

# A conceptual frame with two neural mechanisms to model selective visual attention processes

José Mira<sup>a</sup>, Ana E. Delgado<sup>a</sup>, María T. López<sup>b,c</sup>, Antonio Fernández-Caballero<sup>b,c,\*</sup>, Miguel A. Fernández<sup>b,c</sup>

<sup>a</sup>*Departamento de Inteligencia Artificial, E.T.S.I. Informática, UNED, 28040 Madrid, Spain*

<sup>b</sup>*Departamento de Sistemas Informáticos, Escuela Politécnica Superior de Albacete, Universidad de Castilla-La Mancha, 02071 Albacete, Spain*

<sup>c</sup>*Instituto de Investigación en Informática de Albacete (I3A), Universidad de Castilla-La Mancha, 02071 Albacete, Spain*

Available online 12 November 2007

## Abstract

An important problem in artificial intelligence (AI) is to find calculation procedures to save the semantic gap between the analytic formulations of the neuronal models and the concepts of the natural language used to describe the cognitive processes. In this work we explore a way of saving this gap for the case of the attentional processes, consisting in (1) proposing in first place a conceptual model of the attention double bottom-up/top-down organization, (2) proposing afterwards a neurophysiological model of the cortical and sub-cortical involved structures, (3) establishing the correspondences between the entities of (1) and (2), (4) operationalizing the model by using biologically inspired calculation mechanisms (algorithmic lateral inhibition and accumulative computation) formulated at symbolic level, and, (5) assessing the validity of the proposal by accommodating the works of the research team on diverse aspects of attention associated to visual surveillance tasks. The results obtained support in a reasonable way the validity of the proposal and enable its application in surveillance tasks different from the ones considered in this work. In particular, this is the case when linking the geometric descriptions of a scene with the corresponding activity level.

© 2007 Elsevier B.V. All rights reserved.

**Keywords:** Visual selective attention; Algorithmic lateral inhibition; Accumulative computation; Neurophysiological models; Artificial Intelligence; Computer vision

## 1. Introduction

One of the most important problems in artificial intelligence (AI) and in computational neuroscience (CN) is to find effective calculation procedures that enable connecting the analytic models of the behaviour of individual neurons, typical of the Neurodynamics [6,7], with the formulations in natural language of the concepts and inferences associated to the high-level cognitive processes [49,55,52,59]. We have just celebrated the 50th anniversary of AI and it is evident that there is still no general and satisfactory solution to this problem. The

reasons for this lack of links between the models belonging to Physics and AI reside in the intrinsic complexity of the cognitive processes and in the lack of adequate methodological approaches.

In this work the selective visual attention (SVA) problem is approached, firstly proposing a conceptual model at the knowledge level on the double bottom-up/top-down organization, of the attentional processes in humans, together with a neurophysiological model of the corresponding cortical and sub-cortical structures. Next two complimentary neural mechanisms, algorithmic lateral inhibition (ALI) [9,10,43,54] and accumulative computation (AC) [16,19–21] are selected for the computational rewriting of the different elements of the conceptual model. Finally, the validity of our proposal, by using image sequences in visual surveillance tasks, is analysed.

The ideas underlying the work are: (1) In order to theorize and to solve problems that go beyond the

\*Corresponding author at: Departamento de Sistemas Informáticos, Escuela Politécnica Superior de Albacete, Universidad de Castilla-La Mancha, 02071 Albacete, Spain. Tel.: +34 913987155; fax: +34 913988895.

E-mail address: [caballer@dsi.uclm.es](mailto:caballer@dsi.uclm.es) (A. Fernández-Caballero).

retinotopic projection step it is necessary to propose neurophysiologically plausible synthetic architectures (knowledge models), but without the precision by which the more peripheral structures are known. (2) It is not sufficient to model and interpret the neuronal function at physical (physiological) level of registrations of slow potentials or spikes trains. It is necessary to re-formulate the neuronal mechanisms at the symbolic level in terms of inferential rules and frames, closer to the natural language used by an external observer when describing the attentional processes, or when performing psychophysical experiments [32,33].

The rest of the paper is constructed as follows. In Section 2, the SVA task and some relevant approaches to its computational modelling are presented. Then, our proposal is introduced at a conceptual level (Section 3) and illustrated in Section 4. Afterwards, the two basic mechanisms selected for making this architecture operational, namely ALI (Section 5.1) and AC (Section 5.2) are described. Finally, Section 6 draws some conclusions.

## 2. Selective visual attention task

The name SVA embraces a set of image processing mechanisms for focusing vision and/or a set of effectors on regions of the image where there are relevant local space–time events. The greater or less relevance of a local event has to do with: (1) its differential character in relation to its environment, (2) its analogy or difference with the set of features defining the objects in a pre-defined set to be recognized, and (3) a combination of both criteria. These SVA mechanisms have a twofold purpose: (1) to filter the relevant information from irrelevant information and (2) to modulate the information, pre-selected in accordance with a purpose, using an active search process.

The SVA task has the characteristics of a dynamic selection process based on a combination of data-guided (bottom-up) and knowledge-based (top-down) criteria, as

illustrated in Fig. 1 (see also [7,31,32,58,67,75]). The input role is the set of pixels in the visual scene at every moment and the output role consists of coordinates of the selected items. These are the coordinates of image regions that stand out at that moment (“where” to look). The pixels stand out because they have maximum values of activity in some property or combination of properties or in the analogy measurement of the object detected (“what” to look at) with one or another of the pre-selected set. The static roles consist of the active and passive selection criteria.

This description of the SVA task is static, since it implicitly implies that the visual scene remains constant while the selection criteria are applied. However, in many situations of interest, the selection of coordinates and/or relevant items is done on time (driving, moving targets, surveillance, etc.) which forces us to compare the situation of the visual scene in successive instants of time first seeing “when” it changes in some relevant aspect and, afterwards, seeing “where” it changes. Accordingly, the SVA mechanisms need to have a short-term memory (STM) or the more inferential functionality of a working memory (WM) to store every fact or new event in the time evolution of the scene.

There are two types of criteria: (1) based on the scene (guided by data) and (2) based on the task [71] (guided by knowledge). The first model guided by data was proposed by Koch and Ullman [40] and it is based on Anne Treisman’s conjecture on feature integration [66]. In [38], this bottom-up model is associated with biological data on the anatomy and physiology of the visual pathway in its closest stage to the external environment, before arriving to the cortex. It assumes that a maximum selection operation on the responses of the neural circuitry that has to extract features is sufficient [60]. In SAIM [34], the assumption is that there are three nets (“contents”, “selection”, and “knowledge”) that interact as if they were “agents”.

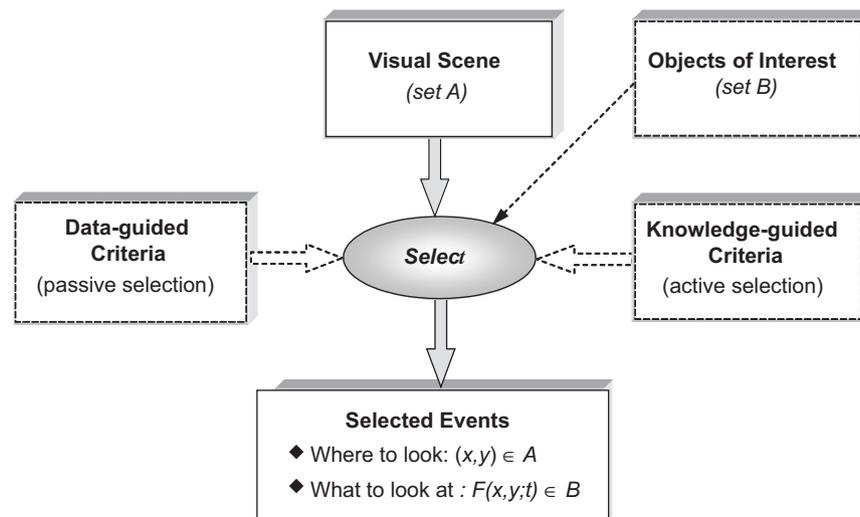


Fig. 1. Selective visual attention task.

In relation to top-down models, based on criteria of selection associated with the internal activity of the system (with its intentions, purposes and objectives), it is necessary to highlight the “Guided Search” model proposed by Wolfe [74]. This model uses the idea of “relevant features map” (saliency map) as a starting point to close the loop and to look for the targets of interest using knowledge about the objects. The attention is directed towards the objects that better correspond with the features that stand out in every temporary interval. Another model [71] can learn to focus its attention on important features depending on the task to be fulfilled by gating the flow of information from the lower to the higher levels of the visual system. Backer and Metsching’s model [2] also abounds with the necessity to combine bottom-up criteria with other intentional top-down criteria. In all cases, in an explicit or implicit form, the idea of “control loop” that connects the perception with the action underlies.

The idea of combining top-down with bottom-up processing is not new. A model for aspects of visual attention based on the concept of selective tuning [68] is perhaps the first known approach. In general, there is now a consensus that visual search is guided by a mixture of bottom-up and top-down influence. An excellent review on computational models—including papers describing the double organization—can be found in [23]. Hamker [31] describes a computational model of attention which explains the guidance of spatial attention by explicitly incorporating top-down feature feedback within a distributed network. Deco and Rolls [7] describe a model of invariant visual object recognition in the brain that incorporates feedback biasing effects of top-down attention mechanisms on a hierarchically organized set of visual cortical areas with convergent forward connectivity, reciprocal feedback connections, and local intra-area competition. A good attempt to mix top-down and bottom-up guided search has been introduced in a recent model by Navalpakkam and Itti [58] in which bottom-up saliency is combined with information about the ongoing task or goal. Task constraints are introduced by user-supplied keywords interacting with knowledge in a long-term memory (LTM) network. In another work [67], a novel model of attentive visual motion processing is presented, addressing both decomposition of the signal into constituent features as well as the binding of those features into wholes. The visual processing architecture is pyramidal in structure, with units within this network receiving both feed-forward and feedback connections. In [32], some simulations with a revised version of the Selective Attention for Identification Model (SAIM) demonstrate the usage of both factors. A very different integrative approach is adopted in another recent image-based model by Zelinsky et al. [75]. In this model, the relative contributions of top-down and bottom-up processes are systematically explored in the context of a realistic search task. Rather than biasing individual bottom-up features to reflect task demands, these authors

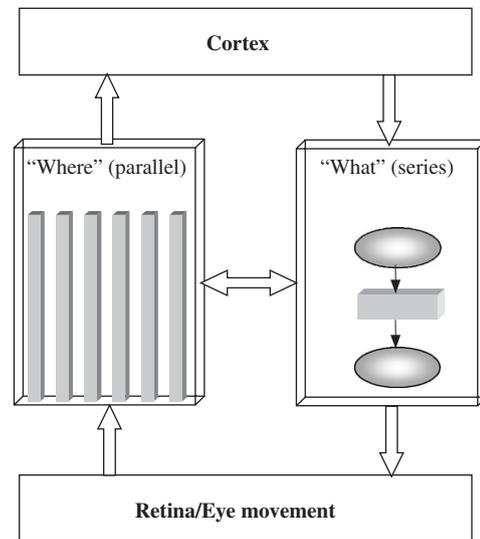


Fig. 2. Bottom-up/top-down double organization of SVA task, as proposed in [7,31,32,58,67,75], among others.

compute separate bottom-up and top-down saliency maps, and then combine the two in various mixtures to derive a guidance signal.

