ViC Model: An Executive Controller of Emotions Based Visual Selective Attention Mechanism for Cognitive Agents

Shehzad Latif [1]

School of Systems & Technology,
University of Management & Technology, Lahore.
Alishba Latif[2]
Faculty of Computer Sciences,
Hajvery University, Lahore.

Abstract
Visual attention is the most significant mechanisms of perception that empowers human to efficiently select the visual data of most prominent interest. Machines confront comparable difficulties as human & need to manage a lot of visual information. In the domain of computer vision, efforts have been made in modeling the mechanism of human like attention, especially the bottom-up attention mechanism. The use of attentive methods has pervaded the computer vision literature demonstrating the importance for reducing the amount of information to be processed. Reducing information in a real time is an extremely daunting task without the assistance of smart systems. Many visual attention models are designed for this purpose. Almost all visual attention models are straightforwardly or by implication motivated from the cognitive architecture. Emotions are the most significant component of cognitive architecture which was ignored in these models. Emotionality is associated with a range of psychological phenomena including temperament, personality, mood, and motivation. Emotions can act as an incredibly powerful force to reduce the visual information. So, this research proposes a ViC Model which will act as an Executive Control Mechanism of Emotions Based Visual Selective Attention for Cognitive Agents. This may allow the underlying agent to act more humanly due to its subjective nature of selection of context and attention of the fore-mentioned percepts.

Keywords:
Bottom-up attention, Top-down attention, Cognitive Architecture, Long Term Memory, Working Memory and Perception

1. Introduction

The world is huge, energized and beautiful with the motivation for human learning and information gathering in result of interaction with the environment [1]. During the process of interaction, human views environmental conditions, learn persistently and produce random reactions. Human communications have multi-modular elements e.g. motions, feelings, actions and many other modules to communicate in the environment. Also, human cooperates and communicates inside conditions on the assumption of attention to learn and acquire information [2] and perform information exchanging [3]. For this reason, human uses vision & natural language [4].Attention plays a key role in the selection of information & context generation. Without attention it is quite difficult for human to manage a lot of visual information every moment. They need to focus on specific visual information as per their interest or requirements. Similarly, machines can learn and obtain information when human nature associates with them. The design of intelligent machines also ought to be viewed as multi-modal. Some multi-modal features e.g. speech are considered in an interactive machines but still many aspects of human's vision (vision based interaction) are neglected [5]. Man and machine both ought to know about each other and cognitive machine ought to perceive each other through vision, communication, actions and motions [6]. So they can cooperate by utilizing these various elements and can affiliate, convey and learn conditions [7]. Thus, with specific end goals to plan human like communications in machine engineering, normal dialect handling and vision selections are basics to interpret and perceive visual and audio information [8].

Visual attention is one of the key components of recognition that allows human to actively pick the visual information of most prominent region. Machines face comparable difficulties as humans. Machines have to deal with a lot of information and need to focus the most prominent regions. To focus the visual information, human have psychological qualities to play a key role in the execution of attention abilities. Similarly, in priority to computational capacity, machine must have the perfect psychological qualities to make progress in visual selection. Salazar et al. [9] suggested that the psychological state a few seconds before the execution of attention abilities is the key factor influencing the execution skills [9]. Furthermore, past researches also demonstrated that human show higher fixation during execution of visual skills [10] which keeps

Manuscript received November 5, 2025
Manuscript revised November 20, 2025
https://doi.org/10.22937/IJCSNS.2025.25.11.19

human from being occupied by other internal and external stimuli. In light of the previously mentioned assessment, psychological states assume to be a critical part of vision. However, the sources of good psychological states are additionally significant elements. As indicated by the investigation led by Vealey [11], sources influencing confidence include power, exhibition ability, physical/psychological preparation, physical self-presentation, social support, alternative experience, environmental comfort and beneficial context. Numerous researchers have examined the sources of self-confidence and explicitly showed the sources of confidence. However, sources of psychological states, for example, focus; inspiration, positive consideration, and so forth have not been explored in details. In addition, the psychological states of human in task performance are influenced by numerous extra inside and outer components, and there are many sources, instead of only one variable. As indicated by past researches, there are many sources of psychological states which are social support , positive self-talk , commitment , and anxiety management skill , and so forth [12][13].

