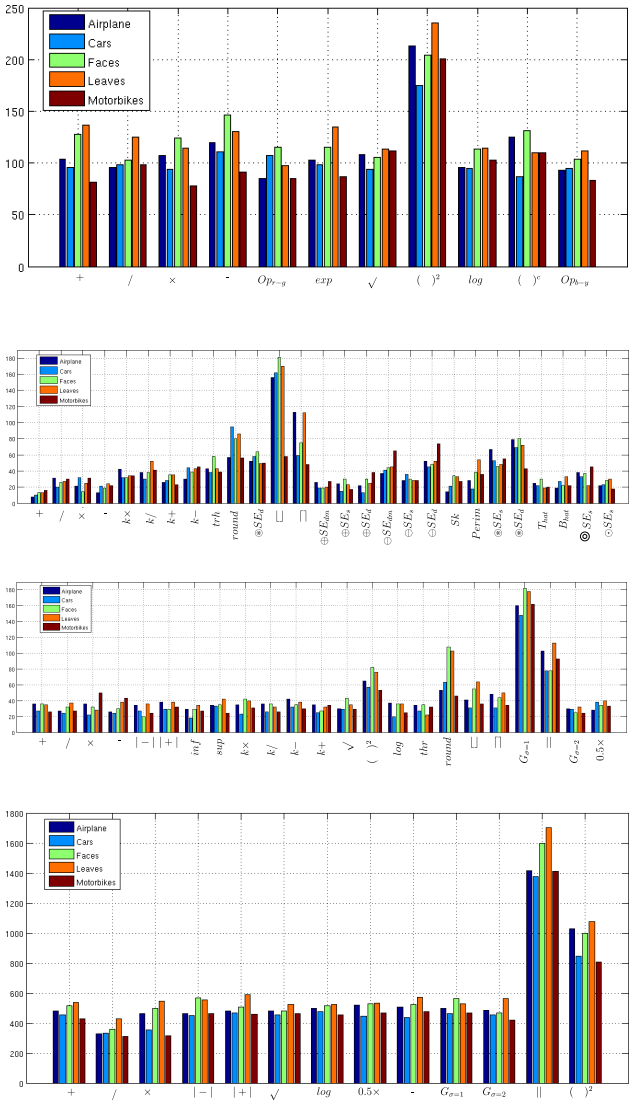

FIGURE 9. The figures show the frequency-of-use of the functions used by the *VOs* along each feature dimension. The charts from top to bottom depict the results for color, shape, orientation, and mental map operations respectively.

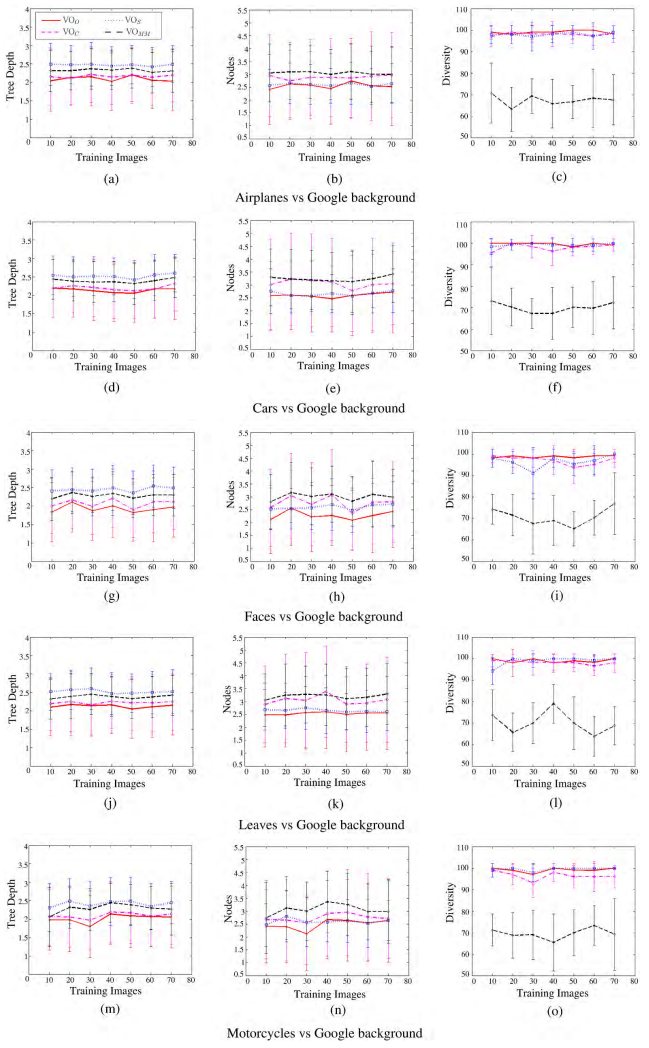

FIGURE 10. The tree representation of the *VOs* can be analyzed from a structural standpoint through their complexity and diversity. The figures on the first and second columns provide the average and standard deviation regarding the tree depth and the number of nodes respectively for different sizes of the training set. Note that the third column depicts the diversity measured as the percentage of the uniqueness of the visual operators' overall solutions.

Frequency-of-use was computed with 3500 solutions, using five classes, by counting the number of appearances of each function in the *VOs* within the AVC, see Figure 9. It is remarkable that, regardless of the image category, the patterns that arise through the computation of the frequency-of-use are very similar along the four *VOs* probably due to the hierarchical process of the AVC. We observe that the quadratic function was commonly applied along the color dimension; see the first graph of Figure 9, while functions like floor, ceiling, round, and hit-miss achieved the highest frequency-of-use for the shape dimension; see the second graph of Figure 9. In the same way, the absolute value, round, Gaussian-blur or convolution against a Gaussian kernel with $\sigma = 1$, and the quadratic functions were primarily used along the orientation dimension; see the third row of Figure 9. Finally, the compound of *VOs* for the mental maps include the quadratic and absolute value functions; see the last graph of Figure 9. For clarity in the charts, the derivatives along the x and y directions were omitted because of their frequency-of-use is very high, since their natural response improves the detection of pixel regions with high variability. We can say that there is no particular set of functions, which classifies a specific image category. Nevertheless, a reduced set of functions was able to classify the five classes. In other words, the response of the functions throughout the AVC structure is different for each input class, which implies a more natural characterization of the system.

In a second experiment, the size and variability of solutions are analyzed through a statistical study of the complexity and diversity of *VOs*. The complexity is measured using depth and number of nodes, while diversity is defined as the percentage of the uniqueness of operators overall solutions;

see Figure 10. Note that the resulting complexity is bounded, irrespective of category and number of training images. Similarly, diversity is above 60%, regardless of category and number of training images. The *VOs* are different in almost every solution, which is consistent with the random search that was used to find them. It is noteworthy that this random process provides solutions despite the large search space. Finally, the length of a solution is calculated as the total number of visual operators, which is determined by the number of mental maps; see Table 4. Note also that the average number of mental maps is constant in general regardless of category and number of training images.

Size of the AVC Search Space

The result of analyzing the solutions from a structural standpoint can be used to devise a solution subspace that is smaller than the search space defined initially by all operations. Thus, the size of this feasible region can be calculated as follows.

The search space is defined as the number of all possible solutions that are achievable through the combination of all possible *VOs*. Its size can be obtained with the set of functions and terminals described in Table 1. In this way, given a particular tree structure i, nT terminals, nF_U unary functions, and nF_B binary functions; the number of possible visual operators nVO_i is calculated as follows:

$$nVO_i = nT^{nln} \times nF_U^{npn_1} \times nF_B^{npn_2},$$

where nln is the number of leaf nodes, npn_1 is the number of parent nodes with one child node, and npn_2 is the number of parent nodes with two child nodes. Hence, the search space S_s is the result of multiplying all possible combinations of visual operators for all dimensions and mental maps. This can be written as follows:

$$S_s = \sum_{i=1}^{l}(nVO_C)_i \times \sum_{i=1}^{l}(nVO_S)_i \times \sum_{i=1}^{l}(nVO_O)_i \times \Big(\sum_{i=1}^{l}(nVO_{MM})_i\Big)^k,$$

where l is given by the depth of the visual operator and k is the number of visual maps. Thus, the size of the search space is around 4.72×10^{87} solutions. Note that the results of the previous statistical analysis could be applied to create a smaller search space. Hence, the search space could be reduced to a new subspace of 1.5×10^{29} solutions. It is remarkable how easy it was to find solutions that score almost perfectly through a random search procedure.

