ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, Issue 2, February 2013

A Review of a Goal Directed Visual Perception System using Object based Top down Attention

Aniket D. Pathak, Priti Subramanium

***Abstract*—The tendency of the human being to apply the selective attention mechanism so as to determine about a truly intelligent perception system, which has the cognitive capability of learning and thinking about how to perceive the environment on its own. There are two attention mechanisms involved one of which is the top–down and the other bottom–up that correspond to the goal-directed and automatic perceptual behaviors, respectively. In this paper we review an artificial system with goal-directed visual perception approach and which uses the object-based top–down visual attention mechanism. This system will able us to determine the perception to an object of interest according to the current task, context and learned knowledge. This system can be mainly divided into three successive stages: first one is preattentive processing, second one top–down attentional selection and last is post-attentive perception. In first stage that is preattentive processing we consider an input scene which gets divided into what we say similar proto-objects, out of these one is then selected by applying the top–down attention and finally it is sent to the post-attentive perception stage for analysis and final outcome.**

***Index Terms*— Visual Perception, Goal Directed approach, Top –down attention, Object-based attention**

I. INTRODUCTION

In the field of artificial intelligence to perceive the environment autonomously is a very crucial and important issue. We know that using the selective attention mechanism the human being can coordinate the processes of the perception, action and learning and also it helps them to think and learn about how to perceive the environment on their own. For this reason only this paper call for a reviewing of the visual attention based perception mechanism that we say as cognitive visual perception. This mechanism also goes in parallel with the concept or the theme of autonomous mental development (AMD) [12] in the sense that there are two type of cognitive mappings involved: One of which sensory inputs with the internal attentional states, and the second one is attentional states with the actions that includes external and internal actions. We also review the two types of visual perception mechanism that mainly involves: the goal-directed behavior that deals with perception based on the current task and learned knowledge, and automatic behavior that deals with the perception in which any situation can occur at random which might also is unusual.

***Manuscript received Feb, 2013*.**

***Aniket D. Pathak**, Department of Computer Science and Engineering, S.S.G.B.C.O.E.T. Bhusawal, India,*

***Priti Subramanium**, Department of Computer Science and Engineering, S.S.G.B.C.O.E.T . Bhusawal, India,*

II. BACKGROUND

A. Psychological theories of visual attention

Feature integration theory is mainly used to explain the space-based bottom–up attention [1]. It states that the visual scene is at beginning is initially coded with a variety of feature dimensions that are present, only after then is attentional competition performs in a location-based serial fashion by combining all features spatially, and focal attention finally provides a way to integrate the initially separated features into a whole object.

The biased competition hypothesis states that selection based on attention, without considering space-based or object-based, is said to be a biased competition process [3]. Competition is biased in part by the bottom–up mechanism that favors a local in homogeneity in the spatial and temporal context and in part by the top–down mechanism that favors items relative to the current task.

The *Integrated Competition hypothesis* [4], [5] was further presented to explain the object-based attention mechanism. This hypothesis says that any property of an object can be used either as a task-relevant feature so as to guide the top–down attention and the whole object can be attended once the task-relevant feature successfully captures the attention.

Guided Search Model was then afterwards proposed so as to model the space-based top–down attention mechanism in accordance with bottom–up attention [2]. The GSM states that the top–down request for a given feature will activate the locations that might contain that feature.

B. The Goal Directed Approach

The term "goal-directed" used over here refers to either of two separate process types – motor processes or decision making processes. Goal-directed process encourages action by the integration of an expectation that a specific action will have a specific outcome and desire for that outcome.

C. Visual Perception: A Cognitive Process

To perceive information in this world, the brain gathers and processes information it receives from the five senses. Visual perception is also a critical part of this process and it should

not be considered as simply a passive recording of visible material. We do not always see things the way they are or as they relate to their environment. Only a part of what is perceived derives straight from our visual system.

As far as the visual system is concerned, perception is purposeful and selective. The selectivity of our visual perception is greatly dependent two distinct things one is 'attention' and another is 'visual search'. The attention part involves a kind of focalization on important concepts and key field of the visual field and the periphery of the visual field, whereas the visual search includes the process of linking several fixed parts on the same visual scene to allow more detailed view to explore. The integration of all these fixations is an immediate and instinctive process that creates what we call our vision of an image.

The main elements that we visually perceive are as follows

- *Luminosity:* defined as the response of the visual system to the actual quantity intensity of light sent out by an object.
- *Contrast*: defined as the response of the visual system to the interaction of luminosity and edges.
- *Color Vision:* defined as the response of the visual system to the wavelength of light rays sent out or reflected by objects.
- *Visual Edges:* defined as the response of the visual system to the spatial distribution of light, meaning the spatial limits of objects, their visual edges, outlines.