"Emotions" are the sum of psychological changes, including psychological examination, sentiments, activity lack of caution, and express conduct, which are produced keeping in mind the end goal to manage stimuli [14]. Therefore, when human's emotions during task execution can't be quickly controlled and dealt with, their task execution will be effectively influenced. There were few sources of emotional administration including "mental relaxation" and "reaction to a challenge ordinarily". At the point when human react ordinarily to a challenge they are more opposed to encounter dynamic change. Hence, it we can conclude that the role of emotions in visual selection process cannot be ignored.

2. RELATEDWORK:

Visual attention is most significant mechanisms of perception that empowers us to efficiently select the visual data of most prominent interest. Machines confront comparable difficulties as human: they need to manage a lot of information and need to choose the most encouraging parts. The idea of selective attention alludes to the fact as describe by [Aristotle]: "It is hard to see two articles in the same sensory act". In spite we usually have the impression to hold a rich depiction of our visual world and that gigantic changes to our environment will pull in our attention.

The principle concerns in demonstrating attention are the manners by which, when, and why we select behaviorally important regions. A common approach is getting motivation from the life structures & usefulness of the human visual framework, which is profoundly developed to take care of mentioned issues. On the other hand, a few examinations have estimated what work visual attention may serve & have detailed it in a computational structure. For an instance, it has been assumed that visual attention is pulled in to the most encouraging, the most amazing scene locales, or those areas that maximize reward regarding a task.

2.1 Top-Down versus Bottom-Up Models

A noteworthy difference among models is whether they depend on stimulus driven approach or goal driven approach or combination of both. Bottom - up cues are fundamentally in view of attributes of a visual scene (stimulus-driven), while top - down cues (objective - driven) [15] are dict ated by cognitive concept like learning, desires, reward, and current objectives. Areas of interest that pull in attention in a bottom- up way should be adequately particular with respect to environment. This attention component is likewise called exogenous, programmed, reflexive, or incidentally signaled. Bottom up attention is quick, automatic, and no doubt nourish forward. A prototypical case of bottom-up attention is taking a gander at a scene with just a horizontal bar among several vertical bars where attention is instantly attracted to the horizontal bar While many models fall into this class, they can just clarify a little part of eye movements since the greater part of obsessions are driven by task [16].

Whereas top down attention is slow. A standout among the most acclaimed cases of top-down attention is from Yarbus in 1967 [17], which demonstrated that eye movement rely upon the current task with following experiment: Subjects were made a request to watch a similar scene (a live with a family and a surprising guest going into the room) under various conditions (questions, for example, "evaluate the physical conditions of the people," "people of what ages are there?", or freely observe the environmental conditions. Models have investigated three noteworthy wellsprings of top-down impacts in light of this inquiry: How might we pick where to look? A couple of models address visual hunt in which attention is drawn toward parts of an objects we are looking for. Some unique models investigate the piece of scene setting or hugeness to constrain areas that we investigate. It is difficult to say where or what we are looking. Design of scene has additionally been suggested as a source of goal driven attention and is considered here in combination with scene setting.

In computer vision domain, endeavors have been done in displaying various systems about human like attention; particularly the stimulus driven attention approach. The utilization of cognitive techniques has invaded the

computer vision exhibiting the significance for lessening the measurement of data to be handled.

2.2 Related Cognitive Models

All attention models are straightforwardly or by implication motivated from cognitive architecture.

2.2.1 Itti et al's. Model [18]

His model is based on Color, intensity, & orientation of the objects. This model has been the premise of future models & can be used as a standard for other models. An information picture is subsampled into a Gaussian pyramid and each pyramid level o is crumbled into separate channels for Red (R), Green (G), Blue (B), Yellow (Y), Intensity (I), and neighborhood presentations (O&). Using these separate channels, center-surround "feature maps" *fl* for variety of features *l* are proposed and normalized.

2.2.2 Le Meur et al Model [19]

Le Meur et al. presented a mechanism for bottom up approach based on (HVS) Human Vision System. Perceptual decomposition, center-surround interactions, Contrast sensitivity functions, & visual masking are the key features focused in the model proposed by Le Meur et al. Afterwards: He modified his model to Spatio-temporal domain by intertwining colorless, colorful, & temporal information. In Le Meur et al. Model, early visual parts are removed from the visual commitment to a couple of isolated parallel channels. A component guide is procured for independent channels, and after that recognized saliency guide worked from the blend of these channels. The significant novelty suggested here lies in the consideration of the temporal measurement in addition to coherent normalization approach.

