

Interactive Learning of Top-down Attention Control and Motor Actions

Ali Borji¹, Majid Nili Ahmadabadi^{1,2}, Babak Nadjar Araabi^{1,2}

¹School of Cognitive Sciences, Institute for Studies in Theoretical Physics and Mathematics
Niavaran Bldg. P.O.Box 19395-5746, Tehran, IRAN

²Department of Electrical and Computer Engineering, University of Tehran, Tehran, IRAN
borji@ipm.ir, {mnili, araabi}@ut.ac.ir

Abstract. Like humans and primates, artificial creatures like robots are limited in terms of allocation of their resources to huge sensory and perceptual information. Thus attention is regarded as the same solution as humans in this domain. While bottom-up attention is determined by the image statistics, top down attention is dependent on behavior and the task an agent is doing. This work attempts to consider a task based top-down visual attention control when resources of the agent are limited. Particularly attention control is formulated as an optimization problem in which the agent has to gain maximum reward while satisfying a constraint which is its information processing bottleneck. Reinforcement learning is then used to solve that optimization problem. A driving environment is simulated in that agent has to learn how to drive safely by attending to the right spatial locations and performing appropriate motor actions.

1 Introduction

Possible confusion because of huge sensory space, limited response time, dynamicity of perceptual space and environment, and accuracy and reliability of sensors causes a bounded rationality for an agent instead of full rationality. In this regard attention control could act like enhancing the rationality of the agent by proposing small amount of information for further processing. For the agent to be successful in terms of receiving maximum cumulative reward, it should be able to perform perceptual actions or attentions as well as motor actions at the same time. These perceptual actions are available at several forms like where to look and what to look in visual modality. In situations where the environment is unfamiliar or not clearly defined or there is a small training dataset, fixed design of a control strategy is helpless. Therefore interactive learning methods like reinforcement learning are the first options in these cases.

In order to better understand the mechanisms of attention, it is not just enough to examine simple synthetic or natural stimuli in laboratory because attention behavior under top down control is mostly determined by task demands and characteristics which change through time in spatial and temporal manner. To better cope with this

type of attention, we should consider embodiment and situatedness which are two main concepts of new trends in artificial intelligence [1]. The main claim behind this concept is that behaviors like attention, emotion, etc evolve depending body characteristics (dynamics, form, etc) and the environment which agents live in. Thus attention could not be studied and understood in an abstract manner. Therefore the study of attention is much influenced by actions which are based on visual information like grasping, etc. For instance to drive safely needs very complex behaviors like detecting and monitoring stop signs, traffic lights, other cars as well coordinating actions and attentions. That's why novice drivers usually consider driving in crowded big cities hard and challenging.

In this work we consider attention control as an optimization problem. In this optimization framework an artificial agent has a certain limited amount of processing power which he should devote them to the most appropriate items from the sensory space. Particularly our aim is to propose a mechanism for learning top top-down visual attention control for an agent performing tasks which needs motor actions in natural interactive environments like driving. Both top down attention control signals and motor actions will be learned concurrently to fulfill task demands. Our approach is motivated by studies in neuroscience and psychology of attention.

2 Related Works

A broad range of works in modeling visual attention are reported in the literature. In this section we review those works which are related to ours, especially those which have considered learning aspects of attention control. Unfortunately, there are a few researches on learning and formation of attention control; rather they are mostly related to the attention modeling. Our focus will be to review those works which has followed a learning based approach to develop visual routines specially attention. In this regard we look at those works which have merged those two topics.

One of the earliest studies which showed that attention is task based and is controlled top-down is done by Yarbus [2]. He showed that task has a great effect on the pattern of eye movements. Bottom-up cues are important in free-viewing tasks and in such situations attention is attracted to more salient regions of the scene. However top-down influences come to play role when subjects are doing a specific task or process the scene based on their knowledge and motivations. Some previous works have tried to implement top-down influences. One drawback of those approaches is that they have tackled the problem in a very limited and abstract way and have not considered the temporal aspects of top-cues such as task demands on attention which happens in our every day behavior. In brain these processes are done mainly in frontal cortex where short and long term memories are stored and our intentions and plans are formed. It is clear that attention is not such an episodic process but is also dependent on temporal manner the stimuli are presented and also temporal needs and internal states of the agent. This approach has been followed up in active vision systems for building robots able to control their gazes according to their tasks.