Thus, there is experimental evidence and consensus on theoretical developments about the existence of a double bottom-up (from the photoreceptors up to the sensory cortex) and top-down (from the prefrontal cortex to motor areas and effectors) organization [6] in SVA mechanisms (see Fig. 2). It is crucial to consider the function of interaction between both organizations, which is present in almost all cortical levels. Most theories of visual search explicitly acknowledge [64,74] or tacitly assume [13] the contribution of both bottom-up and top-down processes in guiding search behaviour.

This approach involves a set of hierarchically biased cooperative/competitive attentional processes. It is cooperative because it relies on the double bottom-up/top-down process. And it is hierarchically biased as there is a hierarchy of levels and the mechanism is globally controlled by the high level cognitive processes. Biased competition suggests that neuronal responses are determined by competitive interactions that are subject to a number of biases, such as ‘bottom-up’ stimulus information and ‘top-down’ cognitive requirements. Important in the theory is the idea that a WM template of the target object can bias competition between objects and features such that the target object is given a competitive advantage and other objects are suppressed (cf. [11]). The dominance is provided by the purposes or objectives in a concrete task, in a concrete domain and for a concrete observer.

### 3. Conceptual and neurophysiological models of SVA: a conjecture of correspondences

In CN it is always difficult to establish a correspondence between the neurophysiological data and the anatomic

organization of the information flow, on the one hand, and the external formulation in terms of natural language of the emergent functions at conscious cognitive level. That is why we talk about conjecture when referring to the content of this section, where a possible correspondence between the functions that we consider necessary for the synthesis of the SVA processes and the underlying anatomic structures when the attentional functions are performed by a human is proposed.

SVA is a global synthetic activity based on the cooperative action of the sensory, motor and association-decision brain structures [8,47,48,56,57]. The different brain structures cooperate as contributing factors to the functions proper of the bottom-up and top-down organizations; and this cooperation is modified with development and learning.

From the point of view of an external observer, the distinguishing features of the SVA bottom-up organization, as shown in Fig. 3, are as follows:

1. It is pre-attentive and automatic [65,66], although it can be modulated by voluntary activity and reflex mechanisms dependent on innovation, conflict, surprise and uncertainty of the stimuli.
2. Computationally, it possesses the architecture of a parallel process (connectionist) on the whole visual scene [65,66], also concurrent with the rest of the sensory modalities.
3. It is generally accepted that this parallel process consists of extracting a set of local features [36] (intensity, intensity contrast, colour, orientation of the shapes, speed, direction, and so on) from which we construct a representation of the scene [63,68].
4. There are transverse lateral inhibition (LI) mechanisms [12,28], which enable the different features to be integrated to represent local situations characterized by the conjunction of two or more properties.

5. The feature detectors must be able to adapt (learn) so that the computational counterpart of the receptive fields of ganglion cells, LGB cells and columns in cortex can adjust [25] in shape and size.

In summary, the SVA bottom-up organization enables the calculation in parallel of local features of differential type, which detect the coordinates of the points in which there are strong discontinuities in the value of each property. It also enables the subsequent combination of these individual maps in a single map of relevant coordinates (feature integration).

The main justification for the bottom-up organization schema of the SVA in Fig. 3 is of computational nature. It includes the sequence of steps that we consider necessary and sufficient to dynamically obtain the coordinates of the attention focus. Therefore it is a synthetic architecture corresponding to a computable inferential schema, which has demonstrated its validity in segmentation and tracking tasks [16,18,19,44] where attentional processes participate. On the other hand, when operationalizing the different functions of feature extraction, evaluation (maxima selection), integration (configuration evaluation), transitory storage (FIFO-like memory) and presentation of the coordinates of the dynamic focus, the necessary logical-mathematical calculations are sufficiently ergonomic to enable their processing in real time [42].

The anatomic structures underlying the bottom-up organization of the SVA (Fig. 3) correspond to the three stages of the sensory pathway shown in Fig. 4 [8,47,48,51,56,57]: (1) sensation, (2) perception, and (3) conceptualization by multisensory association (integration of the different sensory modalities). The first stage is repeated for each sensory modality and includes receptor, corresponding thalamic nucleus and primary cortical area. Here the stimulus is detected but not identified. For the visual pathway it would be: (a) receptor (retina, optic nerve, and chiasm), (b) thalamic nucleus (external geniculate body), and (c) primary cortical area. The second stage (perception) constitutes the first gnostic level. Here there is already semantics. It is where the stimulus is coded and identified. It is the process that the sensory information undergoes from the thalamic nucleus of the previous level to the secondary cortical areas, passing through the corresponding associative thalamic nucleus. The third stage (multisensory association and stimulus conceptualization) includes the confluence and interaction of secondary areas corresponding to the different sensory modalities. Here the supramodal (symbolic) schema takes place.

From the computational viewpoint, the top-down organization of SVA corresponds to a task-oriented guided search with the following characteristics (Fig. 5):

1. It is a sequential process [68,74].
2. The system initially knows the features of the objects and events on which it wants to focus its attention [30,50]. It knows the “objects and events of interest” that it tries to identify [39].

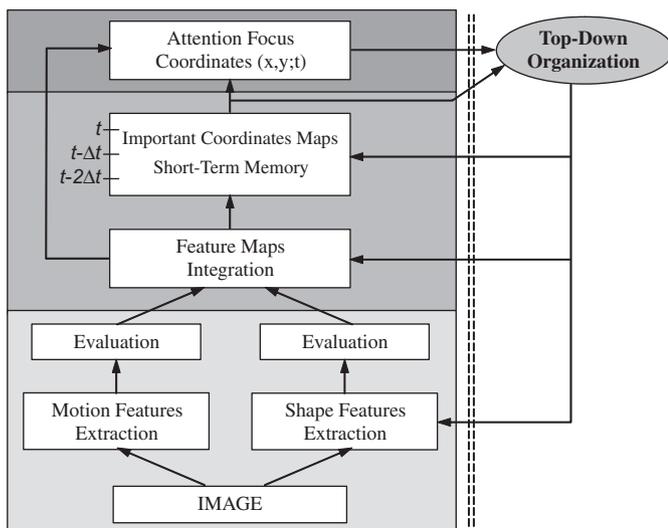


Fig. 3. Conceptual model of SVA bottom-up organization.

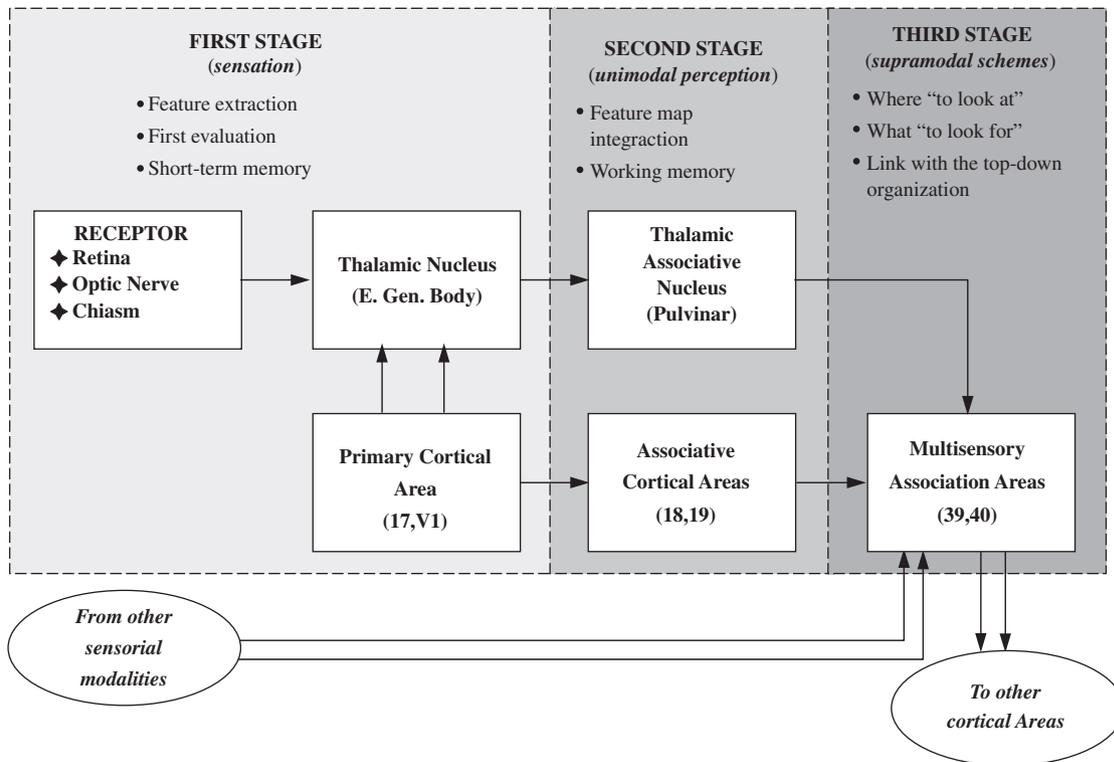


Fig. 4. Neurophysiological model with the anatomical substratum of the bottom-up organization of sensory functions.

- The inferential schema of this active selection task includes sub-tasks of (a) retrieval (activation) of the sensory and motor patterns associated to the “knowledge base” of objects and events of potential interest, (b) pattern matching, (c) motor planning and control, and (d) learning, which reinforces or disables the pre-selected attention focus [35,62,70].
- This top-down organization is in charge of controlling the interaction between both organizations and refining the selection of the attention focus starting from the specific features of an object-target [29], that is to say, of a class properties combination. According to a recent study by Chen and Zelinsky [5], when top-down and bottom-up guidance signals are placed in competition, then top-down guidance clearly prevails. Previous studies have argued for a similarly dominant role of top-down processes during search [23].