2.2.3 Navalpakkam & Itti Model [20]

Navalpakkam & Itti demonstrated visual inquiry in terms of a top-down gain issue by exploring Signal-to-Noise Ratio (SNR) of the object versus distractors as opposed to learning unequivocal combination capacities. They learned straight weights for feature collection by boosting the ratio b/w target saliency & distractor saliency.

2.2.4 Kootstra et al. Model [21]

Kootstra et al. created three symmetry-saliency operators & compared them with the information gathered by human eye. This technique depends on isotropic symmetry & spiral symmetry operator by Reisfeld et al. He also used the shading symmetry of Heinemann. Kootstra modified these operators to multi scale symmetry-saliency models and demonstrated that their mechanism performs essentially well with symmetric stimuli as compared to the Itti et al. mechanism.

2.2.5 Marat et al Model [22]

Marat et al. created a Bottom-Up mechanism for Spatio-temporal saliency forecast in video stimuli. He concentrates on two signs from the videos relating parvo-cellular & magno-cellular cells of the retina. Using above signs, static & dynamic saliency maps are determined & melded into Spatio-temporal guide. Expectation consequences of his mechanism were good for the initial couple of frames of each clip snippet.

2.2.6 Murray et al. Model [23]

Murray et al. presented a model in light of a low-level vision framework in three stages:

1) Visual Stimuli are handled by early human visual pathway (shading rival and luminance channels, trailed by multi scale decay),

2) A stimulation of the restraint systems introduced in cells of the visual cortex standardize their reaction to stimulus differentiation, and

3) Information is incorporated at different scales by playing out a backwards wavelet change specifically on weights processed from the non-linearization of the cortical yields.

The greater part of the past research has centered on the Bottom Up segment of Visual Attention. But researchers today also accept this reality that Goal driven approach should be considered in Visual Selective Attention. Therefore, there is a need to design a mechanism that will serve as an executive control for Visual Selective Attention for Cognitive Agents using psychological states. This may allow the underlying agent to act more humanly due to its subjective nature of selection of context and attention of the fore-mentioned percepts.

3. Proposed Model

The design of the agent is inspired by the real time communication patterns among humans to reduce the gap in man-machine communication. In real world when humans interact with each other they are attentive towards each other and meanwhile they can also communicate with others. It means humans can interact with multiple persons

at the same time i.e. multiparty interaction. In our daily life, we can communicate with multiple persons in multiple scenarios at once. Similarly, we want to introduce multiparty interaction in our agent so that our agent can communicate like us.

Human beings have multiple senses e.g. hear, smell, vision etc. which helps in interaction within environment. In addition to senses attention also play a significant role in the process of interaction. Without attention it is difficult to focus & understand things clearly. Interactive agents need same senses along with attention for better interaction. Working Memory model for interactive agent perceives the environment on the behalf of external stimuli i.e. audio and visual through sensors.

For designing the interactive system that can perform human-like multiparty communication we need to design a cognitive architecture / model. So the system is designed by keeping into account the above mentioned criteria.