One of the outstanding works on visual attention which is based on an earlier model of visual attention developed by Koch and Ullman [3] is proposed in [4]. Simplicity and little processing time are two main advantages of this model. It is continuously updated and has been the basis of many other models. It's called saliency based model and have been used to explain behavioral data on simple synthetic static search arrays to dynamic natural stimuli like movies and interactive ones like game playing [5].

In [6], a reinforcement learning method is proposed to model the reading task. A model for mimicking an expert reader is introduced by predicting where eyes should look (spatial attention) and how long they should stop there in order to lead to the best comprehension from the text. This is an interesting work which proposes an optimization framework for attention and then uses RL to solve it. Our approach is the same, but with the difference that we are aim to apply it to complex tasks where agents have to perform physical actions while controlling their sensory acquisition behavior. Furthermore this research has the limited generalization and could only explain reading experiments.

In [7, 8], a 3-step-architecture is presented which first extracts attention focus according to information theoretic saliency measures. Then, by searching in pre-specified areas from the first step, decides whether the object is available in the image or not and finally suggests a shift for attention. The final step is done using Q-learning with the goal to find the best perceptual action according to the search task. This research is related to our work because it also couples decision making and attention control and uses a reinforcement based learning approach. But, due to the simple task of search in a limited image database, the approach is very simple and the application domain is limited.

In [9], an approach for attention control is presented in a robotic platform with neck, eyes and arms. The first approach is a simple feed forward method which uses back-propagation learning algorithm while the second one uses reinforcement learning and a finite state machine for state space representation. The robot has 3 types of actions: attention shift, visual improvement and haptic improvement. Their results confirm that the second approach generates a better performance in terms of finding previously observed objects even with fewer movements in head and neck and also shifting the attention.

In [10], a new model for human eye movements is proposed that directly ties eye movements to the ongoing demands of behavior. The basic idea is that eye movements serve to reduce uncertainty about environmental variables that are task relevant. A value is assigned to an eye movement by estimating the expected cost of the uncertainty that will result if the movement is not made. If there are several candidate eye movements, the one with the highest expected value is chosen. The model is illustrated using a humanoid graphical figure that navigates on a sidewalk in a virtual urban environment.

An approach for learning gaze control for a mobile robot is proposed in [11], which proposes a model of selective attention for visual search tasks. They have implemented their model using a fixed pan-tilt-zoom camera in a visually cluttered lab environment, which samples the environment at discrete time steps. The agent has to decide where to fixate next purely based on visual information, in order to reach the region where a target object is most likely to be found. The model consists of two

interacting modules. A reinforcement learning module learns a policy on a set of regions in the room for reaching the target object, using as objective function which is the expected value of the sum of discounted rewards. By selecting an appropriate gaze direction at each step, this module provides top-down control in the selection of the next fixation point. The second module performs “within fixation” processing, based exclusively on visual information. An interesting point with this work is that it has incorporated learning where to look in a simple visual search task. Main feature of this work is its implementation on a real robotic agent working in a natural environment.

A study reported in [12], supports the idea that attention could be learned by past experience. In a behavioral task authors observed that subjects showed better performance in successive trials, which evidently is in favor of learning attention control. A modeling work trying to explain such data is done in [13]. They have proposed an optimization framework to minimize an objective function which is the sum over reaction time of each state weighted by the probability of that state to occur. Then by means of a Bayesian belief network they minimized this objective function.

A biologically motivated framework for learning attention control and decision making is reported in [14]. Physical actions and attentions are integrated in augmented actions and reinforcement learning is used to solve an optimization problem which is to maximize the cumulated reward. Mirror neurons are proposed for concept extraction and knowledge abstraction to deliver to other environments and tasks. Our work follows the integration of attention control and action selection for an artificial agent performing a complex cognitive task, but with the distinction of adding rich visual sensory information, achieving complex tasks in natural interactive environments, coping with biology and incorporating other sources which influence visual attention.