The top-down organization has to do with task-oriented purposes and intentions and with the control in a closed loop of the activity of the sensors and effectors directed to reaching these purposes. Here we start with objects and events, which are entities of a superior level of semantics than the vectors of properties of the top-down organization (blobs, shape, colour, velocity, and so on). Thus, the top-down attentional processes involve: (a) an object recognition process that heightens the combination of best fit properties for the proposal of the search and (b) reinforcement learning to refine the initial bottom-up selection by tuning the parameters of the initial filtering and by

eliminating those items that are less similar to the stored prototypes [45].

Although the justification of the schema proposed in Fig. 5 is purely of computational nature, the neurophysiological model shown in Fig. 6 as anatomical support to the top-down attentional processes in humans is proposed in this paper. The decision on the interest events emerge from the activity of the prefrontal areas that programme and control the execution of all the voluntary motor acts (eye, head and rest of the body movements, to focus on the coordinates proposed by the sensory system). These prefrontal areas are connected to the contents of sensory and motor memory, and, consequently, have access to the sensory patterns of the objects and events of interest and to the motor patterns (the “kinetic melody”) adequate to cross the visual pathway and to focus on the environment of the coordinates where the object is present. The scanning of the scene is performed by means of saccadic, pursuit and compensatory movements that give place to successive overlapped observation windows. The motor patterns are in charge of turning the eyes, and/or moving the head, and/or moving the rest of the body, or performing displacements looking for other perspectives. The final result of this scanning phase enables to focus on regions where a more refined search is initiated.

From the very beginning of the search process the feedback loops between the motor outputs and the bottom-up sensory organization are active in order to compare the stored pattern with the current ones (see Fig. 6).

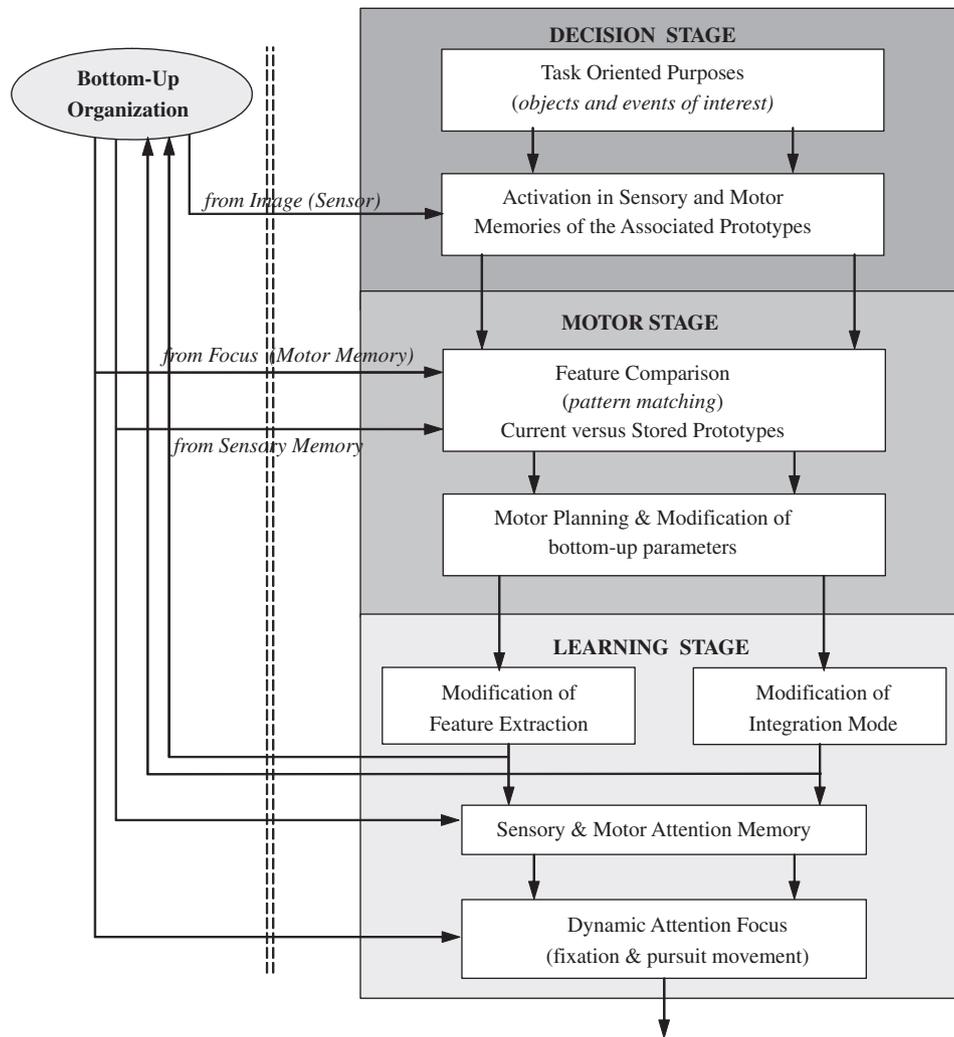


Fig. 5. Task-driven top-down organization model.

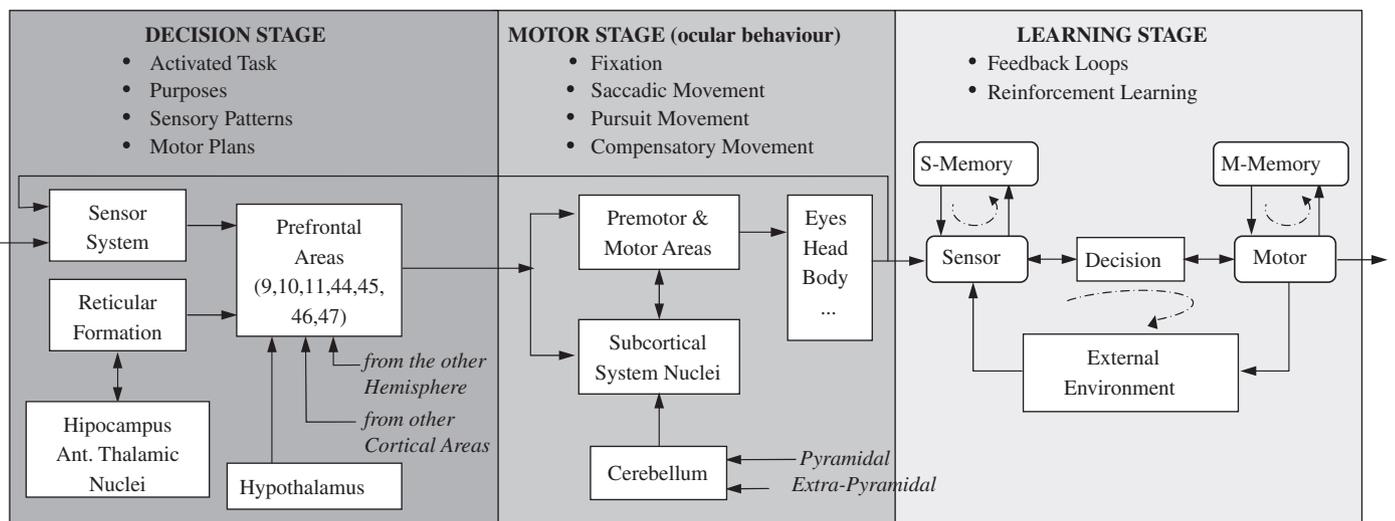


Fig. 6. Neurophysiological model of the top-down organization. In each one of the three phases (decision, motor, and learning) the correspondence between the functions of the conceptual model of Fig. 5 and the involved cortical and subcortical structures is included.

Finally, there is a learning process that enables to refine the matching operation, by adding new components to the prototypes base, or by modifying the distance measure, or

by modifying the parameters of the bottom-up organization, which have led to the selection process that has activated the top-down organization.

#### 4. Application in surveillance

Similarly to [30], where the author proposes a model for generic computer vision, which, in the long run, could be used for object recognition and tracking, our approach is task-oriented towards visual surveillance. Hamker states that there seems to be no generic data-driven solution to computer vision, but rather the model itself (by a cognitive control structure) has to provide appropriate top-down knowledge for each task. In our surveillance works, not all the functionalities of the biological systems can be used. Indeed, we do not use the movement of a camera situated on a robot, or the saccadic movements, the convergent, and divergent and adaptation processes of the eyes, etc. We limit to anything that can be extracted computationally from one or various image sequences captured by fixed cameras or cameras performing an easy two-dimensional translational movement. This limits the richness of the attention concept, as shown in Fig. 7, where an example is depicted using as input a sequence containing different

moving objects that may be interesting in surveillance. The final objective is the capacity of paying attention on pedestrians and/or vehicles. Thus, the proposal for the surveillance task is limited to:

1. The starting point is an image sequence (video sequence) in the bottom-up part of our double organization proposal. As you may notice in Fig. 7a, the input to this first step is a couple of images captured at two consecutive time instants  $t-\Delta t$  and  $t$ .
2. In the first stage (Fig. 7a) the extraction and evaluation of motion and shape features takes place. The result of *Motion Features Extraction and Evaluation* is shown as an image where white pixels represent image pixels where motion has been detected between two consecutive frames. This process consists in motion detection by image differencing.
3. Now, in *Shape Features Extraction and Evaluation* the values used for the evaluation of the blobs present in the image are shown. In our approach a blob is defined