These elements are never perceived in isolation but always in relation to each other, they are produced simultaneously and therefore, the perception of each has an effect on the perception of the others.

Visual attention selectivity can be either covert to drive and guide eye movements, picking up useful information over time, or overt, internally shifting the focus of attention from one locus to another without eye movements.

To model a visual attention is a challenging problem for machine vision. Three closely-related basic questions can be asked so that the problem is immediately identifiable:

1) How the visual system can know what information is important enough to capture attention?

2) How does the visual system know when and how to direct attention and choose important information rather than doing so at random times and by random selection

3) Where is (are) the next potential target(s) of visual attention shifts? That is, how does attention know where to go and what to do next?

Mobile robot vision systems require simple and efficient algorithms as they carry limited computational resources in order to reduce energy requirements and construction cost. Visual attention decides to mimic the ability of natural vision systems to select just the relevant aspects from the visual input [7]. This helps reducing the processing time needed for high level object recognition tasks for set of given input images. Thus attention can play an important role in making vision systems work in real time.

Generally, we begin an attention algorithm with the computation of saliencies in an image with respect to different features such as symmetry, eccentricity, color contrast, and orientation, etc. Artificial visual attention has been a topic of interest for many computer vision applications. Visual attention algorithms have been incorporated in image compression techniques such as JPEG 2000 and many more, in order to improve view quality of important objects in compressed images.

We see a variety of models of that uses the space-based attention in machine vision have been proposed. [8] Shows built a space-based bottom–up attention model based on Feature Integration Theory. The only surprise mechanism [9], [13] was then further proposed to model the bottom–up attention in terms of both spatial and temporal context.

Various people have proposed or either prepared various models some of them are as follows. Frintrop [11] suggested a visual attention system that was meant for robots by combination of bottom–up and top–down attention. Tsotsos in [14] has presented a model that selectively tunes the visual processing networks by a top–down hierarchy of winner-take-all processes. Belardinelli in [15], [16] then showed a visual attention model for robots in the spatial and temporal context by integrating both bottom– up and top–down attention. Recently, Chikkerur [17] suggested an interesting computational method that models attention as a Bayesian inference process.

Alternatively, Sun and Fisher [11] have also a proposed computational method for object-based bottom–up attention. Since the preattentive segmentation is manually achieved in the original work, Sun's model was further improved in [19] by integrating an automatic segmentation algorithm. Some object-based visual attention models [7], [20] have also been presented. However, the top–down attention is not fully achieved in these existing object- based models, e.g., how to get the task-relevant feature is not realized.

It is concluded in [21] that attentional control in computer vision has a strong influence from research on natural attention and remain the same in future also. One of the most substantial results of the attention research is the statement that covert visual attention examines potential candidates for a gaze shift and eye movements are only executed to interesting regions. Following the information from [22] we get that for saliency based visual attention model, visual input is first decomposed into a set of topographic feature maps. Different spatial locations then compete for saliency within each map, such that only locations which locally stand out from their surround can persist. All feature maps feed, in a purely bottom-up manner, into a master "saliency map," which topographically codes for local conspicuity over the entire visual scene.

III. SYSTEM OVERVIEW

On reviewing the system we came to a conclusion that the system framework is mainly divided in three stages.

1. The *preattentive processing* stage
2. The *top–down attentional selection* stage
3. The *post-attentive perception* stage

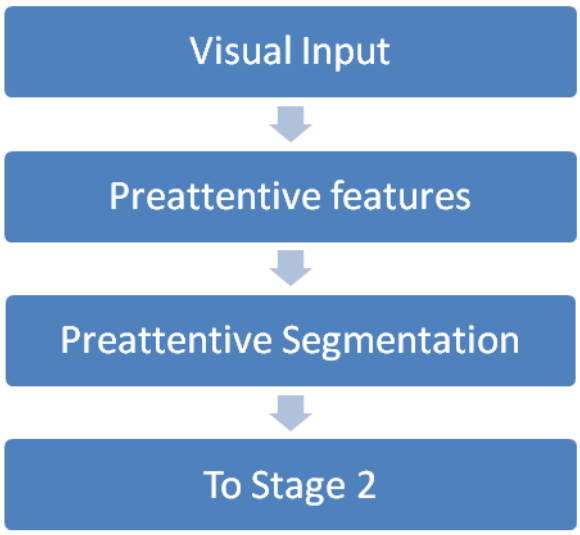

Figure 1 Preattentive Processing

The ***preattentive processing stage*** includes two back to back steps. The first step is the extraction of low-level preattentive features such as luminosity, and other features at multiple scales. The preattentive features include intensity, red–green, blue–yellow, orientation energy with, and contour. The second step is the preattentive segmentation that divides the scene into proto-objects in an unsupervised manner.

