International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 11 (2015) pp. 30153-30166
© Research India Publications
http://www.ripublication.com

Bottom Up Approach For Modelling Visual Attention Using Saliency Map In Machine Vision A Computational Cognititve Neuroscience Approach

Manjunath R Kounte
PhD Scholar, JAIN University
School of Electronics and Communication Engineering
Bengaluru, India, manjunath.kounte@gmail.com
Dr B.K. Sujatha
Professor, Department of Telecommunication Engineering
M.S. Ramaiah Institute of Technology and Management
Bengaluru, India, bksujatha@msrit.edu

Abstract

The Modelling of Visual Attention using Saliency Map based technique and Bottom-up approach is presented in this paper. In Recent years, the detection of visual attention regions (VAR) is becoming more noteworthy due to its valuable applications in the area of multimedia. In this paper, we provide the Saliency Map hypothesis and results for identification of Visual Attention Regions in Machine Vision using Computational Cognitive Neuroscience. We also present how Computational Cognitive Neuroscience approach is the best approach in order for the implementation of Saliency Map technique for identification of Visual Attention Regions (VAR) in Machine Vision in real time scenarios where the systems are evolved for exhibiting intelligence in real time scenarios. And also, we focus on learning how visual attention is beneficial to the machine vision community and explain why attention is considered a selective process by highlighting how in the last 25 years research on attention has characterized into multiple ways.

Index Terms: Computational Cognitive Neuroscience, Saliency Map, Visual Attention, Visual Attention Region.

Introduction

Despite the perception that we “see everything around us”, there is significant gap between the amount of visual information that is received at the retinae, and the section of this visual information that reaches later processing or influences on conscious cognizance. Attention is crucial in determining visual understanding. The spirit of attention is perhaps best captured by William James [1]:

“Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others”.

One of the most severe problems of perception is information overload. Peripheral sensors generate afferent signals more or less continuously and it would be computationally costly to process all this incoming information all the time. Thus, it is important for the nervous system to make decisions on which part of the available information is to be selected for further, more detailed processing, and which parts are to be discarded. Furthermore, the selected stimuli need to be prioritized, with the most relevant being processed first and the less important ones later, thus leading to a sequential treatment of different parts of the visual scene. This selection and ordering process is called selective attention. Among many other functions, attention to a stimulus has been considered necessary for it to be perceived consciously.

What determines which stimuli are selected by the attentional process and which will be discarded? Many interacting factors contribute to this decision. It has proven useful to distinguish between bottom-up and top-down factors. The former are all those that depend only on the instantaneous sensory input, without taking into account the internal state of the organism. Top-down control, on the other hand, does take into account the internal state, such as goals the organisms has at this time, personal history and experiences, etc. A dramatic example of a stimulus that attracts attention using bottom-up mechanisms is a fire-cracker going off suddenly while an example of top-down attention is the focusing onto difficult-to-find food items by an animal that is hungry, ignoring more "salient" stimuli.

Attention provides a mechanism for selection of particular aspects of a scene for consequent processing while disregarding interference from competing visual events. A common misunderstanding is that attention and ocular fixation are one and the same phenomenon. Attention focuses processing on a selected region of the visual field that needn't coincide with the center of fixation.

Visual attention is an appliance of the human visual system to highlight on certain portions of a scene first, before attention is fatigued on to the other parts. Such areas that capture primary attention are called visual attention regions (VARs). For various multimedia applications, the Visual Attention Region (VAR) should then indeed be the regions of interest and an automatic practice of extracting the VARs becomes necessary. Documentation of VARs has been presented to be useful for object recognition and region based image retrieval. Similarly, images can be adapted for different users with different device proficiencies based on the VARs extracted from the image, thus improving viewing choice. Examples of such version comprise involuntary surfing for large images, image resolution adaptation and automatic thumbnail generation [2, 3, and 4].

Section II has a detailed assessment of Visual attention modelling using various conceivable methods which include Psychophysics, Computational Methods and Neurophysiology. We will emphasize on computational modelling which can be further classified into filter models, neural models and computational cognitive neuroscience models.