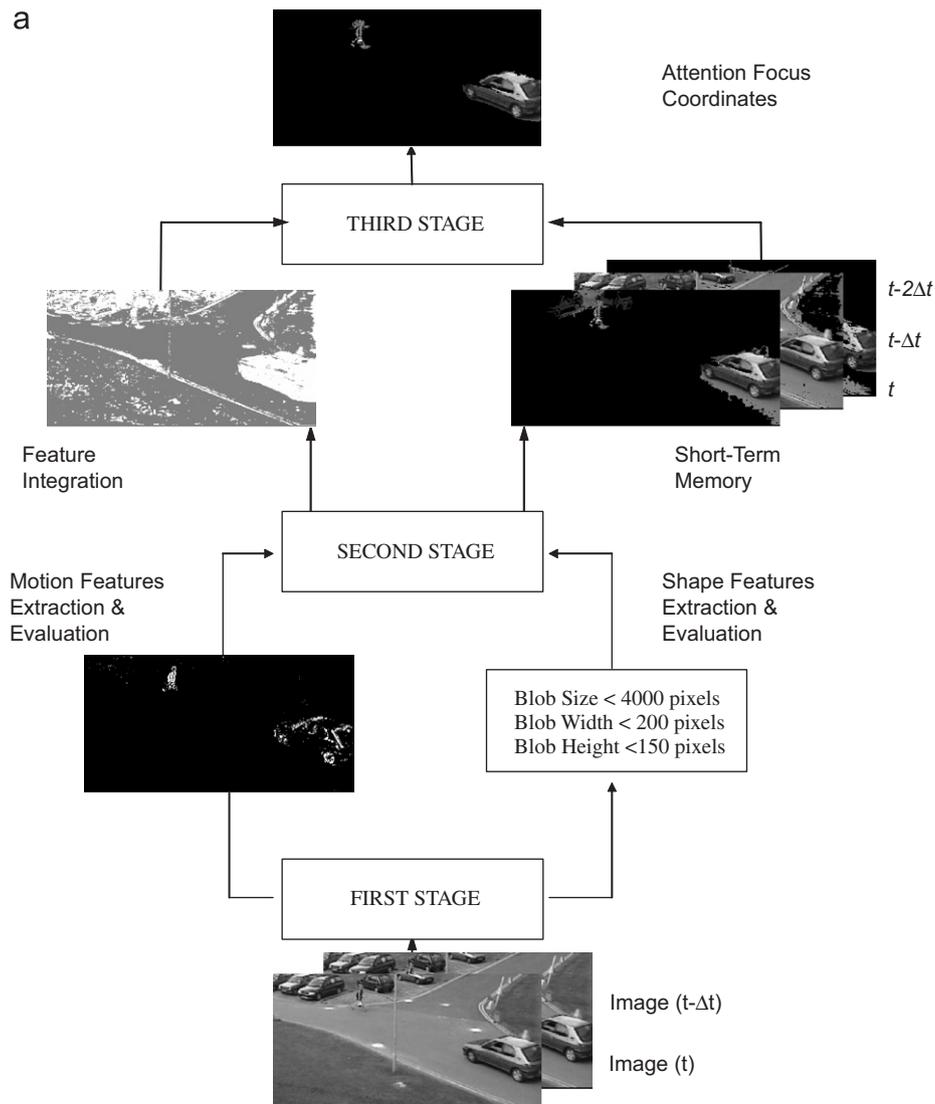


Fig. 7. Results of the bottom-up and top-down organizations in a surveillance scenario. (a) Bottom-up. (b) Top-down.

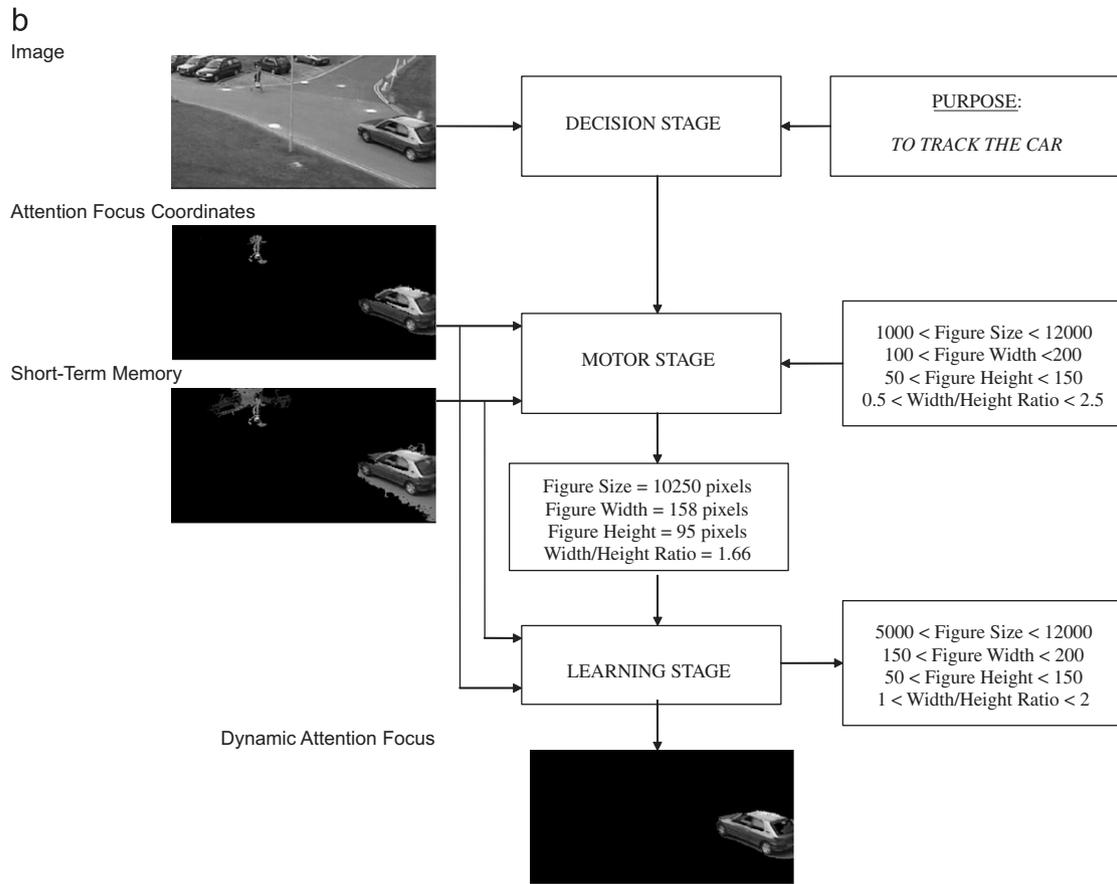


Fig. 7. (Continued)

as a set of connected zones with similar colours (generally, we work with grey levels). For the extraction of the blobs the ALI mechanism is used.

4. In the second stage (Fig. 7a) the result of *Feature Integration*, depicted as an image, is obtained. The colour white in the *Feature Integration* image references image blobs that include some pixel of image *Motion Features Extraction and Evaluation* and fulfil the shape conditions indicated in *Shape Features Extraction and Evaluation*. The colour grey indicates the zones that do not fulfil the shape features. Hence, the zones where there is motion associated to candidate objects of interest are filtered out.
5. Also, in this second stage the *STM* stores in diverse time instants the image zones that fulfil the initially required conditions of shape and motion. In our example the pedestrians as well as the vehicles fulfil the conditions.
6. In the bottom-up organization (we are still in Fig. 7a) in the third stage we keep the selected regions with the blobs of interest in the so-called *Attention Focus Coordinates*. The *Attention Focus Coordinates* have precisely been obtained by performing an AC process on the *STM*. We may observe that the desired objects appear in the *Attention Focus Coordinates*; these will be refined in the top-down organization (see Fig. 7b).

7. The first objective of the process shown in Fig. 7b is to refine the attention focus by means of ALI's competition mechanism of maxima selection, similar to winner-take-all (WTA), to obtain a unique object according to the semantics associated to the task proposal. In our case, the proposal is to focus attending to the semantic memory "to track the car". The decision stage receives the proposal related to the surveillance task and the current input image.
8. In the motor stage (also in Fig. 7b) a comparison between the *Attention Focus Coordinates* and the *STM*—both coming from the bottom-up organization—and the stored prototypes is performed. In our current proposal the prototypes are stored as ranges of values that the objects (figures) of interest have to comply with. Precisely, we are working with shape parameters related to size, width, height (given in pixels), and width/height ratio.
9. The output of the motor stage is the values obtained in the associated *Dynamic Attention Focus* figure (the vehicle, according to our proposal).
10. Lastly, the learning stage, as well as offering the *Dynamic Attention Focus* according to the values generated during the motor stage, it also activates the reinforcement learning mechanisms for the construction of a shapes base annotated with the parameters

that define them and with the potential semantics associated for its incremental usage in the following iterations of the bottom-up component of our proposal.

## 5. Operationalization of the inferences

In the previous sections the inferential schemes corresponding to the bottom-up (Fig. 3) and top-down (Fig. 5) organizations of the SVA together with the corresponding neurophysiological model (Figs. 4 and 6, respectively) have been described. In order to pass from these general descriptions to more operational ones, the inferences have to be substituted by analytic operators, by logical-relational rules, or by instructions of a programming language. For instance, Tsotsos [68] uses WTA updating rules and analytic operators, Deco and Rolls [6] uses a spiking dynamics approach, modelling each neuron by an integrate-and-fire model (first-order differential equation), and Walther [72] uses linear threshold units for the bottom-up attention and “if-then” rules for the top-down attentional modulation.

In this work we explore the conjecture that both organizations, bottom-up and top-down, may be made operational by essentially using two biologically inspired methods called algorithmic lateral interaction (ALI), a generalization of LI anatomical circuits [10] and AC. The ALI method may be compared to centre-surround approaches used by other authors in visual attention by considering that WTA algorithms are an extreme case of LI. Analogously, the AC method can be considered a generalization of STM, WM, and semantic memory.

The reason underlying this conjecture is that the first of these mechanisms the (ALI) enables the discrimination and detection of novelties and spatial-temporal changes, and the second (AC) permits to detect and to accumulate persistence in the structure of the optic flow. In formal terms the first mechanism permits the differential calculation and

the second the integral calculation. And in terms of organizational principles, the first mechanism permits the *cooperation* and the *competition* among the different neural assemblies and the second permits to filter, to store transitorily and to process the information associated to the neuronal activity in a similar way as the working memories (WM) are supposed to do.

To examine the results of the use of ALI and AC in SVA, a running example consisting of a flock of birds flying in the sky, as shown in Fig. 8, will be used throughout the remainder of the text.

### 5.1. Algorithmic lateral inhibition (ALI)

There are some neuronal structures that are repeated at all integration levels in the nervous system and which emerge again in the global behaviour of human beings. This is the case of the LI circuits. LI may be found in levels such as neurogenesis, dendro-dendritic contacts, neuronal circuits in the retina, lateral geniculate body (LGB), and in cerebral cortex, in the interaction between groups of neurons (ocular and orientation dominance columns) [54]. At the physical level, in terms of interconnected neurons, there are two basic connectivity schemes: (1) non-recurrent (feed-forward) LI and (2) recurrent LI, with feedback from the central and peripheral parts of the axonal output spaces (see Fig. 9). In non-recurrent LI, the modification of a unit response depends on the inputs of the neighbouring units, and in recurrent LI it depends on the outputs of the neighbouring units.

When extending the formulation to the continuous case (Fig. 9b) the terms of interaction become nuclei of a convolution integral. Then, the output,  $\Phi(x, y)$ , is the result of accumulating the direct excitation of each unit,  $I(x, y)$ , with the inhibition from modulating the excitation received by the connected neighbouring units,  $I(\alpha, \beta)$ , via the weight factors  $K(x, y; \alpha, \beta)$ . The difference nucleus,  $K(x, y; \alpha, \beta)$ , is now responsible for the specific form of the calculation which, in all cases, acts as a detector of contrasts (Fig. 10).