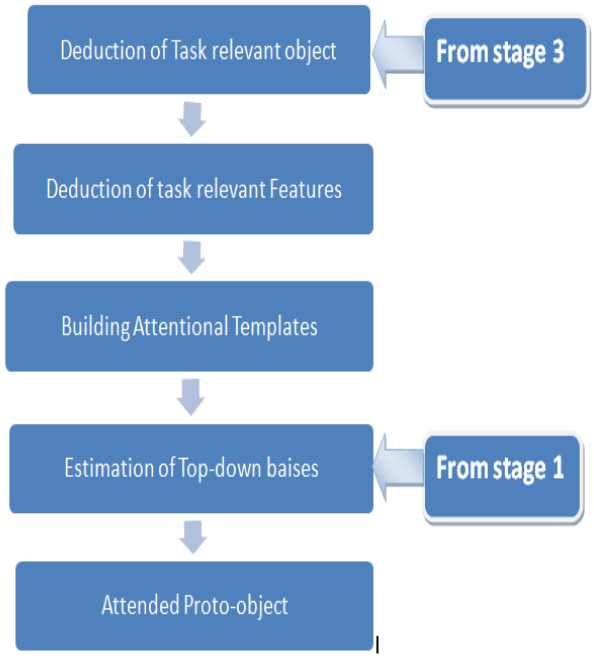

Figure 2 Top-down attentional selection

The ***top–down attentional selection*** stage main job is to autonomously allocate the attention through the top–down way. This stage is implemented by extending the Object Visualization Attention (OVA) model [6] based on the integrated competition(IC) hypothesis. It consists of four steps.

The ***post-attentive perception stage*** makes easily understand the data with attended object in more detail. The detailed interpretation aims to produce an appropriate action and learn the corresponding LTM object representation at the current time as well as to guide the top–down attention at the next time.

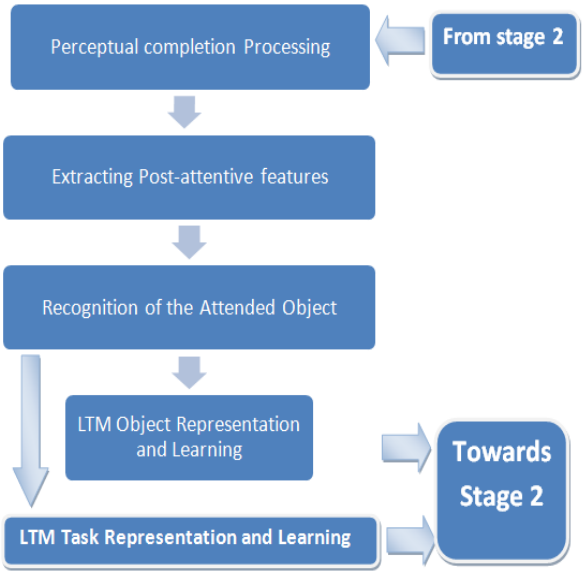

Figure 3 Post attentive perception

In this stage we have four steps are present. The first step is perceptual completion processing. This step it works around the attended proto-object, to achieve the complete object region. It consists of two parts. The second step is the extraction of post-attentive features that are a type of high-level representation of the attended object in Working Memory and used for the following two steps. The third one is object recognition. The fourth and the last are learning of Long Term Memory object representations.

IV. TERMINOLOGY USED

A. *Perception*

For the cognitive science we can view perception as a process of not only passively receiving the sensory input (i.e., a bottom–up process), but also being guided by learning, memory and expectation (i.e., a top–down process). The perception system with a goal directed approach main focus is on the top–down effects, i.e., how to organize and interpret the sensory input through the top–down attention and post-attentive perception. Thus the system is has a relation to the concept of perception in cognitive science.

B. *Top–Down Attention and Bottom–Up Attention*

In system, the integrated competition hypothesis mainly forms the base of the top–down attentional selection stage which is set up computationally. The salience descriptors in the Long Term Memory object representation are determined by using Itti's bottom–up attention model [8] which itself is derived from the Feature Integration Theory. Thus, they can be connected to the corresponding cognitive concepts of top–down and bottom–up attention.

C. WM and LTM

World memory is denoted as WM and Long-Term Memory as LTM and whenever an attended object is presented, it remains in WM over a brief period of time for deciding post-attentive features and leaning the associated LTM representation in this system. This is similar to the dual-store memory mode which we assume in cognitive science, i.e., each time an attended object is present in WM; its strength in LTM is also increased correspondingly.

V. CONCLUSION

This paper reviews the goal-directed visual perception system using the object-based top–down visual attention mechanism. The perception system reviewed has better performance than other methods in the following two aspects. The first one is that the system is robust to noise. This is because the system is object-based. The second one is the effectiveness of the target detection. The top-down attentional selection stage and post-attentive perception stage work together to achieve this performance.