Examples of AVC Solutions

This section presents three examples of discovered solutions that illustrate the information flow through the AVC structure. The first example was selected since its accuracy in classification scores highest for all solutions during testing. The second example corresponds to the solution of the motorcycle class that exhibits the lowest accuracy in classification during the testing stage. Finally, a third example illustrates

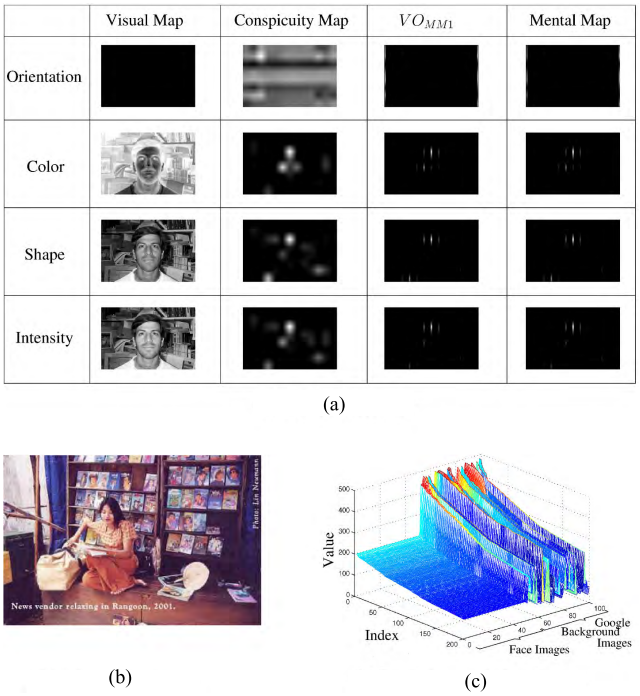

FIGURE 11. These figures show the functionality of the solution AVC_{F1}. Figure (a) depicts the image transformation along the *AVC* structure by applying the *VOs* of AVC_{F1}. As a result, an image classified as false positive can be seen in Figure (b) while Figure (c) illustrates the descriptors that result after applying the solution AVC_{F1} to the testing image set.

thought-provoking behavior since it uses eight mental maps while focusing on specific regions.

The AVC_{F1} solution, see Table 5, was discovered for the face class. Figure 11 (a) shows the behavior after applying an input image to the AVC. The color, shape and intensity dimensions highlight the forehead region of the face. Note that the orientation dimension is eliminated since the input image is mapped to zero; i.e., black color. The only image that was classified as false positive belongs to the Google background class; see Figure 11 (b). This image could be classified as a face because the image contains a person. However, in this image the response of the algorithm is located outside of the face. The reason is that the pattern of the features for the class "faces" is different since in this class the faces are centered over the image while covering a higher area. We observe that the values for the faces' descriptors are well defined in comparison to those of the Google background class. This behavior is described in Figure 11 (c).

The second result corresponds to the AVC_{M1} that scored the highest value for the problem of classifying the motorcycles vs. the Google background class; see Table 5. Figure 12(a) depicts the information flow process for the AVC_{M1} structure. In general, the visual operators that are applied to the image highlight the motorcycle contour and shape. Nevertheless, when the motorcycle was cluttered with the background, the image was regularly confused with the Google background class; an example of such a case is depicted in Figure 12(b). This behavior produces a lower accuracy

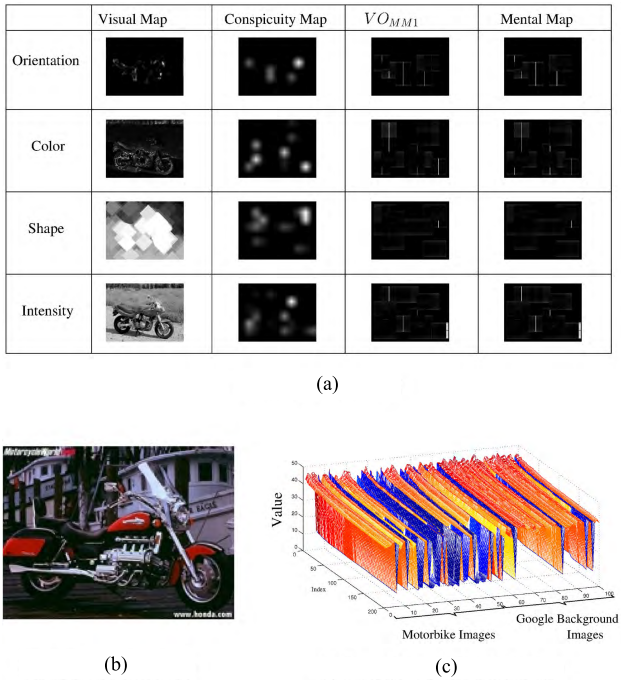

FIGURE 12. These figures provide details about the behavior of the AVC_{F2} solution. Figure (a) shows the image processing through the *AVC* structure; Figure (b) shows some examples that were classified correctly and the hilly regions found after applying the AVC_{F2} solution, and Figure (c) depicts the descriptors obtained for the testing image set.

in classification during the testing stage. Hence, we can say that for this example, of the motorcycle class, the range of the descriptors' values is not well defined; this effect can be seen in Figure 12(c). Finally, as the last example the AVC_{F2} solution was selected due to its ability for detecting a specific region, see Table 5. In this case, the diagram of the faces' category illustrates the information flow through the AVC structure as depicted in Figure 13(a). Note that the shape dimension can be eliminated and the visual operators along color and orientation highlight the lip region on the face. This behavior has been observed on several testing images; see Figure 13(b), whose final descriptors are shown in Figure 13(c).

B. COMPARISON WITH OTHER METHODOLOGIES

This section provides a comparison with several approaches using more challenging databases to illustrate the capacity and limitation of the random search.

1) COMPARISON BETWEEN AVC AND HMAX MODELS

The HMAX model was used in the second series of tests based on the experimental design proposed in [19], in order to compare our results with the state-of-the-art. The solutions from the first experiment were tested, in the object present/absent experiment, with a new random set of images considering 50 positive images for the object classes selected earlier, as well as 50 negative images from the Caltech-5 background database. The goal is to investigate the effect on the 700 final solutions per class using accuracy. In this

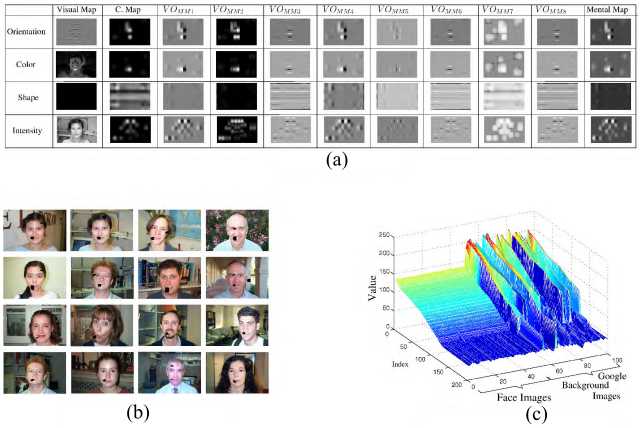

FIGURE 13. These figures provide details about the behavior of the AVC_{F2} solution. Figure (a) shows the image processing through the *AVC* structure; Figure (b) shows some examples that were classified correctly and the hilly regions found after applying the AVC_{F2} solution, and Figure (c) depicts the descriptors obtained for the testing image set.

TABLE 5. This table shows the best solutions that were discovered after a random process.