Section III gives information on extracting early visual features and visual processing from the retina.

Section IV describes Saliency Map detection and hypothesis for extraction of saliency map and identifying the visual attention region followed by Conclusion and future work in last section

Finally Section V gives the results and discussion on the saliency map based bottom up modeling of visual attention.

Modelling Visual Attention

A. Taxonomy

The advantages of Modelling Visual Attention includes in the field of Human-robot interaction, Robotics like Active vision, Robot Navigation, Robot Localization, Synthetic vision for simulated actors. In various other fields like Advertising, Finding tumor's in mammograms, Retinal prostheses etc.

In the field of Computer Vision and Graphics like Image segmentation, Image re-targeting, Image matching, Image rendering, Image and video compression, Image thumb nailing, Image quality assessment, Image super-resolution, Image super resolution, Video summarization, Scene classification, Object detection, Salient object detection, Object recognition, Visual tracking, Dynamic lighting, Video shot detection, Interest point detection, Automatic collage creation Face segmentation and tracking [5].

In human beings, the attention is facilitated by a retina that has evolved a high-resolution central fovea and a low resolution periphery. The important parts of scene gathering and collection of important information is guided by the anatomical structure of the retina. We focus on the Computational Modelling of this interesting field.

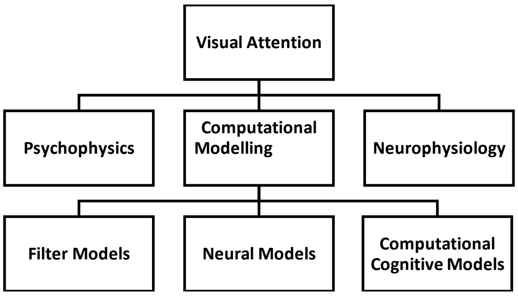

Figure 1: Taxonomy of Visual Attention Models.

As Shown in the figure 1, there are various approaches to the modelling of the visual attention. Research is being carried out by scientists from numerous domain of science. Foremost among them are the Psychologists who have studied behavioral correlates of visual attention such as change blindness [6] [7], inattentional blindness

[8], and attentional blink. Whereas Neurophysiologists have shown that how the neurons accommodate themselves to better represent objects of interest.

In Computational Modelling approach, the filter models have built models that can compute saliency maps and realize the Visual attention based on Top-Down Attentional Models and Bottom-Up Models. In neural network models, the approach is to simulate and explain attentional behaviors. [14]

Although there are many models available now in the research areas mentioned above, here we propose a new model based on computational Cognitive Neuroscience to build systems capable of working in real-time.

Extracting Early Visual Features

A. Visual Processing

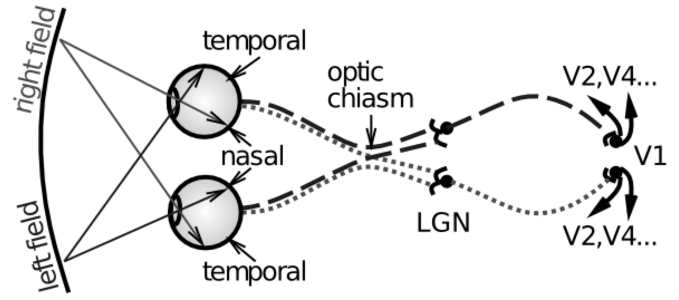

Figure 2: Visual Processing from the retina through lateral geniculate nucleus of the thalamus to primary visual cortex.

Figure 2, shows the graphical representation of basic optics and transmission corridors of input visual signals, which entre through the retina, and headway to the lateral geniculate nucleus of the thalamus (LGN), and further to primary visual cortex (V1).[9] The primary organizing principles at work here, and in other perceptual modalities and perceptual areas more generally, are: i) Transduction of different information -- in the retina, photoreceptors are sensitive to different wavelengths of light (red = long wavelengths, green = medium wavelengths, and blue = short wavelengths), giving us color vision, but the retinal signals also differ in their spatial frequency (how coarse or fine of a feature they detect -- photoreceptors in the central fovea region can have high spatial frequency = fine resolution, while those in the periphery are lower resolution), and in their temporal response (fast vs. slow responding, including differential sensitivity to motion).