In regard with the above researches, we consider attention control and decision making in a unified framework and formularize them as an optimization problem where agent has a limited number of processing resources while facing a lot of sensory information. Then by incorporating reinforcement learning, we try to solve that optimization problem. Some advantages of such a learning approach are easy generalization, knowledge transfer and no need for exact modeling of the environment in advance.

3 Proposed Method

Agent is to learn performing physical actions and visual attentions simultaneously in a working environment in order to optimize an objective (goal) function. Agent captures visual information from the scene in the form of still images or video through its visual sensor. Captured visual scene undergoes an early visual processing. This stage contains two units which act in parallel. According to psychophysical findings, we are not blind outside our focus of attention, and information outside the focus of attention is also processed but not to the extent of the attended area. A bottom-up visual attention unit selects a subset of image and GIST extractor performs a rough categorization of the image [15].

Visual information from attended area, visual sensor and GIST extractor enters to the higher visual processing unit. This unit is aimed to derive cognitive information useful for decision making (eg. scene interpretation, object recognition, etc). Processing in this unit are dependent on the environment the agent is living and the type of tasks it is supposed to do. Although other higher processes like emotions, internal events, etc are also important in determination of visual attention, we limit ourselves here to those cues produced only based on visual scene and task demands.

Next, based on this information a decision has to be made. A state extractor unit in decision making state gets input from the higher cognitive unit as well as GIST extractor which determines spatial context, and outputs a well defined state. Reinforcement learning is used to find the best policy which optimizes an objective function. An augmented action is generated which consists of a motor action and perceptual action (top-down attention signal). A reinforcement signal which is the effect of the motor action and implicitly attention is fed back to RL unit by a critic, which evaluates this action. Reinforcement learning is used here because of its flexibility, power where critic's information is limited as well as its biological plausibility.

Theoretically any bottom-up visual attention unit could be used in early visual processing stage. For example to extend the model to work with natural scenes, saliency based model of visual attention could be used. Overall model is summarized in figure 1.

3.1 Reinforcement Learning

In RL [16], the environment is traditionally modeled as a Markov Decision Process (MDP) [17]. An MDP is a 4-tuple (S, A, P, R) , where S is a finite set of states, A is a finite set of actions, P is a probabilistic transition function from $S \times A$ to S , and R is a reinforcement function from $S \times A$ to \mathbb{R} . An MDP obeys the following discrete-time dynamics. If at time t , the agent takes the action a_t , while the environment lies in a state s_t , the agent perceives a numerical reinforcement, then reaches some state s_{t+1} with probability $P(s_{t+1}|s_t, a_t)$. The definition of MDPs assumes the full observability of the state space, which means that the agent is able to distinguish between the states of the environment using only its sensors. Let us define the perceptual space P as the set of possible percepts that sensors can return. In visual tasks, P is a set of images. So, from the point of view of the agent, an interaction with the environment is defined as a quadruple, where p_t (also p_{t+1}) is the percept furnished by its sensors in the presence of the state s_t (also s_{t+1}). A percept-to-action mapping is a fixed probabilistic function from percepts to actions. A percept-to-actions mapping tells the agent the probability with which it should choose an action when faced with some percepts. In RL terminology, such a mapping is called stationary Markovian control policy. For an infinite sequence of interactions starting in a state s_0 , the discounted return is the discount factor that gives the current value of the future reinforcements. The Markovian Decision Problem is to find an optimal percept-to-action mapping that maximizes the expected discounted return.