Fig. 8. Flock of birds running example.

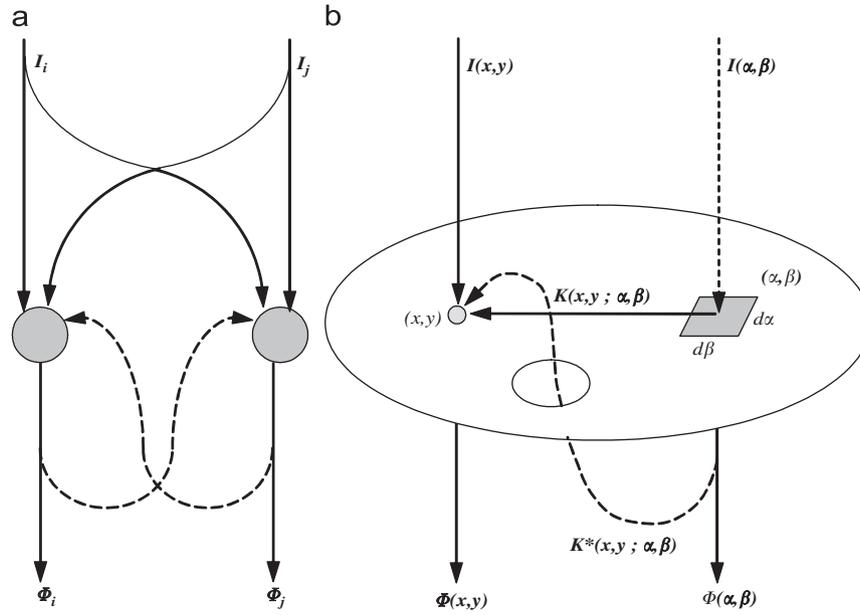


Fig. 9. Non-recurrent (continuous line) and recurrent (discontinuous line) LI connectivity schemes at the physical level: (a) discrete case and (b) continuous case.

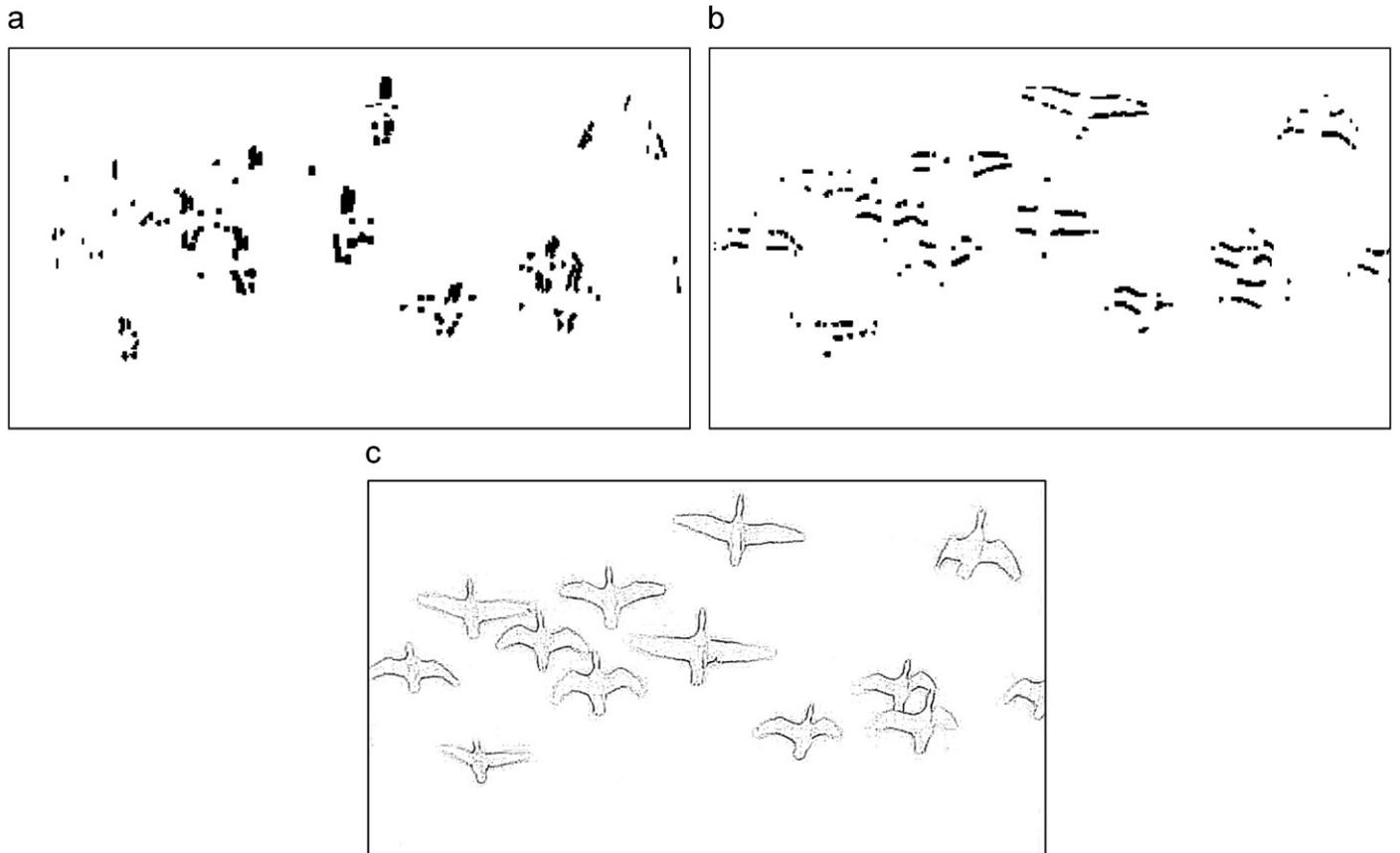


Fig. 10. Application of non-recurrent (feed forward) LI to the running example: (a) horizontal contrasts, (b) vertical contrasts and (c) contrasts on both axes.

In the recurrent LI it is the direct output,  $\Phi(x, y)$ , which accumulates the inhibition from the responses of the neighbouring units,  $\Phi(\alpha, \beta)$ , weighted by an interaction

coefficient  $K^*(x, y; \alpha, \beta)$ , basically different from that of the direct path ( $K$ ). Here again, in recurrent LI, the shape and size of the receptive field (the interaction nucleus  $K^*$ )

specifies the network connectivity (cooperation–competition area) and the calculation details (syntony, orientations, shapes, speeds, etc.). The most usual shape in  $K$  and  $K^*$  is obtained from subtracting two Gaussians. Thus, the calculation structures inherent in the entire LI network are obtained: (1) a central area (ON or OFF), (2) a peripheral area (OFF or ON), and (3) an excluded region (outside the receptive field).

This operationalization of the LI model is limited to a language of physical signals as functions of time (adders followed of sigmoids or integrate-and-fire models), but the same structure of calculus can be reformulated in terms of “if-then” rules. Now feed forward and feed back receptive fields with an excitatory centre and an inhibitory periphery are defined over the data fields of a FIFO-like local STM.

With this interpretation of the calculation performed by a net of LI the associated semantics are augmented [52]. Now the “inputs” to the LI network may be not only signals of the photoreceptors but also properties detected by the ganglionar cells, or the activity of the neuronal groups of the LGB, or “excitatory and inhibitory pools” in cortex [6], for instance. In an analogous way, the “outputs” of the LI network may be considered as “decisions” associated to the state of activity of specific neuronal groups. We call this generalization of the LI ALI, as it maintains invariant the structure of the calculation of the LI but it eliminates the analytic restrictions of the contents of the “condition” and “action” fields of the rules [9,10,52,54].

With this interpretation of ALI, each element of calculus samples its data in the centre and periphery of the volume specified by its receptive field in the input space (STM) and also samples in the centre and periphery of the volume which specifies its receptive field in the output space (STM). By specifying the nature of the *decision rules*, the different types of calculus attainable by a network of algorithmic LI are obtained.

A simple example is to use ALI to select coordinates where a given individual feature reaches its maximum value in a certain local and/or temporal environment, as a WTA mechanism. There, ALI behaves as a competitive algorithm which enables comparing the values of a property in a certain region of the image,  $p_j$ , with the values in the rest of the image in order to retrieve only the coordinates of the maximum. The recurrent ALI units dialogue until only those  $x_j^{\max}$  that have a maximum input respond:

$$x_j^{\max} = \text{position } x \text{ where the measure of property } p_j(x) \text{ is maximum.}$$

For this reason, the calculation is performed on the central ( $C^*$ ) and the peripheral ( $P^*$ ) part of the output receptive field by means of the following rule:

1. Select maximum in  $C^*$ , that is  $Cm$ .
2. Select maximum of absolute values in  $P^*$ , that is  $Pm$ .
3. If  $\max(Cm, |Pm|) = Cm$  then go to (4) else go to (5).

4. Write  $Cm$  in the output space.
5. Write  $-Pm$  in the output space.

The justification for the third point of the ALI algorithm is the need that each module possesses to distinguish if an “opinion” is a proper one (+) or if it belongs to a neighbour (–). The aim of the algorithm is to detect and to propagate a local maximum ( $Cm$ ) within the receptive field and to distinguish if this maximum is a proper one (belongs to  $C^*$ ) or belongs to the neighbours (is within  $P^*$ ). If it is a proper one it is retained and propagated (write  $Cm$  in the output space). In the case of belonging to a neighbour, it is assumed and propagated (write  $-Pm$  in the output space). The essential thing is to maintain the centre/periphery distinction, to assume the maximum value (proper or belonging to the neighbours) and to propagate it with a label. In order to carry out this function the distinction between module ( $|Pm|$ ) and sign (+, –) may be used, as it is done here, or complex numbers might be used, as for instance in [10]. The usage of the sign or the module is physiologically more appropriate than the usage of complex numbers.

The output space of an ALI “circuit” represents the temporal evolution for the decision of choosing the “winning” neuron for that input. Evidently, when the receptive field covers the whole net (total connectivity in the feedback pathways), the decision is reached in a single step. But this is not the usual case in the brain. Moreover, this would mask the cooperative process inherent to a local dialogue and the diffusion of the maximum, made possible by the overlap of the receptive fields in the output space.