The brain regions that participate in the deployment ofvisual attention include most ofthe early visual processing area.Visual information enters the primary visual cortex via the lateral geniculate nucleus[13], although smaller pathways,for example,to the superior colliculus (SC), also exist.From there,visual information progresses along two parallel hierarchical streams.Cortical areas along the ‘dorsal stream’ (including the posterior parietal cortex;PPC) are primarily concerned with spatial localization and directing attention and gaze towards objects of interest in the scene. The control of attentional deployment is consequently believed to mostly take place in the dorsal stream. Cortical areas along the 'ventral stream' (including the in ferotemporal cortex; IT) are mainly concerned with the recognition and identification of visual stimuli. Although probably not directly concerned with the control of attention, these ventral stream areas have indeed been shown to receive attentional feedback modulation, and are involved in the representation of attended locations and objects (that is, in what passes through the attentional bottleneck).In addition, several higher-function areas are thought to contribute to attentional guidance, in that lesions in those areas can cause a condition of 'neglect'in which patients seem unaware of parts of their visual environment[10]. From a computational view point, the dorsal and ventral streams must interact, as scene understanding involves both recognition and spatial deployment ofattention. One region where such interaction has been extensively studied is the prefrontal cortex (PFC).Areas within the PFC are bidirectionally connected to both the PPC and the IT [11].So,in addition to being responsible for planning action (such as the execution ofeye movements through the SC), the PFC also has an important role in modulating, via feedback, the dorsal and ventral processing streams.

Hypothesis For Saliency Map Detection

A. Taxonomy of Saliency Detection Methods

Saliency estimation methods can broadly be classified as into three categories as shown in fig 2, namely General Computational modelling, Hybrid Modelling and Computational Cognitive Neuroscience Modelling.

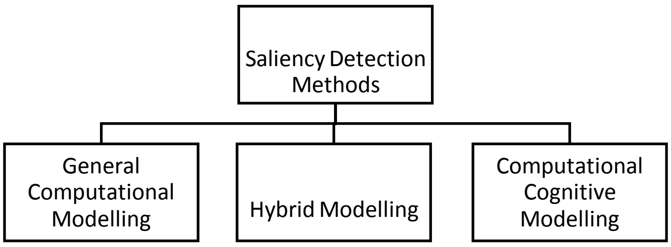
Saliency Detection Methods
General Computational Modelling
Hybrid Modelling
Computational Cognitive Modelling

Figure 2: Taxonomy of Saliency Detection Methods

The goal of any method, in general, is to detect properties of contrast, rarity, or unpredictability in images, of a central region with its surroundings, either locally or globally, using one or more low-level features like color, intensity, and orientation. General computational methods mainly rely on principles of, spectral domain processing, information theory or signal processing for saliency detection. In Hybrid individual feature maps are created separately and then combined to obtain the final saliency map, while in some others, a combined-feature saliency map is computed. Computational Cognitive Neuroscience model based methods attempt to mimic known models of the visual system for detecting saliency. Some of these algorithms detect saliency over manifold scales, while others work on a single scale.

B. Itti and Koch Model

The model proposed by Itti and Koch[12, 16] is shown in Fig.3. The Model is inspired from the "feature integration theory," which explains human visual search strategies. First, Visual input is decomposed into a set of topographic feature maps. All feature maps which then in a bottom-up manner, combine into a "saliency map". This model consequently represents a complete interpretation of bottom-up saliency and does not require any top-down mechanism to change the attention. This framework provides an immensely parallel method for the fast selection of a small number of attention-grabbing image locations to be explored by object-recognition processes.