Name	VO	VO_{MMk}	Evaluation
AVC_{A1}	$VO_O = round(D_{xy}(I_k))$ $VO_C = I_r$ $VO_S = \lfloor I_r \rfloor$	$VO_{MM1} = D_{yy}(CM_d)$ $VO_{MM2} = G_{\sigma=2}(D_{xx}(CM_d))$ $VO_{MM3} = \sqrt{D_{xy}(CM_d)}$ $VO_{MM4} = D_{xx}(CM_d)$ $VO_{MM5} = D_y(CM_d)$ $VO_{MM6} = \lvert D_{xy}(CM) - D_x(CM_d) \rvert$ $VO_{MM7} = G_{\sigma=2}(D_{yy}(CM_d))$ $VO_{MM8} = D_{xx}(CM_d)$ $VO_{MM9} = log(D_{xx}(CM_d))$	$Tr = 100\%$ $Tst = 95\%$
AVC_{C1}	$VO_O = D_y(I_v)$ $VO_C = \frac{I_b}{I_r}$ $VO_S = \lfloor I_c \rfloor$	$VO_{MM1} = D_x(CM_d)$ $VO_{MM2} = log(CM_d)$ $VO_{MM3} = D_{xx}(CM_d)$ $VO_{MM4} = D_{xy}(CM_d)$ $VO_{MM5} = D_{xy}(CM_d)$ $VO_{MM6} = D_{yy}(CM_d)$ $VO_{MM7} = G_{\sigma=1}(D_y(CM_d))$ $VO_{MM8} = D_y(CM_d)$ $VO_{MM9} = CM_d$ $VO_{MM10} = D_{xy}(CM_d)$ $VO_{MM11} = log(D_y(CM_d))$ $VO_{MM12} = G_{\sigma=2}(D_{yy}(CM_d))$	$Tr = 100\%$ $Tst = 77\%$
AVC_{F1}	$VO_O = round(D_{xx}(I_m))$ $VO_C = \sqrt{I_k}$ $VO_S = 0.45 * (I_b)$	$VO_{MM1} = (D_{xx}(CM_d))^2$	$Tr = 100\%$ $Tst = 99\%$
AVC_{L1}	$VO_O = \sqrt{D_x(I_b)}$ $VO_C = I_b * I_g$ $VO_S = \lfloor I_h \rfloor$	$VO_{MM1} = 0.5 * (D_y(CM_d))$ $VO_{MM2} = Dyy(CM_d) - D_y(CM_d)$	$Tr = 100\%$ $Tst = 97\%$
AVC_{M1}	$VO_O = G_{\sigma=1}(D_x(I_y)^2)$ $VO_C = \frac{I_b^c}{I_g - I_r}$ $VO_S = (I_v - 0.21) \oplus SE_{dm}$	$VO_{MM1} = \frac{D_{xx}(CM_d)}{D_{xx}(CM_d)} - log(D_x(CM_d))$	$Tr = 100\%$ $Tst = 75\%$
AVC_{F2}	$VO_O = D_{yy}(I_h)$ $VO_C = I_g - I_r$ $VO_S = \lfloor I_y \rfloor$	$VO_{MM1} = D_x(CM_d)$ $VO_{MM2} = \lvert D_x(CM_d) \rvert$ $VO_{MM3} = G_{\sigma=1}(D_{yy}(CM_d))$ $VO_{MM4} = G_{\sigma=2}(D_x(CM_d))$ $VO_{MM5} = D_{xx}(CM_d)$ $VO_{MM6} = 0.5 \times (D_{yy}(CM_d))$ $VO_{MM7} = log(CM_d)$ $VO_{MM8} = G_{\sigma=1}(D_{yy}(CM_d))$	$Tr = 100\%$ $Tst = 92\%$

test, the background images are in grayscale; therefore, all color bands were initialized with the same value. The results summary is shown in Table 6. The experiment includes a comparison with LeNet and a basic convolutional neural network (from scratch CNN), whose results were computed with 100 runs for each class and size of the training set. The comparison between our model and the HMAX model is provided in Table 7. We report the error rate at the equilibrium point as the measure performance in these experiments. For the sake of showing that the differences between the performances of the proposed AVC and the HMAX-SVM models are statistically significant, we used two non-parametric statistical tests: the Wilcoxon rank sum [83] and a two-sample

TABLE 6. This table summarizes the classification results achieved on testing using the background Caltech-5 database as the negative class.

Class	No. Images on Training	Mean	Std	Max	Min	LeNet	From scratch CNN
Airplane vs background	10	0.65	0.10	0.95	0.50	0.73 ± 0.08	0.76 ± 0.05
	20	0.64	0.09	0.90	0.49	0.67 ± 0.11	0.72 ± 0.06
	30	0.63	0.10	0.86	0.48	0.75 ± 0.10	0.78 ± 0.09
	40	0.62	0.10	**0.97**	0.50	0.77 ± 0.10	0.78 ± 0.08
	50	0.62	0.10	0.96	0.49	0.76 ± 0.10	0.78 ± 0.08
	60	0.60	0.10	0.95	0.50	0.80 ± 0.10	0.77 ± 0.07
	70	0.65	0.11	0.96	0.50	0.86 ± 0.09	0.86 ± 0.07
Cars vs background	10	0.60	0.08	0.78	0.47	0.79 ± 0.07	0.72 ± 0.07
	20	0.62	0.12	0.97	0.45	0.85 ± 0.06	0.88 ± 0.04
	30	0.60	0.10	0.97	0.50	0.85 ± 0.05	0.92 ± 0.03
	40	0.60	0.09	0.92	0.45	0.92 ± 0.03	0.92 ± 0.04
	50	0.61	0.16	0.96	0.47	0.95 ± 0.03	0.93 ± 0.02
	60	0.58	0.10	0.92	0.46	0.94 ± 0.02	0.93 ± 0.03
	70	0.60	0.11	**0.98**	0.46	0.95 ± 0.02	0.96 ± 0.02
Faces vs background	10	0.66	0.15	0.97	0.43	0.74 ± 0.08	0.76 ± 0.07
	20	0.67	0.15	0.98	0.47	0.79 ± 0.09	0.79 ± 0.07
	30	0.66	0.14	0.99	0.47	0.79 ± 0.09	0.87 ± 0.08
	40	0.64	0.13	0.97	0.46	0.86 ± 0.09	0.92 ± 0.06
	50	0.66	0.14	**1.00**	0.48	0.88 ± 0.07	0.91 ± 0.07
	60	0.61	0.13	0.97	0.42	0.89 ± 0.08	0.88 ± 0.06
	70	0.63	0.13	0.99	0.47	0.90 ± 0.08	0.94 ± 0.05
Leaves vs background	10	0.70	0.14	**0.98**	0.50	0.87 ± 0.06	0.88 ± 0.04
	20	0.68	0.13	0.97	0.50	0.83 ± 0.07	0.84 ± 0.06
	30	0.65	0.12	0.91	0.47	0.83 ± 0.09	0.87 ± 0.08
	40	0.67	0.14	0.95	0.46	0.81 ± 0.08	0.93 ± 0.04
	50	0.59	0.11	0.97	0.45	0.92 ± 0.04	0.94 ± 0.04
	60	0.64	0.13	0.96	0.46	0.94 ± 0.04	0.93 ± 0.05
	70	0.62	0.11	0.92	0.47	0.96 ± 0.04	0.95 ± 0.03
Motorcycles vs background	10	0.63	0.12	0.95	0.43	0.78 ± 0.06	0.81 ± 0.07
	20	0.62	0.12	0.98	0.48	0.76 ± 0.07	0.86 ± 0.08
	30	0.58	0.09	0.97	0.46	0.75 ± 0.10	0.86 ± 0.06
	40	0.59	0.11	0.96	0.47	0.85 ± 0.07	0.84 ± 0.07
	50	0.59	0.12	0.96	0.45	0.90 ± 0.05	0.91 ± 0.04
	60	0.56	0.08	0.90	0.45	0.90 ± 0.05	0.89 ± 0.06
	70	0.61	0.13	**1.00**	0.41	0.91 ± 0.04	0.93 ± 0.03

TABLE 7. This table shows a comparison of the performance achieved by the HMAX model, considering the boost and SVM classifiers, against the AVC model. Note that in the case of the HMAX model a learning process was applied in order to identify the best patches. However, only a random sampling was used to discover the best solution with the AVC model.