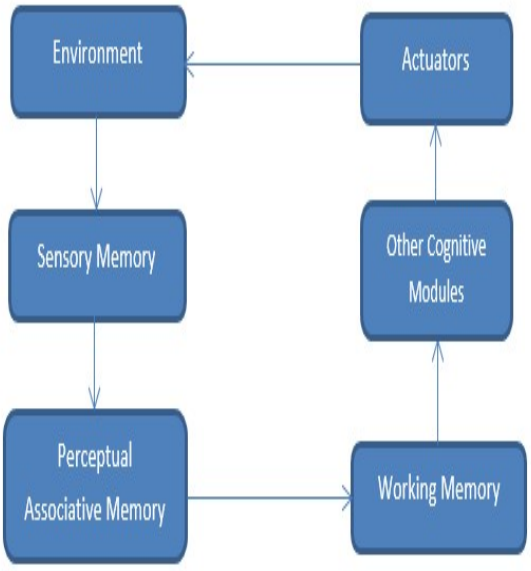

Fig 1 : Cognitive Architecture

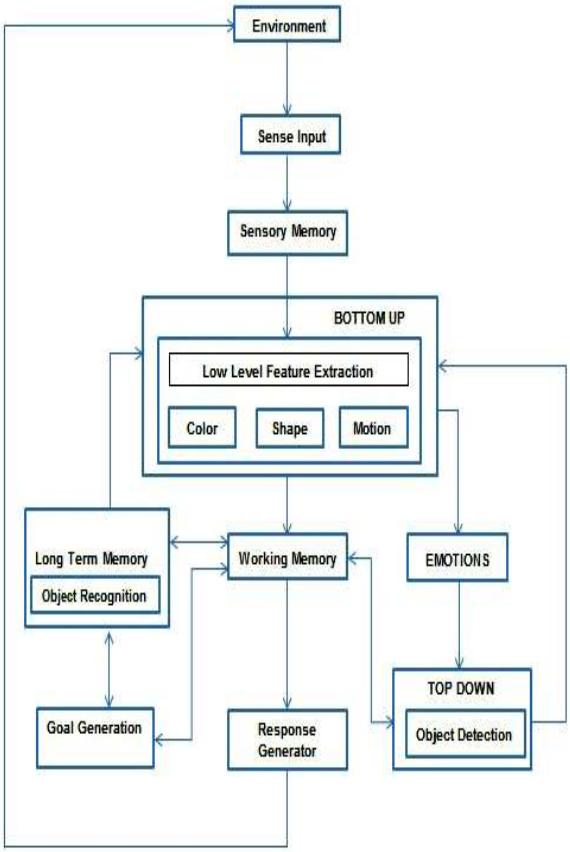

Fig 2. ViC Model of Visual Selective Attention

3.1 Environment

It is assumed that cognitive agent is placed in any interactive environment e.g. real time environment of a road side where multiple objects are present around. High definition camera is placed at the head top of the agent.

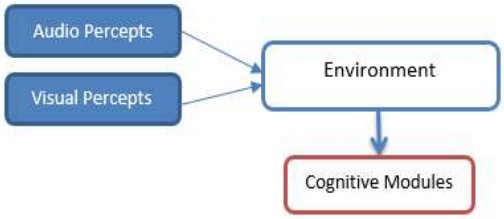

Fig 3: Environment

3.2 Sensory Input

The cognitive agent visualizes the environment using its camera and receives the Sensory Input which may be an image or clip.

3.3 Sensory Memory

The basic function of sensory memory is to transmit the visual information. The sensory signals from visual sensor come into the sensory memory. Sensory memory produces selective attention on the approaching signals. Sensed visual signal from the internal or external sensors are buffered in sensory memory for a shorter time period. The visual signals are stored in the visual sensory memory. Images produced by the camera (or other imaging frameworks) are buffered in visual sensory memory for further processing. Visual sensory memory is the subtype of sensory memory that captures the image frames continuously and holds them in memory. This phase of memory is temporarily reserved.

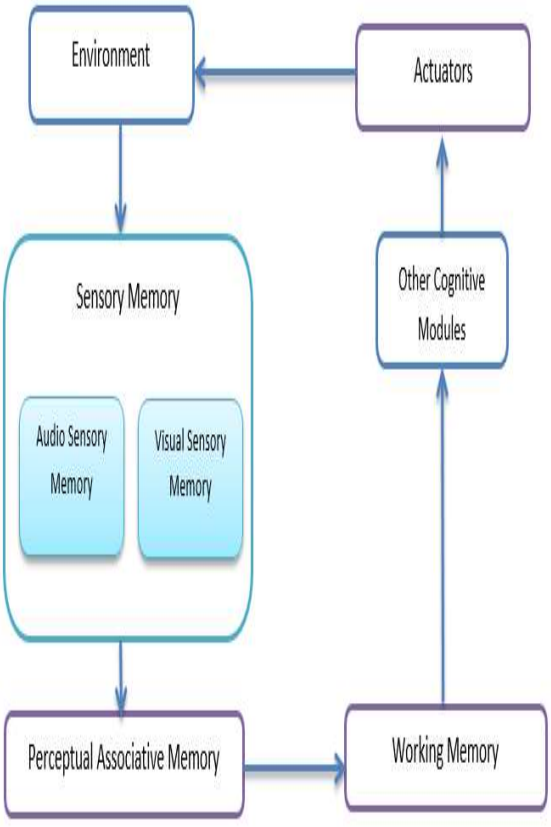

Fig 4: Sensory Memory

3.4 Low Level Feature Extraction

Perceptual memory detects low-level features from input percepts. It localizes the visual percepts and integrates them as a unified signal. Perception can be done in this module. Assigning meanings to the sensory data is done here. It plays a significant role in the interpretations of incoming stimuli.