Fig. 11 illustrates the calculation schema and the results of the simulation for a network of 201 units with receptive field size 11 and convergence in the maximum value of the property ( $x = 41$ ) after 12 iterations. Note that the cooperative decision of ALI needs a lower number of iterations when increasing the size of the receptive field. Also note that any ALI net needs to perform on two time scales; one external for the inputs  $\{p_j(t)\}$  and another one for the competition process. While the detection and local maximum diffusion algorithm operates, the input has to remain constant, waiting for the centre-periphery dialogue of each unit with its neighbours.

The application of the previously introduced competition algorithm to our running example is shown in Fig. 12. The crosses are the result of looking for the maximum grey level value on the negative of the image, by applying the competitive ALI with a  $100 \times 100$  pixels receptive field size. The cross within the box is the result of using a greater receptive field in order to obtain one single maximum in the whole image.

WTA neural networks, a particular case of ALI, have been extensively discussed in the literature as a way of making decisions. The idea of using mutually inhibiting networks of neurons is now 30 years old [26,27]. The first use of a WTA network for SVA goes back to Koch and Ullman’s work [40]. In his selective tuning model of visual

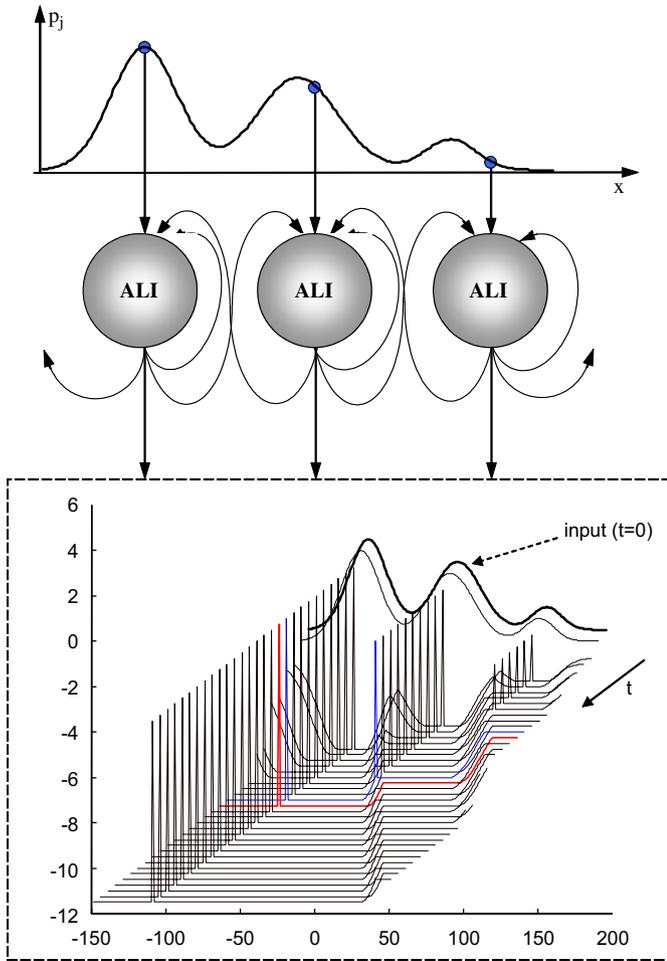


Fig. 11. Selection of coordinate  $x_j^{\max}$  where property  $p_j$  takes the maximum value in the time interval considered. Result of the simulation after 30 iterations and convergence in the maximum value starting from iteration 12, with 201 calculation units and a receptive field of 11 units (1 central and 10 peripheral). The coordinate where the value of  $p_j$  is maximum is  $x = 41$ .

attention, Tsotsos [69] uses WTA networks at multiple levels of visual processing to overcome computational complexity barriers for visual search. Also, Itti and Koch [37] introduced a detailed saliency map model to explain how bottom-up visual information could be used to direct attention towards potentially interesting stimuli in the visual field. In their model, visual features are extracted from an image and separated into various feature maps. Within each map, centre-surround competition weakens the representation of features occurring abundantly in the image and strengthens unique features. Koike and Saiki [41] proposed that a stochastic WTA enables the saliency-based search model to vary the relative saliency in order to change search efficiency, due to stochastic shifts of attention.

We have seen how, when substituting the weighted sum with a maxima selector, ALI circuits project into competitive WTA-like algorithms. In an analogous way, if substituting the maxima selector with the more general inferences of “evaluate” an expression in the centre and the

periphery, “compare” the results of these evaluations and “select” the result that best satisfy a certain criterion, then the ALI circuits convert into problem-solving methods (PSM) in the usual sense in AI [61]. In [43] we applied this PSM to a SVA system with the objective of moving objects detection, labelling, and further tracking.

### 5.2. Accumulative computation (AC)

The memory in ALI is supported by the networks of delays in the receptive fields, which store in a STM the input and output configurations of the network during the last calculation intervals. Computationally this is equivalent to simulating, in a discrete manner in terms of finite differences equations, the analogical models of “integrate-and-fire” neurons. The ALI model greatly increments its calculation capacity when substituting this STM by another transitory data storage method that admits more possibilities to manage its contents. We call this sort of memory “accumulative computation” (AC) [15–17,53], which is formed by a set of “dynamic state variables” whose activity state (attentional state) can be controlled in function of the presence, persistence or extinction of the following events:

- Evaluation of the presence of specific features in the central and the peripheral parts of a receptive field.
- Accumulation of the persistency of these features.
- Generation of an output pattern function of the accumulated value.
- Action of specific control algorithms of the charge/discharge process, including the possibility of learning, understood as a change in the value of a set of parameters (increment and decrement steps, maximum and minimum value, and time scale).

The upper part of Fig. 13 shows the AC model’s inferential schema. The model works in two time scales, a macroscopic,  $t$ , associated to the external data sequence to be processed by the net, and a microscopic one,  $\tau$ , associated to the set of internal processes that take place while the external data (an image) remain constant. The lower part of Fig. 13 illustrates the temporal evolution of the state of the charge in an AC WM in front of a particular one-dimensional stimuli sequence.

When applying this AC process to a sequence of images of the flock of birds running example, the results of Fig. 14a are found. The control rules used to calculate the persistency of motion through time are:

$$Q[x, y; t + \Delta t] = \begin{cases} \max(Q[x, y; t] - \delta Q, \min) & \text{if } C[x, y; t] = 1, \\ \min(Q[x, y; t] + \delta Q, \max) & \text{otherwise,} \end{cases}$$

where  $C[x, y; t] = 1$  means that motion has been detected on pixel  $[x, y]$  at  $t$ . Note that white colour is associated to pixels where no motion has been detected

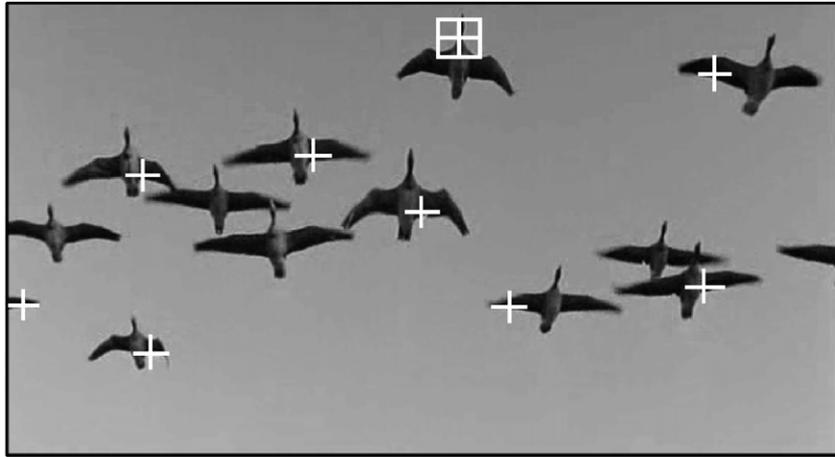


Fig. 12. Application of the ALI competition algorithm to a flock of birds.

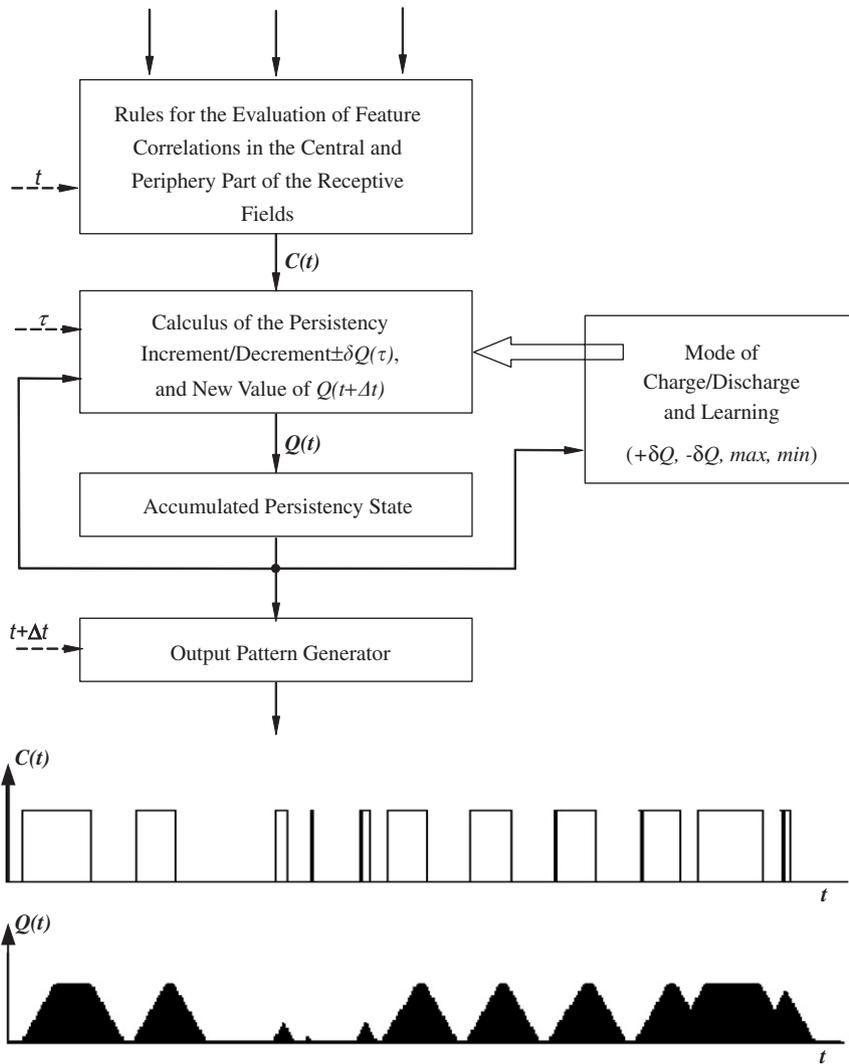


Fig. 13. The AC inferential schema (upper part) and an example of the temporal evolution of the accumulated persistency state,  $Q(t)$ , in response to a specific sequence of values of a detected feature,  $C(t)$ , (lower part).

and black is associated to pixels where motion has been detected at the current time instant. Intermediate grey colours indicate motion detected during the last previous

time instants. The result of this AC process followed by threshold provides the *Attention Focus* (Fig. 14b). Note that almost the whole background has been eliminated.