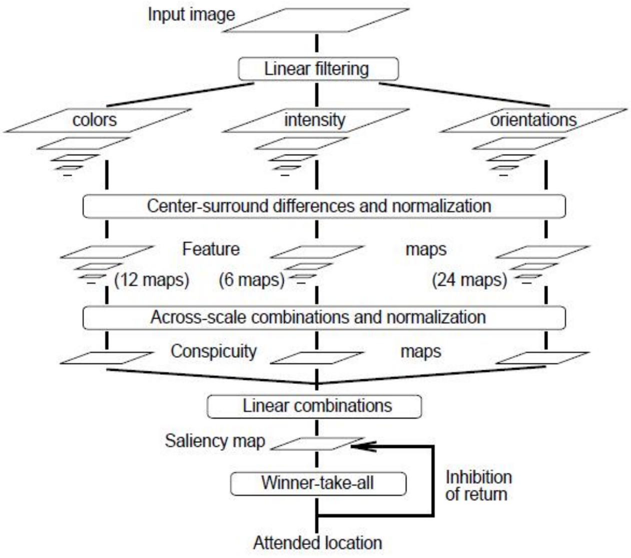

Figure 3: Koch and Itti Model

A. Framework for Saliency Map Extraction

C.1 Visual Preprocessing

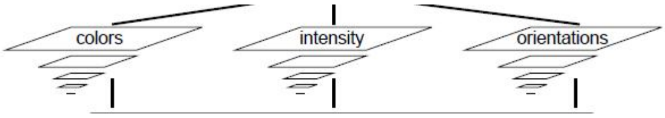

Figure 4: Visual Preprocessing

The fig 4 explains the early visual features are extracted from the visual input image into several separate parallel channels. After this extraction and an individual treatment, a feature map is achieved for each channel[13].

The first feature map, Intensity Channel is obtained from the red, green and blue channels of the input image. If R, G and B are assumed to be the red, green and blue channels respectively, then the intensity I, is given by taking average of all three channels.

$$I = (R + G + B) /3 \tag{1}$$

The Second feature map, Color Channel is constructed by taking into consideration pyramids as shown below

$$R = r - (g+b)/2 \tag{2}$$

$$G = g-(r+b)/2 \tag{3}$$

$$B = b-(r+g)/2 \tag{4}$$

The third feature map, Local orientation Channels O (σ) are computed by performing convolution of the levels of the intensity pyramid[19] with Gabor filters:$G(\sigma,\theta)$, where $\sigma \in [0 \ldots .8]$ represents the scale and $\theta \in \{0°, 45°, 90°, 135°\}$ is the preferred orientation. The Orientation feature maps, $O(c,s,\theta)$, encode, as a group, with local orientation contrast between the center and surround scales. Later the Center-surround receptive fields are simulated by a cross-scale subtraction among two maps at the center and the surround levels in these pyramids, yielding the third “feature map”. Later saliency map is extracted from the feature maps.

C.2 Center-Surround Differences

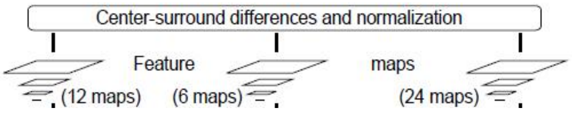

Figure 5: Center-Surround Differences

Each feature is computed by a set of linear “center-surround” operations. The concept is that typically visual neurons are most sensitive in a small region of the visual space (the center), while stimuli presented in a broader, weaker antagonistic region concentric with the center (the surround) inhibit the neuronal response. The operations is aimed at detecting locations which locally stand out from their surround.

Center-surround is implemented in the model as the difference between fine (center) and coarse (surround) scales: The center is a pixel at scale $c \in \{2,3,4\}$, and the surround is the corresponding pixel at scale $s = c + \delta$, with $\delta \in \{3,4\}$. Across-scale difference between two maps, denoted “⊖”below, is obtained by interpolation to the finer scale and point-by-point subtraction.

$$I(c,s) = |I(c) \ominus I(s)| \tag{5}$$

The second set of maps is similarly constructed for the color channels, which in cortex are represented using a so-called “color double-opponent”system: In the center of their receptive field, neurons are excited by one color (e.g., red) and inhibited by another (e.g., green), while the converse is true in the surround. Such spatial and chromatic opponency exists for the red/green, green/red, blue/yellow and yellow/blue color pairs in human primary visual cortex. Accordingly, maps *RG(c,s)* (for red/green, green/red double opponency) and *BY(c,s)* (for blue/yellow, yellow/blue double oppenency) are created as Eq.6 and Eq.7.

