A Bioinspired Approach to Vision

Daniel Camilleri, Luke Boorman, Uriel Martinez,
Andreas Damianou, and Tony Prescott

Psychology Department, University of Sheffield,
Western Bank, Sheffield, United Kingdom
{d.camilleri}
http://www.sheffield.ac.uk

Abstract. *This paper describes the design of a computational vision framework inspired by the cortices of the brain. The proposed framework carries out visual saliency and provides pathways through which object segmentation, learning and recognition skills can be learned and acquired through experience.*

Keywords: Human Visual System, Bioinspired Computing, Computational Model, Vision Model

1 Introduction

Vision processing is the major signal processing pipeline of the human brain since more than 70% of outside information is assimilated and understood through our visual sense [1] thus making it the richest source of information on the immediate surroundings. It is no wonder therefore that this sensory system started very early on in the history of evolution. Fast forwarding to today we find that this visual system is now present in most cognitive creatures not least of all in humans who have a very complex and powerful visual system.

Thus, given that the visual system is so important to the understanding of everyday life, much work has gone into identifying first of all the roles of the different visual cortices in the human brain and secondly the processing occurring within each cortex in the hopes of implementing a computational model of the Human Vision System (HVS). The implementation of such a model would be very beneficial in areas such as mobile robotic navigation and machine vision to name a few.

The layout of this paper is thus as follows. Section 2 contains the investigation of the eye and the various cortices of the human brain together with their individual function and how they are connected together. Section 3 builds on the information in section 2 by identifying computational equivalents to create the final computational model. Section 4 details the work that has been carried out so far and finally section 5 concludes with a recap followed by a description of upcoming work.

2 Visual Processing in the Human Brain

The research being carried out on the visual cortices of the brain in literature takes the form of two distinct but overlapping streams. On one side psychological, neurological and physiological research aims to understand the exact methods by which the brain carries out its day to day visual processing with carefully structured studies based on human, monkey and rat subjects. This research contributes to the understanding of the brain and has been going on since the mid-19th Century.

The other stream is that of Computational Vision Modelling which started in the late 20th Century and whose research focuses on the implementation of visual models which are either derived from the understanding obtained in the other stream, or a simplification thereof.

This computational implementation together with validation against respective datasets constitutes a way of not only improving the understanding derived in the psychological studies but also a way of validating them and their underlying assumptions of operation. This contributes to further refining the initial research and/or point out differences and cases which are irreconcilable with the current model. In this section an overview is given of some of the visual cortices in the brain and the starting point of this overview is the eye itself.

2.1 Mechanical Structure of the Eye

The eye is made up of three pairs of opposing extra-ocular muscles that provide a one to one mapping of muscle configuration to eye gaze [2]. The eye is capable of performing five main types of movements. The first two, related to involuntary reflexes, called the vestibular-ocular and optokinetic reflexes, keep the point of fixation constant in the presence of bodily motion [3].

Another two types of movement, relate to the pursuit and the vergence systems [4] of the eye that track moving objects both in terms of its planar position (eyes moving in parallel to each other) as well as in depth (eyes converging or diverging to focus on an object at a particular depth). The fifth and final eye movement is the saccade which directs attention to specific areas within the surrounding environment.

One can already see a central theme emerging from the movements of the eyes demonstrating that the motion of the eyes is completely dedicated towards focusing attention in a directional manner on separate objects at a time. Moreover as described in [5] the direction depends on the informational and behavioural value of the particular point in space at that point in time.

2.2 Visual Processing in the Retina

The next step in the visual hierarchy is the retina and this is where the processing of the raw input information starts to occur. As shown by [6] the retina is responsible for the low-level image processing within the HVS and is made up of photoreceptors, bipolar cells, and ganglion cells amongst others which are loosely grouped to form two cell layers called the Outer Plexiform Layer (OPL) and the Inner Plexiform Layer (IPL).

Starting at the OPL, this layer performs log luminance equalisation of the incoming scene and applies a non-separable spatio-temporal filter to the input. This both increases the dynamic range of the input and also highlights areas within the image that contain high frequencies either in the spatial or in the temporal domain. This information is then passed to the IPL which applies two different operations. The first operation described as the parvocellular pathway further enhances the textures and contours of the scene. The second operation which occurs in parallel to the first enhances the motion detection in the scene and is carried out in the magnocellular pathway.