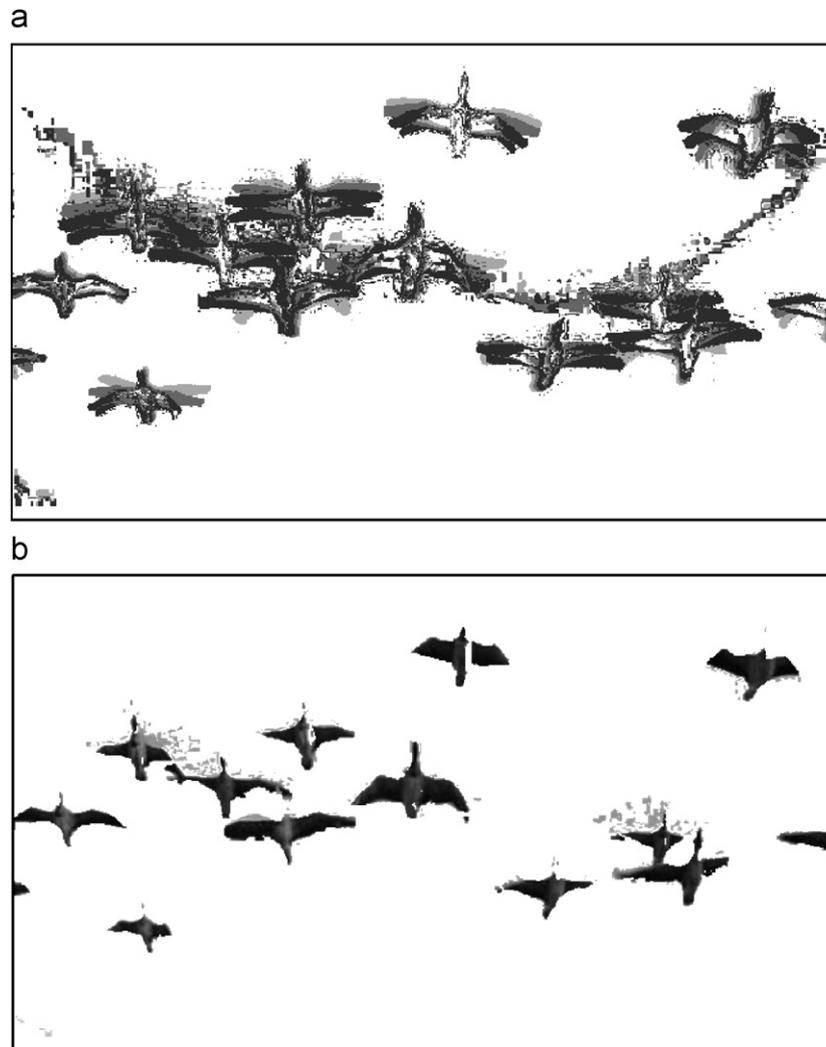


Fig. 14. Application of AC in ALI to the running example: (a) motion persistency measure and (b) attention focus at a given time instant.

The closest concept to our FIFO memory in ALI is the one of the STM. Similarly, the concept closer to our AC is the WM, as proposed by Baddeley and Hitch [4]. From the computational viewpoint WM allows keeping active a limited amount of information for a brief period of time. It was thought to have two functions: storing material that we have to recall in a few seconds and providing a gateway to LTM [1]. WM, as AC, is conceptualized as an active system for temporarily storing and manipulating information needed in the execution of complex cognitive tasks. There are two types of components: storage and central executive functions. WM may also be seen as the ability to transiently hold and manipulate goal-related information to guide forthcoming actions [3,24]. Stimulus-specific persistent neural activity is, apparently, one of the neural processes underlying active memory processes [73].

## 6. Conclusions

In this work we have integrated a great part of the research team's works [14,18,20–22,34,44,46] on SVA

processes in visual surveillance tasks, using a knowledge model based in the acceptance of the existence of a double organization of the bottom-up/top-down attentional processes [7,31,32,58,67,75] and two biologically inspired mechanisms, ALI and AC, as basic functional modules to operationalize the different inferences of the model (“evaluate”, “compare”, “select”, “dialogue”).

In addition to the possible value of the proposal in the specific field of the attentional processes, this paper exemplifies a methodological approach to CN that may be of interest in the modelling of other aspects in the integration of perception and action. We refer to the use of neuronal circuits as a source of inspiration for new algorithms and problem-solving methods of potential utility in Computer Vision and Robotics. The usual strategy in Neurodynamics [6] is to start from an anatomic circuit, to substitute the neurons by analogical integrate-and-fire like models, and to simulate the behaviour of these artificial neural networks (ANN). Our suggestion is to add to these descriptions of the neurophysiological level other symbolic ones coming from keeping the structure of the

neural networks (their connectivity) but substituting the analytic operators by inferential rules, and now simulating the behaviour of these nets of rules, whose semantics is closer to the description in natural language that an external observer emits when he sees the behaviour that emerges of the biological neural networks.

## Acknowledgments

This work has been partially funded by Spanish projects TIN2004-07661-C02 and TIN2007-67586-C02 of the Ministerio de Educación y Ciencia, integrated in AVISA and AVISADOS projects, respectively. This work is also supported in part by Junta de Comunidades de Castilla-La Mancha PBI06-099 Grant.

## References

- [1] R.C. Atkinson, R.M. Shiffrin, Human memory: a proposed system and its control processes, in: *The Psychology of Learning and Motivation: Advances in Research and Theory*, vol. 2, Academic Press, San Diego, CA, 1968.
- [2] G. Backer, B. Mertsching, Two selection stages provide efficient object-based attentional control for dynamic vision, in: *Proceedings of the International Workshop on Attention and Performance in Computer Vision, WAPCV, 2003*, pp. 9–16.
- [3] A.D. Baddeley, *Human Memory*, Lawrence Erlbaum, London, 1990.
- [4] A.D. Baddeley, G.J. Hitch, Working memory, in: *Recent Advances in Learning and Motivation*, vol. 8, Academic Press, San Diego, CA, 1974.
- [5] X. Chen, G.J. Zelinsky, Real-world visual search is dominated by top-down guidance, *Vision Res.* 46 (24) (2006) 4118–4133.
- [6] G. Deco, E.T. Rolls, Attention, short-term memory, and action selection: a unifying theory, *Prog. Neurobiol.* 76 (4) (2005) 236–256.
- [7] G. Deco, E.T. Rolls, A neurodynamical cortical model of visual attention and invariant object recognition, *Vision Res.* 44 (6) (2004) 621–642.
- [8] A.E. Delgado, *Modelos Neurocibernéticos de Dinámica Cerebral*, Ph.D. Thesis, E.T.S. Ingeniería de Telecomunicación, Universidad Politécnica, Madrid, Spain, 1978.
- [9] A.E. Delgado, J. Mira, Algorithmic lateral inhibition as a generic method for visual information processing with potential applications in robotics, in: *Computational Intelligent Systems for Applied Research*, World Scientific, Singapore, 2002, pp. 477–484.
- [10] A.E. Delgado, J. Mira, R. Moreno-Díaz, A neurocybernetic model of modal co-operative decision in the Kilmer–McCulloch space, *Kybernetes* 18 (3) (1989) 48–57.
- [11] R. Desimone, Visual attention mediated by biased competition in extrastriate visual cortex, *Philos. Trans. R. Soc. London B* 353 (1998) 1245–1255.
- [12] R. Desimone, J. Duncan, Neural mechanisms of selective attention, *Annu. Rev. Neurosci.* 18 (1995) 193–222.
- [13] J. Duncan, G.W. Humphreys, Visual search and stimulus similarity, *Psycholog. Rev.* 96 (3) (1989) 433–458.
- [14] M.A. Fernández, A. Fernández-Caballero, M.T. López, J. Mira, Length-speed ratio (LSR) as a characteristic for moving elements real-time classification, *Real-Time Imaging* 9 (1) (2003) 49–59.
- [15] M.A. Fernández, Una arquitectura modular de inspiración biológica con capacidad de aprendizaje para el análisis de movimiento en secuencias de imagen en tiempo real, Ph.D. Thesis, UNED, Madrid, Spain, 1995.
- [16] M.A. Fernández, J. Mira, M.T. López, J.R. Alvarez, A. Manjarrés, S. Barro, Local accumulation of persistent activity at synaptic level: application to motion analysis, *Lect. Notes Comput. Sci.* 930 (1995) 137–143.
- [17] M.A. Fernández, J. Mira, Permanence memory: A system for real time motion analysis in image sequences, in: *IAPR Workshop on Machine Vision Applications MVA'92, 1992*, pp. 249–252.
- [18] A. Fernández-Caballero, J. Mira, M.A. Fernández, A.E. Delgado, On motion detection through a multi-layer neural network architecture, *Neural Networks* 16 (2) (2003) 205–222.
- [19] A. Fernández-Caballero, J. Mira, A.E. Delgado, M.A. Fernández, Lateral interaction in accumulative computation: a model for motion detection, *Neurocomputing* 50 (2003) 341–364.
- [20] A. Fernández-Caballero, M.A. Fernández, J. Mira, A.E. Delgado, Spatio-temporal shape building from image sequences using lateral interaction in accumulative computation, *Pattern Recognition* 36 (5) (2003) 1131–1142.
- [21] A. Fernández-Caballero, *Modelos de interacción lateral en computación acumulativa para la obtención de siluetas*, Ph.D. Thesis, UNED, Madrid, Spain, 2001.
- [22] A. Fernández-Caballero, J. Mira, M.A. Fernández, M.T. López, Segmentation from motion of non-rigid objects by neuronal lateral interaction, *Pattern Recognition Lett.* 22 (14) (2001) 1517–1524.
- [23] C.L. Folk, A.B. Leber, H.E. Egeth, Made you blink! Contingent attentional capture produces a spatial blink, *Percept. Psychophys.* 64 (5) (2002) 741–753.
- [24] J.M. Fuster, *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe*, third ed., Lippincott-Raven, 1997.
- [25] C.D. Gilbert, Learning and receptive field plasticity, *Proc. Natl. Acad. Sci. USA* 93 (20) (1996) 10546–10547.
- [26] S. Grossberg, Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors, *Biolog. Cybernet.* 23 (3) (1976) 121–134.
- [27] S. Grossberg, Adaptive pattern classification and universal recoding: II. Feedback, expectation, olfaction, illusions, *Biolog. Cybernet.* 23 (4) (1976) 187–202.
- [28] F.H. Hamker, M. Zirnsak, V4 receptive field dynamics as predicted by a systems-level model of visual attention using feedback from the frontal eye field, *Neural Networks* 19 (9) (2006) 1371–1382.
- [29] F.H. Hamker, Modeling feature-based attention as an active top-down inference process, *BioSystems* 86 (1–3) (2006) 91–99.
- [30] F.H. Hamker, The emergence of attention by population-based inference and its role in distributed processing and cognitive control of vision, *Comput. Vision Image Understanding* 100 (1–2) (2005) 64–106.
- [31] F.H. Hamker, A dynamic model of how feature cues guide spatial attention, *Vision Res.* 44 (5) (2004) 501–521.
- [32] D. Heinke, G.W. Humphreys, C.L. Tweed, Top-down guidance of visual search: a computational account, *Visual Cogn.* 14 (4–8) (2006) 985–1005.
- [33] D. Heinke, G.W. Humphreys, Computational models of visual selective attention: a review, in: *Connectionist Models in Cognitive Psychology*, vol. 1, Part 4, 2005, pp. 273–312.
- [34] D. Heinke, G.W. Humphreys, G. diVirgilio, Modeling visual search experiments: selective attention for identification model (SAIM), *Neurocomputing* 44–46 (2002) 817–822.
- [35] M.H. Herzog, M. Fahle, The role of feedback in learning a vernier discrimination task, *Vision Res.* 37 (1997) 2133–2141.
- [36] L. Itti, C. Koch, Computational modelling of visual attention, *Nat. Rev. Neurosci.* 2 (3) (2001) 194–203.
- [37] L. Itti, C. Koch, A saliency-based search mechanism for overt and covert shifts of visual attention, *Vision Res.* 40 (2000) 1489–1506.
- [38] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (11) (1998) 1254–1259.
- [39] N. Kenner, J.M. Wolfe, An exact picture of your target guides visual search better than any other representation [abstract], *J. Vision* 3 (9) (2003) 230a.
- [40] C. Koch, S. Ullman, Shifts in selective visual attention: towards the underlying neural circuitry, *Hum. Neurobiol.* 4 (4) (1985) 219–227.