Datasets	Performance of HMAX		Artificial	Statistical Significance	
	boost	SVM	V. C.	K-S test	Wilcoxon test
Airplanes	96.7	94.9	98.6	4.9×10^{-15}	1.1×10^{-7}
Cars	99.7	99.8	98.1	2.1×10^{-13}	3.9×10^{-8}
Faces	98.2	98.1	100	1.8×10^{-15}	1.3×10^{-10}
Leaves	97.0	95.9	96.2	7.9×10^{-17}	6.3×10^{-12}
Motorcycles	98.0	97.4	100	5.4×10^{-13}	9.6×10^{-6}

Kolmogorov-Smirnov test [84]. These last experiments were tested on the 30 best random solutions out of the 700 found for each class.

2) COMPARISON WITH THE GRAZ BENCHMARK

GRAZ is part of the PASCAL object recognition database collection, built by Opelt *et al.*, and it consists of two challenging datasets [85]. First, the GRAZ-01 database contains two classes and background set that are varied in locations, scales and viewpoints. Next, the GRAZ-02 dataset was built to increase the independence of the background context for categorization. Also, the complexity of object appearances in photographs was increased, and the car image class was added as a new category. For testing the AVC model with random runs, we followed the protocol provided in [85]. For the GRAZ-01 dataset, 100 positive and 100 negative images were randomly selected as training samples, and other

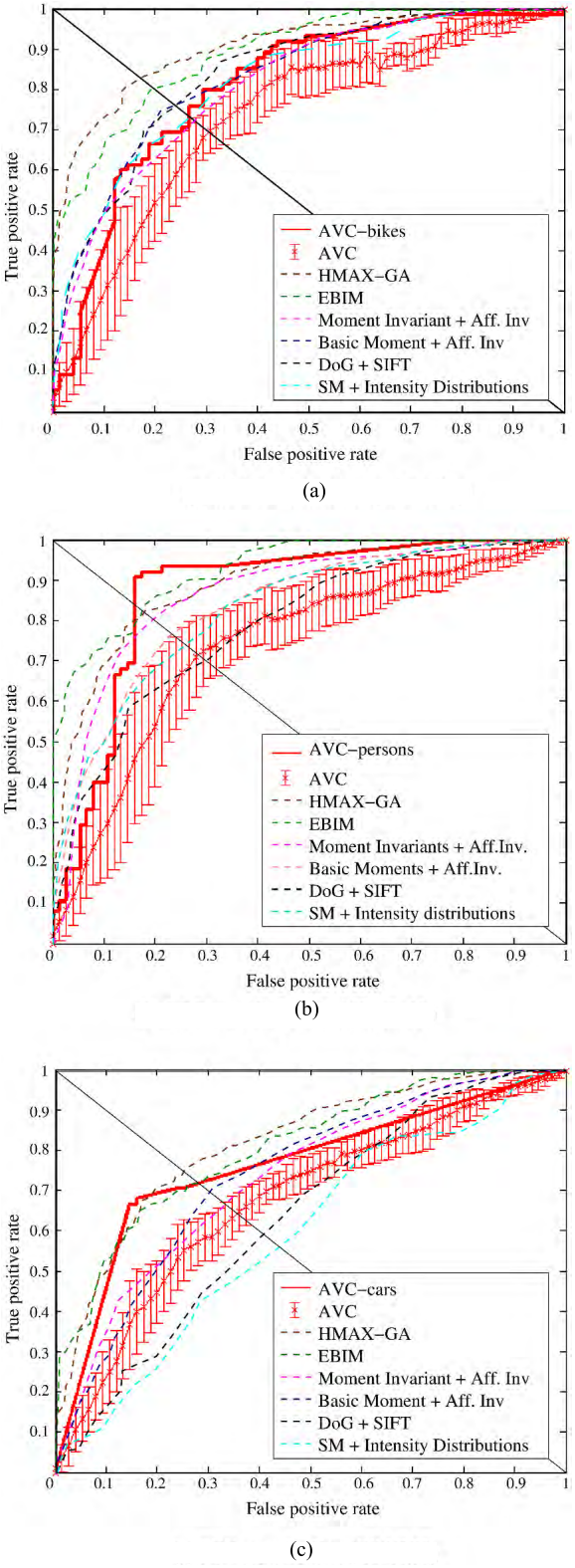

FIGURE 14. Comparison of several approaches on GRAZ-02 database.

50 positive images and 50 negative images were selected as testing samples. For GRAZ-02, 150 positive and 150 negative images were selected at random as training samples, then another set of 75 positive and 75 negative images were

TABLE 8. This table shows a comparison between several feature extraction methods, the AVC average performance, and the best AVC solutions on GRAZ-01.

Methods	Bikes		Persons	
	EER	AUC	EER	AUC
EBIM [88]	84.1	90.5	86.0	91.8
SM [85]	83.5	89.6	56.5	59.1
HMAX-GA [29]	80.2	88.5	84.0	90.8
SIFT [85]	78.0	86.5	76.5	80.8
AVC-bikesG01	**76**	**79.0**	–	–
AVC-personsG01	–	–	**72.0**	**70.0**
Moment Invariants [85]	73.5	76.5	63.0	68.7
AVC	**66.2** ±2.9	**68.8** ±3.2	**66.6** ±2.1	**69.2** ±2.9

TABLE 9. This table shows a comparison between several feature extraction methods and the AVC average performance computed with the EER on GRAZ-02.

Methods	Bikes	Persons	Cars
HMAX-GA [29]	82.6	82.3	75.6
EBIM [88]	80.8	83.2	72.2
Mutch et. al. [20]	80.5	81.7	70.1
Basic Moments [85]	76.5	77.2	70.2
SIFTs [85]	76.4	70.0	68.9
SM [85]	74.0	74.1	56.5
Moment Invariants [85]	72.5	81.1	67.0
AVC	**69.2** ±1.7	**71.5** ±2.9	**64.8** ±1.6

TABLE 10. This table shows the best solution over each class on GRAZ-02.

Class	Solution	eer	auc
Bikes	AVC-bikes	73.3	80.8
Persons	AVC-persons	84.0	87.1
Cars	AVC-cars1	71.7	75.9

randomly selected as testing samples. The experiments were run against the 20 best solutions for the GRAZ-01 and 100 best solutions for the GRAZ-02; considering that the discovered solutions along 3500 random runs were those whose SVM scored a threshold above 75% in classification accuracy during training. All experimental results are reported in Tables 8, 9 and 10; where the average ROC (receiver operating characteristics), the area under the curve (AUC), and Equal-Error rate (EER means the detection rate at equal-error-rate of the ROC curve) are used as performance measurements. Note that we provide results of the best AVC models for the two classes at Table 8 considering GRAZ-01, while the best results for GRAZ-02 are reported in Table 10. The comparison between the AVC model with random runs was made against well-known methods, such as Basic Moments, HMAX-GA, EBIM, SIFT, SM, and Moment Invariants, see Table 9. Figure 14 depicts the results for comparison with Tables 9 and 10 on the GRAZ-02 database. Note that the AVC outperforms other approaches in the Persons class and it achieves a good ranking for the Cars class while achieving lower performance in comparison to other approaches for the Bikes class.