There are three advantages in the system reviewed: First one is the LTM object representation is learned over the entire object. Second is that the LTM object representation includes the salience (i.e., task-relevance) and appearance descriptors, based on both of which the top–down attention can be guided by the conspicuous feature autonomously selected so as to cope with the case that the target and distracters share other features. And third one is the LTM object representation integrates the low-level and high-level representations together, both of which support each other so as to improve the effectiveness and efficiency of both top–down attentional selection and post-attentive recognition.

.

REFERENCES

[1] A. M. Treisman and G. Gelade, “A feature integration theory of attention,” Cogn. Psychol., vol. 12, no. 1–2, pp. 507–545, 1980. 44
[2] J. M. Wolfe, “Guided search 2.0: A revised model of visual search,” Psychonomic Bulletin Rev., vol. 1, no. 2, pp. 202–238, 1994. 48
[3] R. Desimone and J. Duncan, “Neural mechanisms of selective visual attention,” Annu. Rev. Neurosci., vol. 18, pp. 193–222, 1995. 14
[4] J. Duncan, “Converging levels of analysis in the cognitive neuroscience of visual attention,” Philosoph. Trans. Roy. Soc. Lond B: Biol. Sci., vol. 353, no. 1373, pp. 1307–1317, 1998.
[5] J. Duncan, G. Humphreys, and R. Ward, “Competitive brain activity in visual attention,” Current Opinion Neurobiol., vol. 7, no. 2, pp. 255–261, 1997.
[6] Y. Yu, G. K. I. Mann, and R. G. Gosine, “An object-based visual attention model for robotic applications,” IEEE Trans. Syst., Man, Cybernet, Part B: Cybernet, vol. 40, no. 5, pp. 1398–1412, 2010.
[7] M. Z. Aziz, B.Mertsching, M. S. E.-N. Shafik, and R. Stemmer, “Evaluation of visual attention models for robots,” in Proc. 4th IEEE Conf. Comput. Vis. Syst., New York, 2006, pp. 20–20.
[8] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 11, pp. 1254–1259, 1998.
[9] L. Itti and P. Baldi, “Bayesian surprise attracts human attention,” Vis. Res., vol. 49, no. 10, pp. 1295–1306, 2009.
[10] V. Navalpakkam and L. Itti, “Modeling the influence of task on attention,” Vis. Res., vol. 45, no. 2, pp. 205–231, 2005.
[11] S. Frintrop, “Vocus: A visual attention system for object detection and goal-directed search,” Ph.D. dissertation, Univ. Bonn, Bonn, Germany, 2005.
[12] J. Weng, “On developmental mental architectures,” Neurocomputing, vol. 70, pp. 2303–2323, 2007.
[13] W. Maier and E. Steinbach, “A probabilistic appearance representation and its application to surprise detection in cognitive robots,” IEEE Trans. Autonom. Mental Develop., vol. 2, no. 4, pp. 267–281, Dec. 2010.
[14] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, and F. Nuflo, “Modelling visual attention via selective tuning,” Artif. Intell.,vol. 78, pp. 282–299, 1995.
[15] A. Belardinelli and F. Pirri, “A biologically plausible robot attention model, based on space and time,” Cogn. Process., vol. 7, no. Supplement 5, pp. 11–14, 2006.
[16] A. Belardinelli, F. Pirri, andA. Carbone, “Robot task-driven attention,” in Proc. Int. Symp. Practical Cogn. Agents Robot., Perth, Australia, 2006, pp. 117–128.
[17] S. Chikkerur, T. Serre, C. Tan, and T. Poggio, “What and where: A bayesian inference theory of attention,” Vis. Res., vol. 50, no. 22, pp. 2233–2247, 2010.
[18] Y. Sun and R. Fisher, “Object-based visual attention for computer vision,” Artif. Intell., vol. 146, no. 1, pp. 77–123, 2003.
[19] Y. Sun, “A computer vision model for visual-object-based attention and eye movements,” Comput. Vis. Image Understand., vol. 112, no. 2, pp. 126–142, 2008.
[20] F. Orabona, G. Metta, and G. Sandini, “Object-based visual attention: A model for a behaving robot,” in Proc. IEEE Comput.Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), San Diego, CA, 2005, pp. 89–89.
[21] G. Backer, B. Mertsching, and M. Bollmann, “Data-and model-driven gaze control for an active-vision system,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 12, pp. 1415–1429, 2001.
[22] A. Bhattacharyya, “On a measure of divergence between two statistical populations defined by their probability distributions,” Bulletin Calcutta Math. Soc., vol. 35, pp. 99–109, 1943.