$$RG(c,s) = \left|\big(R(c) - G(C)\big) \ominus (G(s) - R(s))\right| \tag{6}$$

$$BY(c,s) = \left|\big(B(c) - Y(C)\big) \ominus (Y(s) - B(s))\right| \tag{7}$$

Local orientation is obtained from I using oriented Gabor pyramids $O(\sigma,\theta)$, where $\sigma \in [0..8]$ represents the scale and $\theta \in \{0°, 45°, 90°, 135°\}$ is the preferred orientation. Orientation feature maps, $O(c,s,\theta)$, encode, as a group, local orientation contrast between the center and surround scales:

$$O(c,s,\theta) = |O(c,\theta) \ominus O(s,\theta)| \tag{8}$$

In total, 42 feature maps are computed:
 for intensity, 12 for color and 24 for orientation.

C.3 Normalization

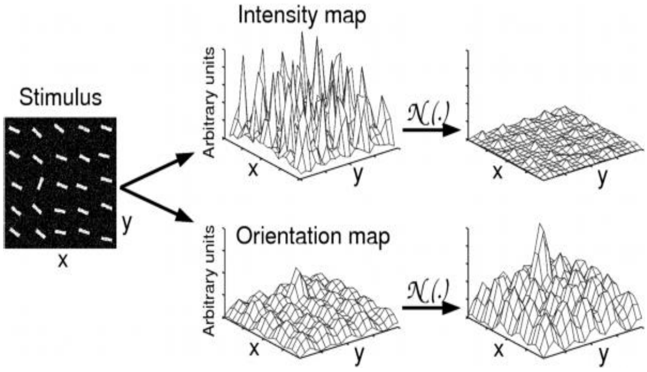

Figure 6: Normalization

The operator is denoted as N(.). Normalize the values in the map to a fixed range [0..M] in order to eliminate modality-dependent amplitude differences. Find the location of the map's global maximum M and compute the average $\overline{m}$ of all its other local maxima; and globally multiplying the map by $(M - \overline{m})^2$.

C.4 Conspicuity Maps

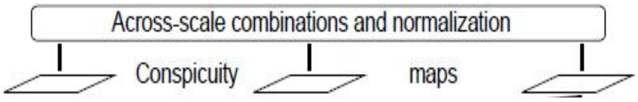

Figure 7: Conspicuity Maps

The feature maps are combined into three conspicuity maps at the scale 4 ($\sigma = 4$). $\bar{\mathrm{L}}$for intensity, $\bar{\mathrm{C}}$ for color , and $\bar{\mathrm{O}}$ for orientation .This is obtained through across-scale addition which consists of reducing of each map to scale 4 and point-by-point addition.

C.5 Saliency Map

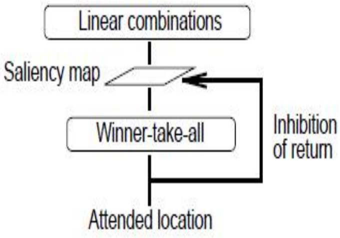

Figure 8: Saliency Map

The three conspicuity maps are normalized and summed into the final input S to the saliency map.

$$S = \frac{1}{3}\left(\mathcal{N}(\overline{I}) + \mathcal{N}(\overline{C}) + \mathcal{N}(\overline{O})\right) \tag{8}$$

Results and Discussion

The MATLAB implementation of this method can be downloaded from http://www.saliencytoolbox.net.[15].

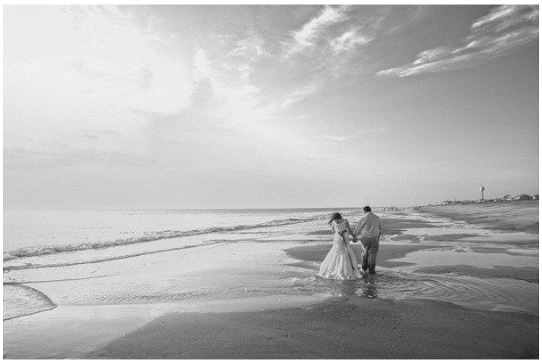

Figure 9: Input Image

Fig 9, Shows the input image in png format, which is scaled down to 600x400 size with bit depth of 32.