- [41] T. Koike, J. Saiki, Stochastic guided search model for search asymmetries in visual search tasks, *Lect. Notes Comput. Sci.* 2525 (2002) 408–417.
- [42] M.T. López, A. Fernández-Caballero, M.A. Fernández, J. Mira, A.E. Delgado, Algorithmic lateral inhibition formal model for real-time motion detection, *Lect. Notes Comput. Sci.* 4739 (2007) 638–645.
- [43] M.T. López, A. Fernández-Caballero, J. Mira, A.E. Delgado, M.A. Fernández, Algorithmic lateral inhibition method in dynamic and selective visual attention task: application to moving objects detection and labeling, *Expert Systems Appl.* 31 (3) (2006) 570–594.
- [44] M.T. López, A. Fernández-Caballero, M.A. Fernández, J. Mira, A.E. Delgado, Motion features to enhance scene segmentation in active visual attention, *Pattern Recognition Lett.* 27 (5) (2006) 469–478.
- [45] M.T. López, Modelado computacional de los mecanismos de atención selectiva mediante redes de interacción lateral, Ph.D. Thesis, UNED, Madrid, Spain, 2004.
- [46] J.M. López-Valles, M.A. Fernández, A. Fernández-Caballero, Stereovision depth analysis by two-dimensional motion charge memories, *Pattern Recognition Lett.* 28 (1) (2007) 20–30.
- [47] A.R. Luria, *The Working Brain: An Introduction to Neuropsychology*, Basic Books, New York, 1973.
- [48] A.R. Luria, *Cerebro y Lenguaje*. Ed. Fontanella, Barcelona, 1974.
- [49] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freeman and Company, New York, 1982.
- [50] R. Milanese, S. Gil, T. Pun, Attentive mechanisms for dynamic and static scene analysis, *Opt. Eng.* 34 (8) (1995) 2428–2434.
- [51] J. Mira, On some of the neural mechanisms underlying adaptive behaviour, *Lect. Notes Comput. Sci.* 4224 (2006) 1–15.
- [52] J. Mira, A.E. Delgado, On how the computational paradigm can help us to model and interpret the neural function, *Natural Comput.* 6 (3) (2006) 207–209.
- [53] J. Mira, M.A. Fernández, M.T. López, A.E. Delgado, A. Fernández-Caballero, A model of neural inspiration for local accumulative computation, *Lect. Notes Comput. Sci.* 2809 (2003) 427–435.
- [54] J. Mira, A.E. Delgado, What can we compute with lateral inhibition circuits?, *Lect. Notes Comput. Sci.* 2084 (2001) 38–46.
- [55] J. Mira, The semantic gap, preface of bio-inspired modeling of cognitive tasks, in: *IWINAC 2007, Lecture Notes in Computer Science*, vol. 4528, Springer, Berlin, 2007.
- [56] P. Muñoz-Sotés, et al., Estudio de los síndromes afásicos, *Ciencias Neurológ.* V (2) (1971).
- [57] P. Muñoz-Sotés, et al., Estudio analítico de los síndromes hemipléjicos, *Ciencias Neurológ.* I (2) (1973).
- [58] V. Navalpakkam, L. Itti, Modeling the influence of task on attention, *Vision Res.* 45 (2) (2005) 205–231.
- [59] A. Newell, The knowledge level, *AI Mag.* Summer (1981) 1–20.
- [60] R.H. Phaf, A.H.C. van der Heijden, P.T.W. Hudson, SLAM: a connectionist model for attention in visual selection, *Cogn. Psychol.* 22 (3) (1990) 273–341.
- [61] G. Schreiber, H. Akkermans, A. Anjewierden, R. de Hoog, N. Shadbolt, W. van de Velde, B. Wielinga, *Knowledge Engineering and Management: The CommonKads Methodology*, The MIT Press, Cambridge, MA, 2000.
- [62] A.R. Seitz, J.E. Nanez, S. Holloway, Y. Tsushima, T. Watanabe, Two cases requiring external reinforcement in perceptual learning, *J. Vision* 6 (2006) 966–973.
- [63] B. Takacs, H. Wechsler, A dynamic and multiresolution model of visual attention and its application to facial landmark detection, *Comput. Vision Image Understanding* 70 (1) (1998) 63–73.
- [64] A.M. Treisman, S. Sato, Conjunction search revisited, *J. Exp. Psychol.: Hum. Perception Performance* 16 (1990) 451–478.
- [65] A.M. Treisman, Features and objects: the fourteenth Barlett memorial lecture, *Quart. J. Exp. Psychol.* 40A (2) (1988) 201–237.
- [66] A.M. Treisman, G. Gelade, A feature-integration theory of attention, *Cogn. Psychol.* 12 (1980) 97–136.
- [67] J.K. Tsotsos, Y. Liu, J.C. Martinez-Trujillo, M. Pomplun, E. Simine, K. Zhou, Attending to visual motion, *Comput. Vision Image Understanding* 100 (1–2) (2005) 3–40.
- [68] J.K. Tsotsos, S.M. Culhane, W.Y.K. Wai, Y. Lai, N. Davis, F. Nuflo, Modeling visual attention via selective tuning, *Artif. Intell.* 78 (1–2) (1995) 507–545.
- [69] J.K. Tsotsos, Analyzing vision at the complexity level, *Behav. Brain Sci.* 13 (3) (1990) 423–445.
- [70] S. Ullman, Aligning pictorial descriptions: an approach to object recognition, *Cognition* 32 (1989) 193–254.
- [71] P. van de Laar, T. Heskes, S. Gielen, Task-dependent learning of attention, *Neural Networks* 10 (6) (1997) 981–992.
- [72] D. Walther, C. Koch, Modeling attention to salient proto-objects, *Neural Networks* 19 (2006) 1395–1407.
- [73] X.-J. Wang, Synaptic reverberation underlying mnemonic persistent activity, *Trends Neurosci.* 24 (8) (2001) 455–463.
- [74] J.M. Wolfe, Guided search 2.0. A revised model of visual search, *Psychonom. Bull. Rev.* 1 (1994) 202–238.
- [75] G.J. Zelinsky, W. Zhang, B. Yu, X. Chen, D. Samaras, The role of top-down and bottom-up processes in guiding eye movements during visual search, *Adv. Neural Inf. Process. Systems* 18 (2006) 1569–1576.



**José Mira** is professor of Computer Science and Artificial Intelligence and Head of the department of Artificial Intelligence at the National University for Distance Education (UNED) in Madrid (Spain). His research interests are neural modelling at the knowledge level and integration of symbolic and connectionist problem-solving methods in the design of Knowledge-Based Systems in the application domains of medicine, robotics and computer vision. Prof. Mira is the general Chairman of the biennial interdisciplinary meetings IWINAC (International Work Conference on the Interplay between Natural and Artificial Computation).



**Ana E. Delgado** is professor of Computer Science and Artificial Intelligence with the department of Artificial Intelligence at the National University for Distance Education (UNED) in Madrid (Spain). Her current research interests are in neural modelling, bio-inspired cooperative agents and computer vision.



**María T. López** received her degree in Physics from the University of Valencia, Spain, in 1991, and received her Ph.D. from the Department of Artificial Intelligence of the National University for Distance Education, Spain, in 2004. Since 1991, she is an Assistant Professor with the Department of Computer Science at the University of Castilla-La Mancha, Spain. Her research interests are in image processing and computer vision. María T. López is member of the IAPR.



**Antonio Fernández-Caballero** received his degree in Computer Science from the Technical University of Madrid, Spain, in 1993, and received his Ph.D. from the Department of Artificial Intelligence of the National University for Distance Education, Spain, in 2001. Since 1995, he is an Associate Professor with the Department of Computer Science at the University of Castilla-La Mancha, Spain. His research interests are in image processing, computer vision, neural networks, and agent technology. A. Fernández-Caballero is member of the IAPR.



**Miguel A. Fernández** received his degree in Physics from the University of Granada, Spain, in 1987, and received his Ph.D. from the Department of Artificial Intelligence of the National University for Distance Education, Spain, in 1995. Since 1991, he is an Assistant Professor with the Department of Computer Science at the University of Castilla-La Mancha, Spain. His research interests are in artificial intelligence, image processing, computer vision, and neural networks.