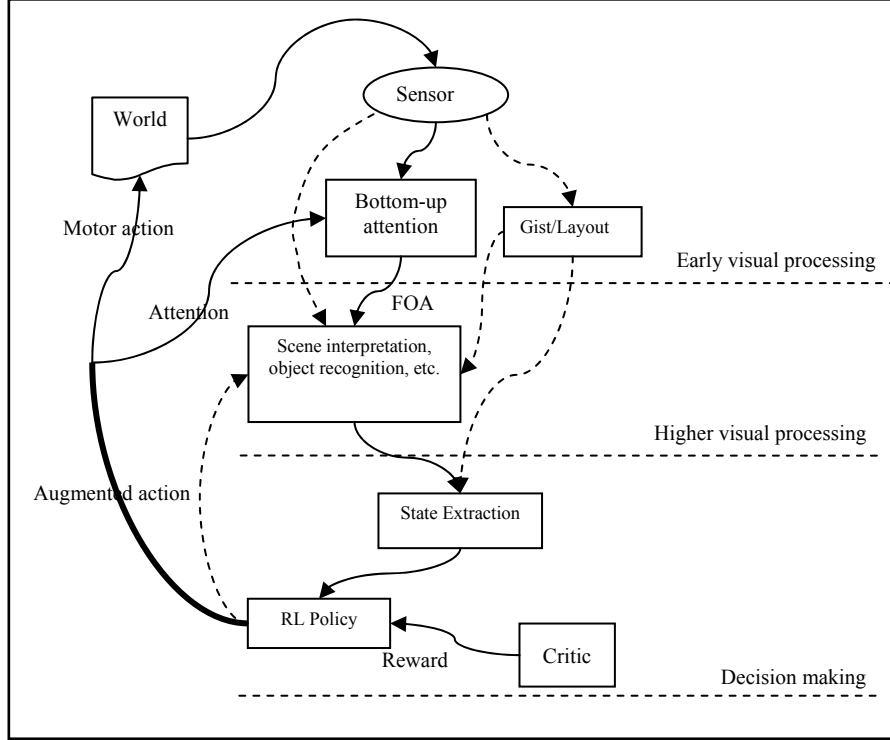


Fig. 1. Our model for learning top-down visual attention control. World is the environment agent is doing a behavior or a task in it. A visual stimulus is captured by the visual sensor of the agent. This stimulus is processed by an early visual system which could be artificial counterpart of V1 and V2 of the primary visual cortex and also by a bottom-up visual attention system. A saliency map is the output of this unit. Evidences exist for availability of such a saliency map in multiple brain areas like FEF, SC, LIP and V4. Another top-down component is based on the idea of “Gist,” which in psychophysical terms is the ability of people to roughly describe the type and overall layout of an image after only a very brief presentation, and to use this information to guide subsequent target searches. Higher cognitive processes which are believed to occur in higher visual areas V4/IT perform a variety of cognitive visual processes like scene understanding, object recognition, etc. Conceptual output of this unit is transferred to prefrontal, posterior parietal, LIP, etc which perform a sensory motor processing and action selection. Two control signals (motor action and attention) are generated which affect the world and bottom-up visual attention system. A critic assigns a reinforcement signal which evaluates the actions. Dashed lines in the figure are not implemented yet.

3.2 Mathematical Formulation

Top-down attention is represented as a vector with n elements:

$$\overline{\text{Attention}} = \{\overline{\text{Att}}_1, \overline{\text{Att}}_2, \dots, \overline{\text{Att}}_n\} \quad (1)$$

with each \overline{Att}_i being the parametr vector of the bottom-up visual attention system. In general extracting area of interest (attended area) is a function f which takes an image I and a top-down attention signal $\overline{Attention}$ as input and outputs a salient region.

$$\overline{AOI} = f(I, \overline{Attention}) \quad (2)$$

Higher level processing is modeled as a function g which takes attended area (AOI) and GIST as input and outputs a state S (higher visual information and information about scene):

$$S = g(\overline{AOI}, GIST, I) \quad (3)$$

S is a cognitive representation of the scene rich enough to help the agent make decisions. If necessary the scene representation should be transformed into a state by another mapping. After these steps, an RL unit produces an augmented action a , which is composed of a motor action m and a perceptual action b . This attention vector biases the bottom up model of attention. This command, could act as an explicit spatial shift toward an object or a position with high value in the saliency map, a shift in weightings of different features or modalities, modulation of a region in the visual field, etc. State of the agent is its subjective representation of the external world and itself at every decision making point. This representation should be rich enough to support bounded rational decision making.