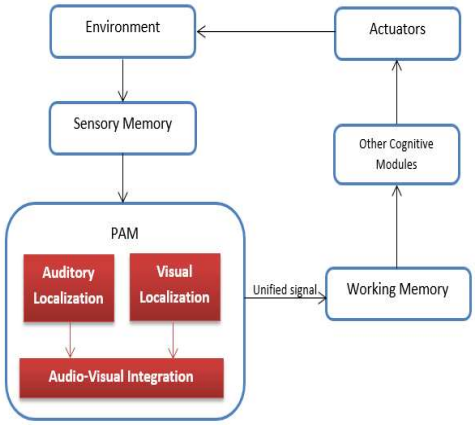

Fig.5: Perceptual Memory

3.5 Working Memory

Working memory is vital for thinking and the guidance of decision making and conduct. Working memory surpass this information coming from sensory memory and perceptual memory module to other modules of the model. Internally, working memory consists of several other modules that are as follows:

3.5.1 Executive Control:

It is the supervisor inside working memory and responsible for the integration of information between other modules of working memory. It controls the cognitive process and regulates information between them.

3.5.2 Episodic Buffer Module:

It the subsystem of the working memory which maintains the cues of the incoming perception. This module also analyzes the context behind the generated cues and interrelate them. It has two sub-modules that manages and analyzes the cues.

3.5.3 Episodic Manager

This sub-module is like a buffer that manages the incoming stimuli. It checks the cues and manages the related cues together.

3.5.4 Encoding

How the incoming stimuli initiate and what type of features in cues can be extracted. It determines the cues and depending upon the relevant feature this module stores it.

Storage

After the proper encoding of the cues within the buffer episodic manager stores them. The stored cues can be removed from storage module.

Retrieval

It is responsible for providing cues that are being used by other cognitive modules. This module defines which type of key is used to trigger the episode.

Context Analyzer

This module generates context and analyze them on the behalf of cues from the episodic manager. It formulates context and update the manager so that it could change the feature of selection of cues. For the analysis, this module uses the active context, knowledge repository or Long Term Memory (LTM) and set of goals. For example, while picking actions, an agent can "play forward" earlier scenes with comparable highlights and expectations, giving an agent a general and autonomous source of information assessment.

3.5.5 Visual Processing Module:

The module processes the visual part of the attentive signal using image processing strategies. It switches the vision processed information to executive control for task switching. Mid-Level image processing is done in this module of the visual processing module to extract the attributes of the objects. It applies high level image processing on the extracted objects and creates association.

3.5.6 Feature Detection

This module gets a part of the image along with the corresponding segmentation mask and returns the features, which are then used for recognition and/or learning.

3.5.7 Visual Concepts/Attributes

This module contains the learned attributes of the images that help the processing system that define the image after the detection.

3.6 Long Term Memory (LTM)

It is also known as Knowledge Source. LTM is used for the recognition of visual input on the behalf of cues. It helps in developing the context and understanding the meaning of the visual input for the purpose of recognition using past experiences & knowledge.

3.7 Response Generator

After detection & recognition of visual input response generator module is responsible for generating the suitable response to the visual input; which may be some action etc.

Color Based Detection of Objects

We utilize colors to show the key information. Colors can be a critical source of information in the detection & recognition of objects. Since colors are important features of objects, they can make this process simpler. In addition, color processing can significantly lessen the measure of false edge focuses. A camera creates a colored picture. This picture is not appropriate in most cases for the detection of road signs colors on the grounds that the RGB color space is worked as Cartesian framework organizes R, G, and B in the x, y, and z axis separately, & directions of the above three colors are profoundly related which results that any variety in the encompassing light power influences the RGB system. An important piece of color-based recognition framework is color space change, which implies changing over the RGB picture into another shape that simplifies the detection process. This implies isolating the color information from the brightness information by changing over the RGB color space into another color space, which gives great detection capacities relying upon the color cue.