3) COMPARISON WITH THE VOC CHALLENGE

The PASCAL Visual Object Classes (VOC) Challenge 2007 and its associated database has become accepted as a benchmark for object detection [87]. The challenge provides the vision and machine learning communities with a standard dataset of images, annotations, and evaluation procedures that are used to fairly compare different image classification systems [86]. The VOC2007 database contains natural images with significant variability in terms of object size, orientation, pose, illumination, positions, and occlusions. Today, the VOC dataset is considered as one of the most challenging databases for object classification [88]–[90]. In this paper, the classification challenge is used to evaluate the performance of the AVC using the random search. The selected challenge is the classification whose goal is to determine the presence/absence of an object from a particular class within an image. This database contains images from 20 difference object classes: person, train, car, aeroplane, horse, boat, motorbike, bicycle, bus, cat, bird, diningtable, chair, tvmonitor, sofa, dog, sheep, cow, pottedplant, and bottle. The VOC2007 database is divided into two subsets: the training/validation set, known as *trainval*, composed of 5011 images; and the testing set, which consists of 4952 images. The *trainval* set is further divided into two subsets, which are used in our two-stage process for the discovery of the best AVC solutions.

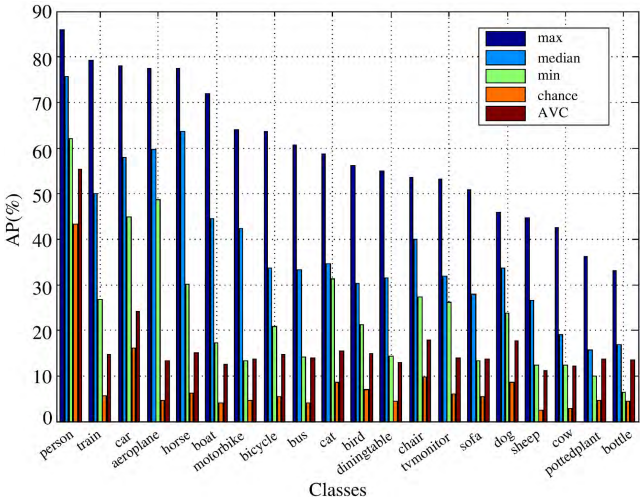

FIGURE 15. This figure provides a performance comparison between the best AVC solutions (AVC) against the maximum AP (max), median AP (median), minimum AP (min) and AP (chance) obtained by a random ranking of the images reported in [87] for the VOC2007 database.

The experiment consists of 500 random tries used to select the best AVC solution for each class. The performance of the solutions is computed with the average precision (AP) measure proposed in the VOC2007 challenge. In this work, we compare the best AVC solutions performances with the best, median, and minimum results reported in [87] over the twenty classes; as well as, the *chance* performance, whose AP value was obtained with a classifier outputting a random confidence value without examining the input image; see Figure 15. Note how the performances of all AVC solutions are better than the *chance* method, even though all solutions were found through a random procedure. While the test

shows the limit of our search process, we note that the AVC random solutions for the bottle, motorbike, pottedplant, and sofa classes outperform the low boundary solutions reported in [87].

V. CONCLUSIONS AND FUTURE WORK

This paper presented a novel computational model of the visual cortex following the hierarchical structure of previous visual attention and object recognition proposals. The overall approach considers that the processes of extraction and description can be enforced by function composition through a set of mathematical operations that are used within the stages as mentioned earlier. According to the results, all functions embedded within the hierarchical structure of the AVC can be quickly discovered through random search while achieving excellent results on the Caltech and GRAZ databases. The results provide evidence about the regularity in patterns related to the optimal size of the training set. The results show that the proposal matches the performance of algorithms in the state-of-the-art according to the results obtained in Caltech and GRAZ testbeds. As a conclusion, we can say that the AVC methodology offers a new perspective to study the development of artificial brains since the structural complexity can be improved because the approach is susceptible of being framed as an optimization problem. In this way, we can synthesize new structures according to the task at hand. In particular, for future research, we would like to test the approach with more complex datasets such as the VOC challenge, ImageNet, and Visual Genome [87], [91], [92]. The methodology is computationally costly, and we propose to change to parallel computing implementations of the AVC model through the application of GPGPU technology [13]. Finally, we would like to continue to explore the application of this new paradigm to problems of humanoid robotics [12].

VI. ACKNOWLEDGMENT

The paper presented in this article is an extension of the material published in [23] which was nominated for the Best Paper Award during EvoStar 2017.

REFERENCES

[1] M. Riesenhuber and T. Poggio, "Models of object recognition," *Nature Neurosci.*, vol. 3, pp. 1199–1204, Nov. 2000. doi: 10.1038/81479.

[2] N. Pinto, D. D. Cox, and J. J. DiCarlo, "Why is real-world visual object recognition hard?" *PLoS Comput. Biol.*, vol. 4, no. 1, pp. 151–156, Jan. 2008. doi: 10.1371/journal.pcbi.0040027.

[3] J. J. DiCarlo, D. Zoccolan, and N. C. Rust, "How does the brain solve visual object recognition?" *Neuron*, vol. 73, no. 3, pp. 415–434, Feb. 2002. doi: 10.1016/j.neuron.2012.01.010.

[4] G. Olague, *Evolutionary Computer Vision—The First Footprints*. Berlin, Germany: Springer, 2016. [Online]. Available: https://link.springer.com/book/10.1007/978-3-662-43693-6

[5] G. Olague, E. Clemente, L. Dozal, and D. E. Hernández. "Evolving an artificial visual cortex for object recognition with brain programming," *EVOLVE—A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation III*, vol. 500. Heidelberg, Germany: Springer, 2014, pp. 97–119. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-01460-9_5

[6] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cogn. Psychol.*, vol. 12, no. 1, pp. 97–136, Jan. 1980. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0010028580900055

[7] M. M. Mishkin, L. G. Ungerleider, and K. A. Macko, "Object vision and spatial vision: Two cortical pathways," *Trends Neurosci.*, vol. 6, pp. 414–417, Jan. 1983. [Online]. Available: https://www.sciencedirect.com/science/article/pii/016622368390190X

[8] D. E. Hernández, E. Clemente, G. Olague, and J.L. Briseño, "Evolutionary multi-objective visual cortex for object classification in natural images," *J. Comput. Sci.*. vol. 17, pp. 216–233, Nov. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877750315300338

[9] E. Clemente, F. Chavez, F. F. de Vega, and G. Olague, "Self-adjusting focus of attention in combination with a genetic fuzzy system for improving a laser environment control device system," *Appl. Soft Comput.*, vol. 32, pp. 250–265, Jul. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494615001647

[10] L. Dozal, G. Olague, E. Clemente, and D. E. Hernández. "Brain programming for the evolution of an artificial dorsal stream," *Cogn. Comput.*, vol. 6, pp. 528–557, Sep. 2014. [Online]. Available: https://link.springer.com/article/10.1007/s12559-014-9251-6

[11] G. Olague and D. E. Hernández, E. Clemente, and M. Chan-Ley, "Evolving head tracking routines with brain programming," *IEEE Access*, vol. 6, pp. 26254–26270, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8352651/

[12] G. Olague and D. E. Hernández, P. Llamas, E. Clemente, and J.L. Briseño, "Brain Programming as a New Strategy to Create Visual Routines for Object Tracking," *Multimedia Tools Appl.*, vol. 78, no. 5, pp. 5881–5918, Mar. 2018. [Online]. Available: https://link.springer.com/article/10.1007/s11042-018-6634-9

[13] D.E. Hernández, G. Olague, B. Hernández, and E. Clemente, "CUDA-based parallelization of a bio-inspired model for fast object classification," *Neural Comput. Appl.*. vol. 30, no. 10, pp. 3007–3018, Nov. 2018. [Online]. Available: https://link.springer.com/article/10.1007/s00521-017-2873-3

[14] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *J. Physiol.*, vol. 148, no. 3, pp. 574–591, 1953. doi: 10.1113/jphysiol.1959.sp006308.

[15] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, Jan. 1962. doi: 10.1113/jphysiol.1962.sp006837.

[16] D. H. Hubel, "Exploration of the primary visual cortex, 1955–78," *Nature*, vol. 299, no. 5883, pp. 515–524, Oct. 1982. doi: 10.1038/299515a0.