In learning component (RL), attention control and motor responses are optimized gradually and interactively. This module is responsible for learning optimal decisions at each state S . The optimal decision is found using an RL method which seeks a decision making strategy (policy) at each state that maximizes agent's expected return. In our model we use discounted expected reward as expected return and Q-learning because of its biological plausibility.

$A(s)$ is the set of all possible actions in each state, $M(s)$ is the set of motor actions and $B(s)$ is the set of perceptual actions. Null is added to both types of actions because in some situations no action or attention is needed.

$$A(s) = \{M(S) \cup \text{Null}\} \times \{B(S) \cup \text{Null}\} \quad (4)$$

For agent to find its optimum action selection policy $\pi(s, a)$ to maximize the expected value of the reward signals r it receives, uses a Q-learning algorithm. The value of taking action a in state s under a policy π is denoted by $Q(s, a)$ and is updated in each time step according to following equation:

$$Q(S_t, a_t) = (1 - \alpha)Q(S_t, a_t) + \alpha(r_t + \gamma \max_a Q(S_{t-1}, a)) \quad (5)$$

in above formula, α is the learning rate and γ is the discount factor, $\pi(s, a)$ (probability of taking action a given state s) is then determined by the values of $Q^{\pi(s, a)}$ in a softmax or ϵ -greedy manner. Algorithm of the discussed attention control system is summarized in table 1.

Table 1. Algorithm for learning top-down visual attention control

Algorithm	
1	Get a visual image from the visual sensor I
2	Apply the top-down attention signal b over the input image (bias the bottom up model) and report the attended area AOI
3	Perform a higher cognitive processing over the attended area and derive a state s.
4	Based on policy $\pi(s)$, select an action $a = (m, b)$, $m \in M$, $b \in B$.
5	Get a reward r from the critic.
6	Update the current policy $\pi(s)$ by updating the value of $Q(s, a)$ by equation x.
7	If learning is not converged go to step one
8	Stop.

4 Experiments and Results

In order investigate the efficiency of the proposed method, we applied it to a simulated driving experiment. In this experiment, agent tries to discover rules and spatial locations of interest of an unknown environment by performing actions and attentions until it learn a set of best actions and attentions in each state. A mobile robot with a limited field of view (FOV) is supposed to learn how to drive in a simulated driving site and extract the rules which are set by the designer in advance. Environment is a simulated driving world with crosses, three ways, turns, etc. Like real world situation some rules for driving are imposed in this environment which if driving agent ignores them will receive punishments. In figure 2 visual field of view of the agent is shown. Visual sensor of the agent captures a scene from this field. The set of locations that agent (his field of view) sees at any moment are:

$$\overline{\text{FOV}} = \{\text{forward left (FL), far forward (FF), forward right (FR), left (L), forward(F), right(R)}\} \quad (6)$$

This large set of visual data makes an attention mechanism more necessary and useful. As it is assumed that one location could be attended in any moment because of agent's limited processing power. Attention and motor actions the agent could perform are as follows:

$$\overline{\text{Attention}} = \{\text{FL, FF, FR, L, F, R, Null}\} \quad (7)$$

$$\overline{\text{Action}} = \{\text{go forward(F), turn right(TR), turn left (TL), Null}\} \quad (8)$$

Each cell in the grid is represented by '0' if it is open (white cell making the way) and '1' if it is occupied. A state is represented by the contents of the cells it observes augmented by another value which is the content of the cell (value of the sign) at the attended location.

$$\overline{\text{State}} = \{\text{FL, FF, FR, L, F, R, sign value at attended location}\} \quad (9)$$

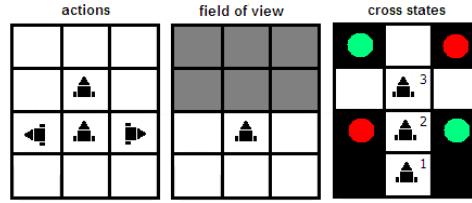


Fig. 2. Motor actions of the agent are shown at the left. Agent at any moment could choose to either turn left or right, go ahead or stay at its current position. Turning moves the agent one block to the direction of turn. At right is shown field of view of robot (FOV). Agent sees all locations in its visual field but could only attend to one location because of limited processing power. Right side shows the states of the agent behind a cross represented at table 2. In four sides of the cross are signs which alarm the status of the sign. Agent must learn to attend to the sign at its forward right and perform the right action in each situation. Signs are randomly changed in our simulation to generate all possible of states.