Shape Based Detection of Objects

Regardless of the wide utilization of colors in the identification of objects, the last can also be detected utilizing shapes. Usefulness of shapes of road sign is supported by many endeavors. The need of standard colors among the countries is one of the focuses supporting the utilization of shape for object recognition. Another point in this contention is about the fact colors fluctuate as sunshine & reflectance feature change. The circumstances in which it is quiet hard to retrieve color information for example, evening and nightfall time, detecting and recognizing shapes is a smart choice. Utilizing shapes to identify objects has certain properties and it might confront a few challenges. Among the properties and troubles of shape-based objects detection are the following. Comparable objects may exist in the environment. Objects can be appeared damaged, occluded by other objects. Objects may seem irregular. As, the span of the objects fluctuates too. At the point where the road sign looks very small, it seems unrecognizable. Shapes are not affected by sunlight or shading variations. Working with shapes requires robust edge recognition and coordinating calculation. This is difficult when the objects appear moderately small in the picture. The result of the detection stage is a list of the candidate objects. This list is sent to the recognizer for advance assessment, and afterward to the classifier to choose whether the objects in the list are either dismissed items or objects. To design a

decent recognizer, numerous parameters should be considered. Firstly, low computational cost and decent discriminative power. Also, it ought to be well aware about the geometrical shapes of road signs. Thirdly, it is expected to be so powerful to overcome the noise. Fourthly, acknowledgment ought to be completed rapidly if planned for the development of real time applications. Besides, the classifier should have the capacity of taking wide number of classes and as much from the earlier learning about object should be utilized into the classifier's plan, as could be expected under the circumstances.

Impact of Colors on Emotions

Color Symbolism is by all accounts evident in how individual associate color with things, objects or physical space. A Color related feelings are majorly subject to individual likings and one's past involvement with the specific colors. We can relate colors with our feelings & emotions. The color blue is associated with the comfort & security [24], [25]. The red color has both negative and positive impressions, for example, active, strong & passionate; but on the other hand aggressive, angry & intense. The green color has retiring and relaxing effect [24], [25]. It is assumed that the light colors have positive whereas dull colors have negative effects on our emotions & feelings.

Impact of Shapes on Emotions

As Bar and Neta [26] proposed that the emotional valence of stimuli is initiated by their semantic importance and by initial level properties of the environment (e.g., shape, symmetry, differentiation, multifaceted nature, or perceptual familiarity [27]. These properties influence people's interests, judgments, behavioral reactions, and choices [28]. For instance looming movements convince human infants to reconsider [29]. Moreover, threat-relevant stimuli presented with a looming movement capture humans' attention [30].

Basic shapes additionally incite passionate responses by individuals. Researches reveal that straight lines and angular shapes (especially a downward V) are viewed as "terrible" and circles and curvilinear shapes are viewed as "good" [31]. This preference for curvilinear shapes seems to be present early during development, before dialect is obtained [32].

Impact of Psychological States on Visual Selective Attention

Psychological states play a significant role in the detection & recognition of objects. In priority to having, physical capacity, and great physical environment, cognitive machine must have the perfect psychological qualities to make progress in visual selection. Emotional States or Feelings are the most prominent psychological states, which influence our attention & have great impact on it. Emotional cues play an important role in the regulation of the attention spotlight. As Matthew and Wells (1999) explain, "Emotion and attention are significantly linked. States of emotion influence both the contents of consciousness and performance on tasks requiring selection of stimuli or intensive concentration". Happiness and sadness have been seen to modify attention in a number of ways. First, happiness and sadness lead to external and internal focuses of attention respectively. While people who are happy tend to focus their attention outward on other people and external stimuli, sad people are more likely to focus their attention inward on themselves and how they are feeling. Happiness generally indicates that everything is going well and there are no problems; sadness, on the other hand, indicates that something is wrong. Salazar et al. proposed that the psychological state a few seconds before the execution of attention abilities is the key factor impacting the execution skill.

Object Detection & Recognition using a ViC Model

An agent is placed in any environment having different objects. An agent is getting the visual input from its surroundings. An agent is searching for the "cell phone". Instantly an object is seen by the agent. Agent has to check whether this object is cell phone or not. Visual sensory memory captures the image frames of that object & its surroundings and holds them. Perceptual memory detects low-level features from input percepts of that object. It localizes the visual percepts and integrates them as a unified signal. Low level features extraction will be performed here based on the shape, color, motion & other physical features of the object. Episodic Buffer analyzes the cues and interrelate them. Context Analyzer generates context and analyze them on the behalf of cues from the episodic manager. It formulates context and update the manager so that it could change the feature of selection of cues. For the analysis it uses the active context, knowledge repository or Long Term Memory (LTM). Visual Processing Module processes the visual part of the attentive signal using image processing strategies to recognize the detected object. If these cues are not enough for the recognition of that object as a cell phone. It checks for more cues related to the context & extract more features for which high level image processing is performed. This process will repeat till the enough cues for object recognition are obtained.