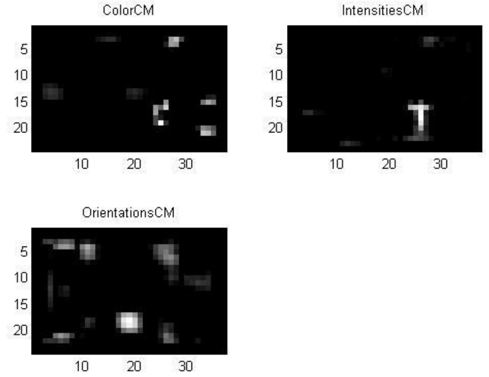

Figure 10: Color, Intensities and Orientations

Fig 10, the identification of the feature maps namely Intensity Channel, Color Channel and Orientations channel

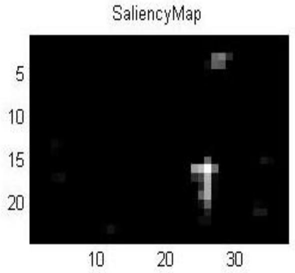

Figure 11: Saliency Map

Fig 11, shows the identification of the saliency map based on the feature maps extracted from the conspicuity maps developed from separate feature map channels like intensity, color and orientations. The computational cognitive neuroscience approach is significant because the saliency map highlights the most important parts of the input image with respect to focus of human attention. As we see in fig 6, the input image consists of two people walking in a beach, with a blue sky, land and sea making the majority of the image. As we normally see that human focus will be on the two people first followed by other things later. The saliency map after feature map extraction does exactly that and that’s why we call them as identification of the visual attention region (VAR).

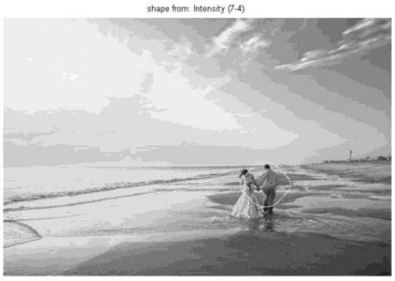

Figure 12.a:

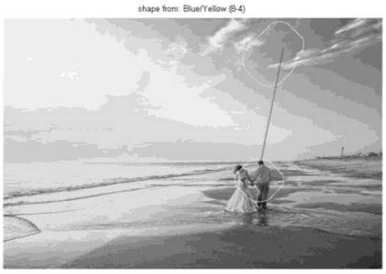

Figure 12.b:

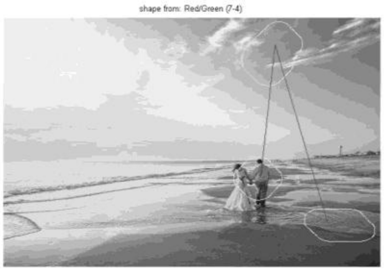

Figure 12.c:

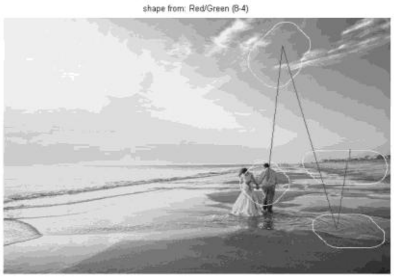

Figure 12.d:

Figure 12: Bottom-up Approach for Identifying Visual Attention Regions using Saliency Map, a) Identification of Most important region using WTA , b) ,c) and d) Identifying next important regions without going back to the Previously identified regions.

Figure 12 shows the complete results of the Bottom-up Approach for Identifying Visual Attention Regions using Saliency Map. In fig 12 a) the most important region i.e highest probability of where the human visual attention can focus is identified. Fig 12b, 12c and 12d highlights the implementation of Inhibition of return for the saliency map based visual attention regions, where the previously identified regions are skipped and next most important regions are identified.