A ‘9’ at the end of a state means the value of the sign is not relevant and does not change the state of the agent. For example, consider the situation of the robot behind a cross shown in the right side of figure 2. Moving from position one to position two, agent’s state changes to 1010004 provided that attention is FR and light is red which is coded here by a ‘4’ and 1010005 if attention is FR and light is green. Driver must attend to the light in its forward right position and stay behind the cross if the light is red and pass if it gets green.

Table 2. States of the robot behind the cross and their right actions and attentions that must be performed as well as their associated rewards at figure 2. Agents was also punished ($r=-2$) whenever it went off-road. ‘Rep.’ stands for ‘Representation’.

Rule	Rep.	Action	Attention	Reward	Comment
1	0001019	F	FR	+3	Position 1 behind cross
2	1010002	N	FR	+3	Position 2 behind cross, red light
3	1010003	F	N	+2	Position 2 behind cross, green light
4	1011012	F	N	+2	Position 3 in cross, red light
5	1011013	F	N	+2	Position 3 in cross, green light

Some of the simulated driving environments are illustrated in figure 3. Figure 4 shows the learning curves of a sample run.

5 Conclusions and Future Works

In this paper, we proposed top-down model of attention control that is in cross roads of neuroscience and engineering. For that, we considered findings from both artificial intelligence and neuroscience. From AI point of view, many previous works (like RL, planning, probabilistic frameworks, etc) has tried to propose methods for performing complex tasks. “How complex tasks could be learned?” has been the question of the past research in AI. Now we state this question as “How complex tasks could be

learned when an agent has bounded rationality because of bounded sensory power and how does it affect its behavior?”

Our proposed solution is in abstract form. Next step will be to apply it to visual complex tasks where agent has to perform a richer set of physical actions in natural or synthetic visual interactive environments. Another extension is to implement the units of the model presented at figure 1 like GIST extractor. Fortunately recently in [15] a fast approach has been proposed for that purpose. It also remains to define and implement higher cognitive processes in higher visual processing unit of the model.

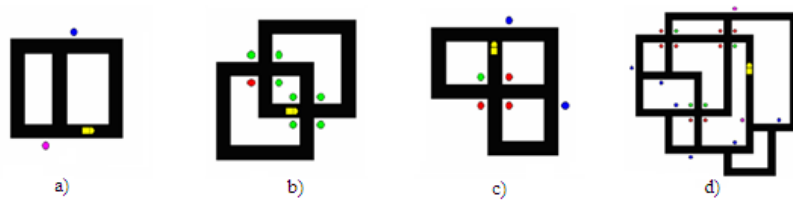


Fig. 3. Examples of simulated driving environments from simple to complex (a to d). Environment could have crosses, turns, three ways etc. Rules for attention and action are associated with each state. Agent has to learn such rules in order to gain maximum reward. Lights were randomly changed in simulation in order to produce all combinations of states. Agent could learn in simple environments and later benefit from that knowledge in more complex ones by transferring its knowledge or it could start by learning in the complex environments.