4. Discussion

In earlier sections related models were discussed. Now it is time to compare the ViC Model to those models in order to check its usefulness. The proposed mechanism works efficiently & accurately as compare to its competitive systems. In our model both top down & bottom up feature extraction is performed; while previous models are based

on either bottom up approach or top down approach. Our model associates psychological states with visual information whereas previous models are developed without association of psychological states. The proposed model memorizes new concepts & relevant context generation can be performed whereas, other systems don't. Using our model agent can easily detect & localize the road signs which are then recognized after analyzing cues using the knowledge source. Agent can work efficiently in any interactive environment & can generate responses accordingly. Here is the comparison of different systems with the proposed system.

Table 1 Comparison of ViC Model with other Cognitive Models

Features	Itti et al's. Model	Le Meur et al Model	Naval pakkam & Itti Model	ViC Model Proposed Model
Bottom Up Feature Extraction	Yes	Yes	Yes	Yes
Top Down Feature Extraction	No	No	Yes	Yes
Visual Perception	Yes	Yes	Yes	Yes
Memorization/Learning	No	No	Yes	Yes
Interaction on the behalf of Visual Perception	No	No	No	Yes
Localization	No	No	Yes	Yes
Emotions based Visual Selection	No	No	No	Yes
Goal Generation	No	No	No	Yes

5. Conclusion

Surveying the literature about the attention based vision it is assumed that Visual attention is the most significant mechanisms of perception that empowers humans to efficiently select the visual data of most prominent interest. Machines confront comparable difficulties as people: they need to manage a lot of information and need to choose the most encouraging parts. Visual Selective Attention is a paradigm that cannot be ignored since it is an evolutionary advantage emerged in cognitive architecture. This indicates that it might be beneficial if applied to cognitive agents. The proposed model is an attempt in this regard. Effective utilization of Visual Selective Attention is entangled with the general cognitive abilities. We have found encouraging results and we are in a better position to find some new ways for solving the problem of Visual Selective Attention in an effective manner.

6. FUTURE WORKS

Several improvements can be made in the future version of the model since only the basic functionality for visual selective attention using bottom-up & top-down approach has been considered in the model. In the future, the application will be modified for the detection & recognition of objects from videos. This research has significant implications in the area of education in various ways and provides more interactivity in the field of education and learning. In future, video processing will be added to the design. It will modify the mechanism of Visual selective Attention to create more interactive cognitive agent. An agent will communicate more efficiently with

environment if it can perceive and express emotions. Affective communication may involve giving agents the ability to recognize emotional expressions as a step toward interpreting the situation of what the user might be feeling or what the environment demands from the agent. That situation directly affects to the system's behavior. So adding emotions will surely make interaction between agent, humans & environment more familiar and human-like. Furthermore, the Output Analyzer module will be added to the design. The purpose of Output Analyzer will to check the accuracy of response (output) in current scenario. If response will considered not being appropriate to the current scenario, then that response will not be sent to Actuator.

References

[1] K. Jokinen, "Natural Interaction in Spoken Dialogue Systems," *Work. Ontol. Multilinguality User Interfaces.HCI Int.* , vol. 4, pp. 730–734, 2003.

[2] M. Yousfi-Monod and V. Prince, "Knowledge Acquisition Modeling Through Dialogue Between Cognitive Agents," *Int. J. Intell. Inf. Technol.*, vol. 3, no. 1, pp. 60–78, 2007.

[3] P. Oborski, "Man-machine interactions in advanced manufacturing systems," *Int. J. Adv. Manuf. Technol.*, vol. 23, no. 3–4, pp. 227–232, 2004.

[4] M. Melichar, "Design of multimodal dialogue-based systems," *Communications*, vol. 4081, 2008.

[5] Karray, F., Alemzadeh, M., Saleh, J. A., & Arab , M. N. (2008). Human-Computer Interaction: Overview on State of the Art. *1*(1).

[6] Arora, S., Batra, K., & Singh, S. (2009). *Dialogue System: A Brief Review.* Punjab Technical University, Kapoorthala.

[7] H. Seidel, "Multimodal Computing and Interaction – Robust , efficient , and intelligent processing of text , speech and visual data," no. September, pp. 3–6, 2009.