Conclusion and Future Work

In this paper, we have given a hypothesis for Computational Cognitive Neuroscience based Saliency Map identification and highlighting of Visual Attention Region's (VAR's) in Machine vision using saliency map based bottom-up approach. Also with done a Critical review of Visual Attention Modelling by different streams arising because of Computational Methods like Filter Models & Neural Models, Psychophysics, and Neurophysiology.The cognitive neuroscience computing is at a verge to give upsurge for many potential conventional applications. We have presented the results for implementation of the saliency map based bottom-up modelling of visual attention.

The future work includes the justification for the Saliency Map Hypothesis for Modelling of Visual Attention using Computational Cognitive Neuroscience using Mathematical modelling like Bayesian Approach and Perception Based approach.

Acknowledgment

We would like to thank JAIN University, REVA University and M S Ramaiah Institute of Technology for providing the necessary support and infrastructure to carry our research work.

References

[1] W. James. The Principles of Psychology. Holt, New York, 1890.

[2] U. Rutishauser, D. Walther, C. Koch, and P. Perona,"Is bottom-up attention useful for object recognition?", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 37–44, Washington, DC, USA, July 2004.

[3] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "On the usefulness of attention for object recognition" *In 2nd International Workshop on Attention and Performance in Computational Vision 2004 Prague*, Czech Republic, May 2004,pp 96–103.

[4] D. Walther, U. Rutishauser, C. Koch, and P. Perona,"Selective visual attention enables learning and recognition of multiple objects in cluttered scenes",*Computer Vision and Image Understanding*,2005, pp 745–770.

[5] Ali Borji, Laurent Itti,," State-of-the-art in Visual Attention Modelling" *IEEE Transactions On Pattern Analysis And Machine Intelligence,* 2010.

[6] D.J. Simons and D.T. Levin, "Failure to Detect Changes to Attended Objects,“ *Investigative Ophthalmology & Visual Science,* vol. 38, no. 4, pp. 3273, 1997.

[7] R. A. Rensink, "How Much of a Scene is Seen - the Role of Attention in Scene Perception*", Investigative Ophthalmology & Visual Science*, vol. 38, 1997.

[8] D.J. Simons and C.F. Chabris, "Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events,"*Journal on Perception,* vol. 28, no. 9, pp. 1059-1074, 1999

[9] O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., and Contributors, " Computational Cognitive Neuroscience". Wiki Book, 1st Edition. URL: http://ccnbook.colorado.edu, 2012

[10] Webster, M. J. & Ungerleider, "L. G. in The Attentive Brain" (ed. Parasuraman, R.) MIT, Cambridge, Massachusetts, 19-34, 1998

[11] Miller, E. K., "The prefrontal cortex and cognitive control", Nature Rev. Neurosci. 1, 59–65, 2000.

[12] Laurent Itti, Christof Koch and Ernst Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", *IEEE transactions on pattern Analysis and Machine Intelligence*, Vol. 20, No.11, November 1998

[13] L. Itti, C. Koch, Computational Modelling of Visual Attention, Nature Reviews Neuroscience, Vol. 2, No. 3, pp. 194-203, March 2001.

[14] Manjunath R Kounte, Dr. B K Sujatha, "A Review of Modelling Visual Attention using Computational Cognitive Neuroscience for Machine Vision " , *International Journal of Advanced Research in Computer and Communication Engineering*,Vol. 2, Issue 9, pp 3558-3563,September 2013.

[15] Dirk Walther and Christof Koch, "Modeling attention to salient protoobjects",*Neural Networks 19*, 1395-1407,2006

[16] Laurent Itti, Christof Koch and Ernst Niebur, “A Model of Saliency-Based Visual Attention for Rapid Scene Analysis”, *IEEE transactions on pattern Analysis and Machine Intelligence*, Vol. 20, No.11, November 1998

[17] Mohammed Riyaz Ahmed, Dr. B K Sujatha, “A Review of Reinforcement Learning in Neuromorphic VLSI Chips using Computational Cognitive Neuroscience”,*International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 2, Issue 8, pp 3315-3320,August 2013.