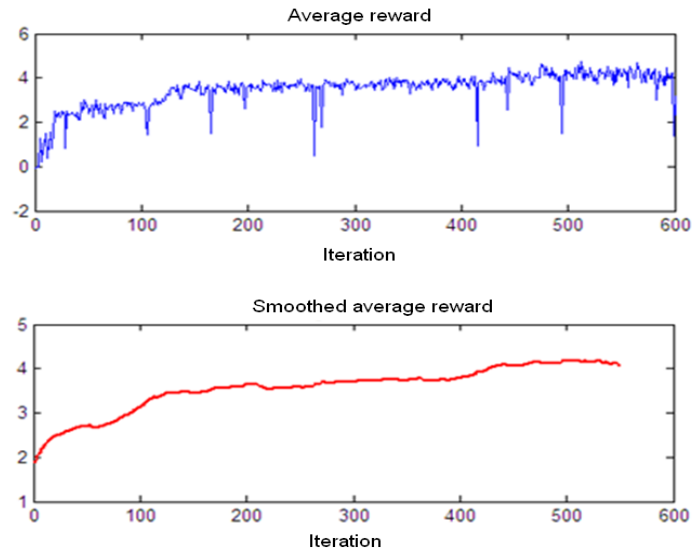


Fig. 4. Learning curves associated with the driving environment c in figure 3. An ϵ -greedy algorithm was used for action selection. Top row shows the average reward in each learning episode and the bottom row shows the average of the top row in a 50 iterations window. Parameters of the Q-value updating formula were $\alpha=0.3$ and $\gamma=0.9$.

References

1. Clark, A., and Grush, R., Towards a cognitive robotics. *Adaptive Behavior*, 7: 5, 1999.
2. Yarbus, A.L. Eye movements during perception of complex objects, in L. A. Riggs, ed., *Eye Movements and Vision*, Plenum Press, New York, chapter VII, pp. 171-196, 1967.
3. Koch, C. and Ullman, S., Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry, *Human Neurobiology*, 4:219–227, 1985.
4. Itti, L., Koch, C., Niebur, E., A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11) 1254-1259, 1998.
5. Peters, R. J., Itti, L., Applying computational tools to predict gaze direction in interactive visual environments, *ACM Transactions on Applied Perception*, Vol. 5, No. 2, p. Article 8, 2008.
6. Reichle, E.D. and Patryk, L.A., Using Reinforcement Learning to Understand the Emergence of Intelligent Eye-Movement Behavior During Reading, *Psychological Review* Copyright 2006 by the American Psychological Association Vol. 113, No. 2, 390- 408, (2006).
7. Fritz, G., Seifert, C., Paletta, L., and Bischof, H., Attentive Object detection Using an Information Theoretic Saliency Measure, *WAPCV 2004*: 29-41, 2004.
8. Paletta, L., Fritz, G., and Seifert, C., Cascaded Sequential Attention for Object Recognition with Informative Local Descriptors and Q-learning of Grouping Strategies , *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.
9. Gonic, L.M.G., Giraldi, G.A., Oliveira, A. AF., and Grupen, P.A., Learning Policies for Attentional Control, *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, pp 294 - 299, 1999.
10. Sprague, N. and Ballard, D.H. Modeling Embodied Visual Behaviors. 2005.
11. Minut, S. and Mahadevan, S., A Reinforcement Learning Model of Selective Visual Attention, *Fifth International Conference on Autonomous Agents*, Montreal, 2001.
12. Maljkovic, V., and Nakayama, K. Priming of pop-out: I. Role of features. *Memory and Cognition*, 22, 657-672. 1994.
13. Mozer, M.C., Shettel, M., and Vecera, S.P. Control of visual attention: A rational account. In Y. Weiss, B. Schoelkopf, & J. Platt (Eds.), *Neural Information Processing Systems 18* (pp. 923-930). Cambridge, MA: MIT Press. 2006.
14. Shariatpanahi, H.F., Ahmadabadi, M.N., A Biologically Inspired Framework for Learning and Abstract Representation of Attention Control, *Proceedings of International Workshop on Attention in Cognitive Systems*, at *IJCAI 2007*, Hyderabad, India, pp. 63-80, 2007.
15. Oliva, A. and Torralba, A. Building the Gist of a Scene: The Role of Global Image Features in Recognition. *Progress in Brain Research: Visual perception*, 155, 23-36. 2006.
16. Sutton, R.S., Barto, A.G., *Reinforcement Learning*. The MIT Press. Cambridge, MA, 1998.
17. Putterman, M.L., *Markov Decision Processes*. John Wiley and Sons, New York, NY, 1994.