[8] Harris, & Marry, D. (1985). *Introduction to Natural Language Processing.* USA: Prentice Hall.

[9] Salazar, W. Landers, D. M., Petruzello, S.J. Han, M.W. Crews, D.J. and Kubitz, K.A. Hemispheric asymmetry, cardiac response and performance in elite archers. Research Quarterly for Exercise and Sport, 61, 351-359, 1990

[10] Jackson, S.A and Roberts, G.C. Positive performance states of athletes: toward a conceptual understanding of peak performance. The Sport Psychologist. 6, 156-171. 1992.

[11] Vealey, R.S , Hayashi, S.W. Graner-Holman, M. and Giacobbi, P. Sources of sport confidence: conceptualization and instrument development. Journal of Sport and Exercise Psychology, 20, 54-80, 1998.

[12] Huang, C.J, The development of sources of confidence scale for athletes. The University of Physical Education and Sports, 5, 91-101, 2003.

[13] DeFransesco, C. and Bruke, K.L. Performance enhancement strategies used in a professional tennis tournament. International Journal of Sport Psychology, 28,185-195,1997

[14] Wang S.L. Emotional Issues, research and strategy of adolescents. Taipei, Taiwan: Ho-Chi, 1995.

[15] H.C. Nothdurft, "Salience of Feature Contrast," Neurobiology ofAttention, L. Itti, G. Rees, and J. K. Tsotsos, eds., Academic Press,2005.

[16] J.M. Henderson and A. Hollingsworth, "High-Level Scene Perception," Ann. Rev. Psychology, vol. 50, pp. 243-271, 1999.

[17] A.L. Yarbus, Eye-Movements and Vision. Plenum Press, 1967.

[18] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," IEEE Trans. Pattern Analysis and Machine Intelligence vol. 20, no. 11, pp. 1254-1259, Nov. 1998.

[19] O. Le Meur, P. Le Callet, and D. Barba, "Predicting Visual Fixations on Video Based on Low-Level Visual Features," Vision Research, vol. 47/19, pp. 2483-2498, 2007.

[20] V. Navalpakkam and L. Itti, "An Integrated Model of Top-Down and Bottom-Up Attention for Optimizing Detection Speed," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2006.

[21] G. Kootstra, A. Nederveen, and B. de Boer, "Paying Attention to Symmetry," Proc. British Machine Vision Conf., pp. 1115-1125, 2008.

[22] S. Marat, T. Ho-Phuoc, L. Granjon, N. Guyader, D. Pellerin, and A. Gue´rin-Dugue´, "Modeling Spatio-Temporal Saliency to Predict Gaze Direction for Short Videos," Int'l J. Computer Vision, vol. 82, pp. 231-243, 2009.

[23] N. Murray, M. Vanrell, X. Otazu, and C. Alejandro Parraga, "Saliency Estimation Using a Non-Parametric Low-Level Vision Model," Proc. IEEE Computer Vision and Pattern Recognition, 2011.

[24] Ballast D. K. 2002. Interior design refrence manual , Belmont CA: Professional Pub

[25] Mahnke, F. 1996. color, environment , human response, NewYork : Van Nostrand Reinhold.

[26] Bar M., Neta M. (2006). Humans prefer curved visual objects. *Psychol. Sci.* 17 645–648. 10.1111/j.1467-9280.2006.01759.

[27] Reber R., Schwarz N., Winkielman P. (2004). Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience? *Pers. Soc. Psychol. Rev.* 8 364–382.

[28] Palmer S. E., Schloss K. B., Sammartino J. (2013). Visual aesthetics and human preference. *Annu. Rev. Psychol.* 64 77–107. 10.1146/annurev-psych-120710-100504.

[29] Ball W., Tronick E. (1971). Infant responses to impending collision: optical and real. *Science* 171818–820. 10.1126/science.171.3973.818.

[30] Hillstrom A. P., Yantis S. (1994). Visual motion and attentional capture. *Percept. Psychophys.* 55399–411. 10.3758/BF03205298

[31] Aronoff J., Barclay A. M., Stevenson L. A. (1988). The recognition of threatening facial stimuli. *J. Pers. Soc. Psychol.* 54 647–655. 10.1037/0022-3514.54.4.647

[32] Amir O., Biederman I., Hayworth K. J. (2011). The neural basis for shape preferences. *Vision Res.* 512198–2206. 10.1016/j.visres.2011.08.015