[17] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Annu. Rev. Neurosci.*, vol. 18, pp. 193–222, Mar. 1995. doi: 10.1146/annurev.ne.18.030195.001205.

[18] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, Apr. 1980. doi: 10.1007/BF00344251.

[19] T. Serre, M. Kouh, C. Cadieu, U. Knoblich, G. Kreiman, and T. Poggio, "A Theory of object recognition: Computations and circuits in the feedforward path of the ventral stream in primate visual cortex," Comput. Sci. Artif. Intell. Lab., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. TR-2005-082, Jul. 2003. [Online]. Available: https://dspace.mit.edu/handle/1721.1/36407

[20] J. Mutch and D. G. Lowe, "Object class recognition and localization using sparse features with limited receptive fields," *Int. J. Comput. Vis.*, vol. 80, no. 1, pp. 45–57, Oct. 2008. doi: 10.1007/s11263-007-0118-0.

[21] B. W. Mel, "SEEMORE: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition," *Neural Comput.*, vol. 9, no. 4, pp. 777–804, May 1997. doi: 10.1162/neco.1997.9.4.777.

[22] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature Rev. Neurosci.*, vol. 2, no. 3, pp. 194–203, Mar. 2001. doi: 10.1038/35058500.

[23] G. Olague, E. Clemente, D. E. Hernández, and A. Barrera. "Brain programming and the random search in object categorization," in *Applications of Evolutionary Computation* (Lecture Notes in Computer Science), vol. 10199, G. Squillero and K. Sim, Eds. Cham, Switzerland: Springer, 2017. doi: 10.1007/978-3-319-55849-3_34.

[24] I. Biederman, "Recognition-by-components: A theory of human image understanding," *Psychol. Rev.*, vol. 94, no. 2, pp. 115–147, Apr. 1987. doi: 10.1037/0033-295X.94.2.115.

[25] D. I. Perrett and M. W. Oram, "Visual recognition based on temporal cortex cells: Viewer-centred processing of pattern configuration," *Zeitschrift für Naturforschung C*, vol. 53, nos. 7–8, pp. 518–541, Jul./Aug. 1998. doi: 10.1515/znc-1998-7-807.

[26] S. Ullman and S. Soloviev, "Computation of pattern invariance in brain-like structures," *Neural Netw.*, vol. 12, nos. 7–8, pp. 1021–1036, Oct./Nov. 1999. doi: 10.1016/S0893-6080(99)00048-9.
[27] S. Ullman, M. Vidal-Naquet, and E. Sali, "Visual features of intermediate complexity and their use in classification," *Nature Neurosci.*, vol. 5, no. 7, pp. 682–687, Jun. 2002. doi: 10.1038/nn870.
[28] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neurosci.*, vol. 2, no. 11, pp. 1019–1025, Nov. 1999. doi: 10.1038/14819.
[29] M. Ghodrati, S. M. Khaligh-Razavi, R. Ebrahimpour, K. Rajaei, and M. Pooyan, "How can selection of biologically inspired features improve the performance of a robust object recognition model?," *PLoS ONE*, vol. 7, no. 2, 2012, Art. no. e32357. doi: 10.1371/journal.pone.0032357.
[30] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," *Hum. Neurobiol.*, vol. 4, no. 4, pp. 219–227, 1985. [Online]. Available: https://cseweb.ucsd.edu/classes/fa09/cse258a/papers/koch-ullman-1985.pdf
[31] J. M. Wolfe, K. R. Cave, and S. L. Franzel, "Guided search: An alternative to the feature integration model for visual search," *Journal Exp. Psychol., Hum. Perception Perform.*, vol. 15, no. 3, pp. 419–433, Aug. 1989. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/2527952
[32] R. Milanese, "Detecting salient regions in an image: From biological evidence to computer implementation," Ph.D. dissertation, Dept. Comput. Sci., Univ. Geneva, Geneva, Switzerland, 1993.
[33] K. Fukushima, "Neural network model for selective attention in visual pattern recognition and associative recall," *Appl. Opt.*, vol. 26, no. 23, pp. 4985–4992, Dec. 1987. doi: 10.1364/AO.26.004985.
[34] B. A. Olshausen, C. H. Anderson, and D. C. Van Essen, "A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information," *J. Neurosci.*, vol. 13, no. 11, pp. 4700–4719, Nov. 1993. doi: 10.1523/JNEUROSCI.13-11-04700.1993.
[35] D. Walther, L. Itti, M. Riesenhuber, T. Poggio, and C. Koch, "Attentional selection for object recognition—A gentle way," *Biologically Motivated Comput. Vis.*, (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2002, pp. 472–479. doi: 10.1007/3-540-36181-2_47.
[36] D. B. Walther and C. Koch, "Attention in hierarchical models of object recognition," *Prog. Brain Res.*, vol. 165, pp. 57–78, Jan. 2007. doi: 10.1016/S0079-6123(06)65005-X.
[37] D. Heinke and G. W. Humphreys, "Attention, spatial representation, and visual neglect: Simulating emergent attention and spatial memory in the selective attention for identification model (SAIM)," *Psychol. Rev.*, vol. 110, no. 1, pp. 29–87, Jan. 2003. [Online]. Available: https://psycnet.apa.org/record/2002-08416-004
[38] A. D. Milner and M. A. Goodale, *The Visual Brain in Action*, 2nd ed. Oxford, U.K.: Oxford Univ. Press, 2006. doi: 10.1093/acprof:oso/9780198524724.001.0001.
[39] G. E. Schneider, "Contrasting visuomotor functions of tectum and cortex in the golden hamster," *Psychologische Forschung*, vol. 31, no. 1, pp. 52–62, Mar. 1967. doi: 10.1007/BF00422386.
[40] G. E. Schneider, "Two visual systems," *Science*, vol. 163, no. 3870, pp. 895–902, Feb. 1969. doi: 10.1126/science.163.3870.895.
[41] L. G. Ungerleider and J. V. Haxby, "'What' and 'where' in the human brain," *Current Opinion Neurobiol.*, vol. 4, no. 2, pp. 157–165, Apr. 1994. doi: 10.1016/0959-4388(94)90066-3.
[42] S. H. Creem and D. R. Proffitt DR, "Defining the cortical visual systems: 'What' 'where' and 'how,'"*Acta Psychologica*, vol. 107, nos. 1–3, pp. 43–68, Apr. 2001. doi: 10.1016/S0001-6918(01)00021-X.
[43] R. Farivar, "Dorsal-ventral integration in object recognition," *Brain Res. Rev.*, vol. 61, no. 2, pp. 144–153, Oct. 2009. doi: 10.1016/j.brainresrev.2009.05.006.
[44] K. Grill-Spector and R. Malach, "The human visual cortex," *Annu. Rev. Neurosci.*, vol. 27, pp. 649–677, Jul. 2004. doi: 10.1146/annurev.neuro.27.070203.144220.
[45] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Netw.*, vol. 19, no. 9, pp. 1395–1407, Nov. 2006. doi: 10.1016/j.neunet.2006.10.001.
[46] M. Corbetta, F. Miezin, S. Dobmeyer, G. Shulman, and S. E. Petersen, "Attentional modulation of neural processing of shape, color, and velocity in humans," *Science*. vol. 248, no. 4962, pp. 1556–1559, Jun. 1990. doi: 10.1126/science.2360050.
[47] M. Ito and H. Komatsu, "Representation of angles embedded within contour stimuli in area V2 of macaque monkeys," *J. Neurosci.*, vol. 24, no. 13, pp. 3313–3324, Mar. 2004. doi: 10.1523/JNEUROSCI.4364-03.2004.
[48] R. Heydt, E. Peterhans, and G. Baumgartner, "Illusory contours and cortical neuron responses," *Science*, vol. 224, no. 4654, pp. 1260–1262, Jun. 1984. doi: 10.1126/science.6539501.
[49] A. Plebe, "A model of angle selectivity development in visual area V2," *Neurocomputing*, vol. 70, nos. 10–12, pp. 2060–2063, Jun. 2007. doi: 10.1016/j.neucom.2006.10.105.
[50] J. Hegdé, and D. C. Van Essen, "Selectivity for complex shapes in primate visual area V2," *J. Neurosci.*, vol. 20, no. 5, pp. 1–6, Mar. 2000. doi: 10.1523/JNEUROSCI.20-05-j0001.2000.
[51] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Boston, MA, USA: Addison Wesley, 1992. [Online]. Available: https://dl.acm.org/citation.cfm?id=573607
[52] G. Olague and L. Trujillo, "Evolutionary-computer-assisted design of image operators that detect interest points using genetic programming," *Image Vis. Comput.*, vol. 29, no. 7, pp. 484–498, Jun. 2011. doi: 10.1016/j.imavis.2011.03.004.
[53] C. B. Perez and G. Olague, "Genetic programming as strategy for learning image descriptor operators," *Intell. Data Anal.*, vol. 17, no. 4, pp. 561–583, Jul. 2013. [Online]. Available: https://content.iospress.com/articles/intelligent-data-analysis/ida00594
[54] R. A. Young, R. M. Lesperance, and W. W. Meyer, "The Gaussian derivative model for spatial-temporal vision: I. Cortical model," *Spatial Vis.*, vol. 14, nos. 3–4, pp. 261–319, 2001. [Online]. Available: https://brill.com/abstract/journals/sv/14/3-4/article-p261_3.xml
[55] S. V. David, B. Y. Hayden, and J. L. Gallant, "Spectral receptive field properties explain shape selectivity in area V4," *J. Neurophysiol.*. vol. 96, no. 6, pp. 3492–3505, Dec. 2006. doi: 10.1152/jn.00575.2006.
[56] A. Pasupathy and C. E. Connor, "Responses to contour features in macaque area V4," *J. Neurophysiol.*, vol. 82, no. 5, pp. 2490–2502, Nov. 1999. doi: 10.1152/jn.1999.82.5.2490.
[57] C. G. Gross, C. E. Rocha-Miranda, and D. B. Bender, "Visual properties of neurons in inferotemporal cortex of the macaque," *J. Neurophysiol.*. vol. 35, no. 1, pp. 96–111, Jan. 1972. doi: 10.1152/jn.1972.35.1.96.
[58] E. Kobatake and K. Tanaka, "Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex," *J. Neurophysiol.*. vol. 71, no. 3, pp. 856–867, Mar. 1994. doi: 10.1152/jn.1994.71.3.856.
[59] C. P. Hung, G. Kreiman, T. Poggio, and J. J. DiCarlo, "Fast readout of object identity from macaque inferior temporal cortex," *Science*, vol. 310, no. 5749, pp. 863–866, Nov. 2005. doi: 10.1126/science.1117593.
[60] K. Tanaka, "Inferotemporal cortex and object vision," *Annu. Rev. Neurosci.*, vol. 19, pp. 109–139, Mar. 1996. [Online]. Available: https://www.cs.cmu.edu/afs/cs/academic/class/15883-f13/readings/tanaka-1996.pdf
[61] R. Desimone, T. D. Albright, C. G. Gross, and C. Bruce, "Stimulus-selective properties of inferior temporal neurons in the macaque," *J. Neurosci.*, vol. 4, no. 8, pp. 2051–2062, Aug. 1984. doi: 10.1523/JNEUROSCI.04-08-02051.1984.
[62] J. Ponce *et al.*, "Dataset issues in object recognition," *Toward Category-Level Object Recognition* (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2006, pp. 29–48. doi: 10.1007/11957959_2.
[63] R. Gopalakrishnan, Y. Chua, and L. R. Iyer, "Classifying neuromorphic data using a deep learning framework for image classification," in *Proc. 15th Int. Conf. Control, Automat., Robot. Vis. (ICARCV)*, Nov. 2018, pp. 1520–1524. doi: 10.1109/ICARCV.2018.8581256.
[64] I. Rocco, A. Relja, and S. Josef, "End-to-end weakly-supervised semantic alignment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6917–6925. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2018/papers/Zhu_End-to-End_Flow_Correlation_CVPR_2018_paper.pdf
[65] J. Ryu, M. H. Yang, and J. Lim, "DFT-based transformation invariant pooling layer for visual classification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 84–99. doi: 10.1007/978-3-030-01264-9_6.
[66] W. Luo, J. Li, J. Yang, W. Xu, and J. Zhang, "Convolutional sparse autoencoders for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 3289–3294, Jul. 2018. doi: 10.1109/TNNLS.2017.2712793.
[67] M. Weber, M. M. Welling, and P. Perona, "Unsupervised learning of models for recognition," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2000, pp. 18–32. doi: 10.1007/3-540-45054-8_2.
[68] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, pp. 18–20, 2003. doi: 10.1109/CVPR.2003.1211479.
[69] Z. Wang and J. Feng, "Multi-class learning from class proportions," *Neurocomputing*, vol. 119, pp. 273–280, Nov. 2003. doi: 10.1016/j.neucom.2013.03.031.

[70] Z. Ji, J. Wang, Y. Su, Z. Song, and S. Xing, "Balance between object and background: Object-enhanced features for scene image classification," *Neurocomputing*, vol. 120, pp. 15–23, Nov. 2003. doi: 10.1016/j.neucom.2012.02.054.

[71] S. Chandra, S. Kumar, and C. V. Jawahar, "Learning hierarchical bag of words using naive Bayes clustering," in *Proc. Asian Conf. Comput. Vis.*, 2012, pp. 382–395. doi: 10.1007/978-3-642-37331-2_29.

[72] B. Chen, G. Polatkan, G. Sapiro, D. Blei, D. Dunson, and L. Carin, "Deep learning with hierarchical convolutional factor analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1887–1901, Aug. 2013. doi: 10.1109/TPAMI.2013.19.

[73] B. Xu, R. Hu, and P. Guo, "Combining affinity propagation with supervised dictionary learning for image classification," *Neural Comput. Appl.*, vol. 22, nos. 7–8, pp. 1301–1308, Jun. 2013. doi: 10.1007/s00521-012-0957-7.

[74] N. S. Kamarudin, M. Makhtar, S. A. Fadzli, M. Mohamad, F. S. Mohamad, and M. F. D. Kadir, "Comparison of image classification techniques using Caltech 101 dataset," *J. Theor. Appl. Inf. Technol.*, vol. 71, no. 1, pp. 79–86, Jan. 2015. [Online]. Available: http://www.jatit.org/volumes/Vol71No1/9Vol71No1.pdf

[75] Y. Xie, F. Porikli, and X. He, "Object-aware dictionary learning with deep features," in *Proc. Asian Conf. Comput. Vis.*, 2016 pp. 237–253. doi: 10.1007/978-3-319-54184-6_15.

[76] S. H. Khan, M. Hayat, M. Bennamoun, R. Togneri, and F. A. Sohel, "A discriminative representation of convolutional features for indoor scene recognition," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3372–3383, Jul. 2016. doi: 10.1109/TIP.2016.2567076.

[77] D. Yu and X. J. Wu, "VLAD is not necessary for CNN," in *Proc. Eur. Conf. Comput. Vis.-Workshops*, 2016, pp. 185–194. doi: 10.1007/978-3-319-49409-8_41.

[78] Y. He, L. Wang, Z. Wu, and H. Zhang, "Object recognition in images via a factor graph model," *Proc. SPIE*, vol. 10615, Apr. 2018, Art. no. 1061517. doi: 10.1117/12.2303409.

[79] Y. Hong and W. Zhu, "Learning visual codebooks for image classification using spectral clustering," *Soft Comput.*, vol. 22, no. 18, 6077–6086, 2018. doi: 10.1007/s00500-017-2937-4.

[80] H. Cholakkal, J. Johnson, and D. Rajan, "Backtracking spatial pyramid pooling-based image classifier for weakly supervised top–down salient object detection," *IEEE Trans. Image Process.*, vol. 27, no. 12, 6064–6078, Dec. 2018. doi: 10.1109/TIP.2018.2864891.

[81] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop*, Jun./Jul. 2007, vol. 106, no. 1, pp. 59–70. doi: 10.1109/CVPR.2004.383.

[82] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," California Inst. Technol., Pasadena, CA, USA, Tech. Rep. CNS-TR-2007-001, 2007. [Online]. Available: https://authors.library.caltech.edu/7694/

[83] F. Wilcoxon, "Individual comparison by ranking methods," *Biometrics Bolletin*, vol. 1, no. 6, pp. 80–83, Dec. 1945. doi: 10.2307/3001968.

[84] F. J. Massey, Jr., "The Kolmogorov-Smirnov test for goodness of fit," *J. Amer. Stat. Assoc.*, vol. 46, no. 253, pp. 68–78, Mar. 1951. doi: 10.1080/01621459.1951.10500769.

[85] A. Opelt, A. Pinz, M. Fussenegger, and P. Auer, "Generic object recognition with boosting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 416–431, Mar. 2006. doi: 10.1109/TPAMI.2006.54.

[86] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. (2007). The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results [Internet]. EU-funded PASCAL Network of Excellence on Pattern Analysis, Statistical Modelling and Computational Learning. Accessed: Mar. 26, 2008 [Online]. Available: http://host.robots.ox.ac.uk/pascal/VOC/voc2007/

[87] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010. doi: 10.1007/s11263-009-0275-4.

[88] Y. Huang, K. Huang, D. Tao, T. Tan, and X. Li, "Enhanced biologically inspired model for object recognition," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 41, no. 6, pp. 1668–1680, Dec. 2011. doi: 10.1109/TSMCB.2011.2158418.

[89] C. Wang and K.-Q. Huang, "VFM: Visual feedback model for robust object recognition," *J. Comput. Sci. Technol.*, vol. 30, no. 2, pp. 325–339, Mar. 2015. [Online]. Available: http://jcst.ict.ac.cn/EN/10.1007/s11390-015-1526-1

[90] X. Bai, Z. Zhang, H. Y. Wang, and W. Shen, "Directional edge boxes: Exploiting inner normal direction cues for effective object proposal generation," *J. Comput. Sci. Technol.*, vol. 32, no. 4, pp. 701–713, Jun. 2017. doi: 10.1007/s11390-017-1752-9.

[91] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, pp. 211–252, Dec. 2015. doi: 10.1007/s11263-015-0816-y.

[92] R. Krishna *et al.*, "Visual Genome: Connecting language and vision using crowdsourced dense image annotations," *Int. J. Comput. Vis.*, vol. 123, no. 1, pp. 32–73, May 2017. doi: 10.1007/s11263-016-0981-7.

GUSTAVO OLAGUE (M'99–SM'17) was born in Chihuahua, Chih., México, in 1969. He received the B.S. and M.S. degrees in industrial and electronics engineering from the Instituto Tecnológico de Chihuahua (ITCH), in 1992 and 1995, respectively, and the Ph.D. degree in computer vision, graphics, and robotics from the Institut Polytechnique de Grenoble (INPG) and the Institut National de Recherche en Informatique et Automatique (INRIA) in France. He is currently a Professor with the Department of Computer Science, Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), México, and also the Director of the EvoVisión Research Team. He is also an Adjunct Professor of engineering with the Universidad Autonóma de Chihuahua (UACH). He has authored over 100 conference proceedings papers and journal articles, co-edited two special issues in *Pattern Recognition Letters* and *Evolutionary Computation* (MIT Press). He has authored the book *Evolutionary Computer Vision* (Springer) in the Natural Computing Series. His main research interests are evolutionary computing and computer vision. He is a member of the Editorial Team of the IEEE Access, *Neural Computing and Applications* (Springer), and served as the Co-Chair of the Real-World Applications track at the main international evolutionary computing conference, GECCO (ACM SIGEVO Genetic and Evolutionary Computation Conference), in 2012 and 2013. He has received numerous distinctions, among them the Talbert Abrams Award–first honorable mention 2003–presented by the American Society for Photogrammetry and Remote Sensing (ASPRS) for authorship and recording of current and historical engineering and scientific developments in photogrammetry; Best Paper Awards at major conferences such as GECCO, EvoIASP (European Workshop on Evolutionary Computation in Image Analysis, Signal Processing, and Pattern Recognition), and EvoHOT (European Workshop on Evolutionary Hardware Optimization); and twice the Bronze Medal at the Humies (GECCO award for Human-Competitive results produced by genetic and evolutionary computation).

EDDIE CLEMENTE was born in Mexico, in 1982. He received the bachelor's degree in mechatronics engineering from the Interdisciplinary Professional Unit on Engineering and Advanced Technologies, one of the schools of the National Polytechnic Institute (UPIITA-IPN), México, and the M.Sc. degree in computer science and the Ph.D. degree in computer science from the Centro de Investigación Científica y de Educación Superior de Ensenada, B.C., (CICESE), México, in 2006 and 2015, respectively. He is also a member of the Robotics and Control Research Team, Instituto Tecnológico de Ensenada. His research interests include evolutionary computer vision, robotics, and evolutionary computation.

DANIEL E. HERNÁNDEZ was born in Tijuana, B.C., México, in 1985. He received the bachelor's degree in computer engineering from the Universidad Autónoma de Baja California (UABC), México, the M.Sc. degree in computer science from from the Centro de Investigación Científica y de Educación Superior de Ensenada, B.C., (CICESE), México, in 2011, and the Ph.D. degree in computer science from CICESE. He is currently with the Instituto Tecnológico de Tijuana and also the Director of the Master Program on Intelligent Manufacturing. His research interests include computer vision, robotics, evolutionary computation, and bio-inspired algorithms.

AARON BARRERA was born in Mexicali, México, in 1987. He received the bachelor's degree in electrical engineering from the Instituto Tecnológico de Mexicali, México, in 2010, and the M.Sc. degree in computer science from the Centro de Investigación Científica y Educación Superior de Ensenada (CICESE), México, in 2017. He has been with the electronics manufacturing industry for more than eight years working in test design engineering departments. He is currently a member with the EvoVisión Research Team, CICESE. His research interests include artificial intelligence, neurotechnology, and the psychedelic culture and philosophy.

MARIANA CHAN-LEY was born in Mérida, Yucatan, México, in 1988. She received the bachelor's degree in mechatronics engineering from the Universidad Autónoma de Yucatan, México, and the M.Sc. degree in computer science from the Centro de Investigación Científica y de Educación Superior de Ensenada, B.C., (CICESE), México, in 2017, where she is currently pursuing the Ph.D. degree in computer science and also a member of the EvoVisión Research Team. Her research interests include evolutionary computer vision, projective geometry, robotics, and evolutionary computation.

SAMBIT BAKSHI received the Ph.D. degree in computer science and engineering, in 2015. He is currently with the Centre for Computer Vision and Pattern Recognition, National Institute of Technology at Rourkela, Rourkela, India, where he is also an Assistant Professor with the Department of Computer Science and Engineering. He has more than 50 publications in reputed journals, magazines, and conferences. His area of interest includes surveillance and biometric authentication. He is the Technical Committee Member of the IEEE Computer Society Technical Committee on Pattern Analysis and Machine Intelligence. He received the prestigious Innovative Student Projects Award 2011 from the Indian National Academy of Engineering (INAE) for his master's thesis. He currently serves as an Associate Editor for the *International Journal of Biometrics*, since 2013, the IEEE Access, since 2016, *Innovations in Systems and Software Engineering* (A NASA Journal), since 2016, *Expert Systems*, since 2017, and *Plos One*, since 2017. From 2016 to 2017, he has served/serving as the Guest Editor for reputed journals, such as *Multimedia Tools and Applications*, the IEEE Access, *Innovations in Systems and Software Engineering* (A NASA Journal), *Computers and Electrical Engineering*, and *IET Biometrics*.