Once this processing has been carried out the information is then passed to the rest of the brain via the optic nerve which branches and enters both the Lateral Geniculate Nucleus and the Superior Colliculus. [7]

2.3 The Lateral Geniculate Nucleus (LGN) and the Superior Colliculus (SC)

The LGN and the SC are both situated in the mid-brain and each cortex executes a different processing function. The LGN on one hand is responsible for extracting features such as colour and contrast [8] from the information extracted by the parvocellular pathway and it is thought to do so through the extraction of colour-antagonists [9]. This compresses the information from the 3 colours detected by 3 separate cone types [10] in the retina to 2 streams of antagonistic colours, namely red-green and blue-yellow.

The SC on the other hand is subdivided into 7 cellular layers which are divided into the superficial and the deep layers. The latter receives sensory input from multiple senses including vision and sound but the former exclusively receives visual input directly from the retina. This retinal input has been shown by Sabes et al [11] to form retinotopic maps on the surface of this cortex which preserve the link to the spatial location of each input region.

This input to the superficial layers is then processed to extract visually transient information such as flicker (temporal changes in light intensity) as well as motion stimuli [12]. The output of the SC has been demonstrated in primates to stimulate the generation of spatially averaged saccades through the combination of multi-sensory information in the deep layers [13] but it also performs saccadic suppression [14] resulting in the inhibition of saccades towards uninteresting regions of the visual field.

2.4 Visual Processing Streams in Biology

From the LGN and SC onwards, the information flow splits into two main streams which are the Dorsal Stream and the Ventral Stream. These streams have been advocated and refined since their initial proposition by Mishkin et al [15] in 1983 and can be seen in Figure 1. The dorsal and ventral streams are also colloquially called the 'where' and 'what' streams respectively. As described by [16], the dorsal stream is concerned with processing the spatial features of an image and is the main driver of saccadic movements and thus visual attention.

On the other hand the ventral stream is tasked with identifying objects in the scene through the high level representation of said objects in the subject's memory. This stream drives saccades in a more indirect manner and often only in the availability of task specific demands. The following paragraphs will describe in more detail the functionality and role of the cortices for the dorsal and ventral stream which are relevant to the scope of this project.

Dorsal Stream

As stated above, the Dorsal stream is the 'where' stream and thus its main role is in the redirection of gaze as part of the oculomotor system. The structure of this stream developed in [18] depicts the connections between the different cortical regions in the human brain for the dorsal stream, of which the Frontal Eye Fields (FEF) and Supplementary Eye Fields (SEF) are of particular interest with relevance to visual saliency, which although not the sole function of these cortices, is a principal component.

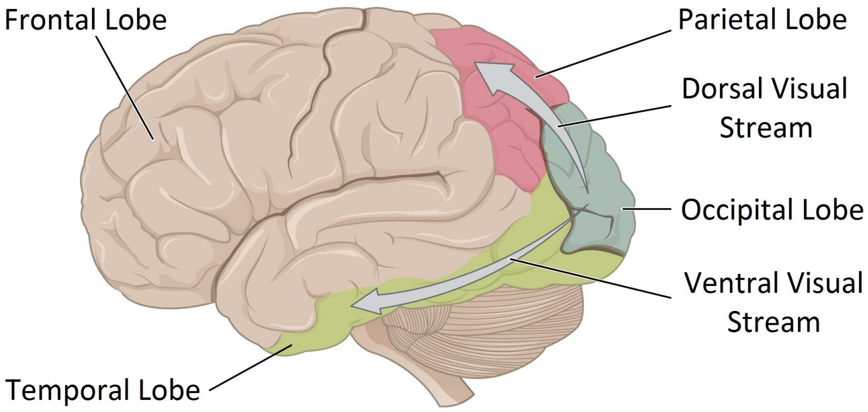


Fig. 1: Diagram of the position of visual cortices in the brain [17].

Frontal Eye Fields (FEF)

The FEF are described by literature as the "principle saccadic decision structure together with the SC" [16] and just like the SC, this cortex keeps a retinotopic map of the field of vision. Furthermore the work in [19] has shown that this cortex combines not only the information from both the dorsal and ventral stream to decide on a saccade target but has also been shown to accept input biases from areas in the Pre-Frontal Cortex (PFC) cortex that modify the selectiveness of certain properties providing a path where a high-level cortical process can tweak the functionality or priority of lower-level cortical operation such as saliency.

Supplementary Eye Fields (SEF)

The SEF are linked heavily with the FEF but while the latter generates saccadic targets, the former's task is in keeping a craniotopic [20] (relative to the head) mapping of the environment around the subject and thus hints towards a region of the brain that keeps track of the environment and provides the required data to fill in gaps in instantaneous knowledge with historic data.

Ventral Stream

The Ventral or 'what' stream starts off at the Primary Visual Cortex (V1) and was one of the first cortices to be investigated in the seminal work of Hubel [21] which describes the functionality of V1 as a hierarchical arrangement of neurons capable of extracting local features from an image such as bars or edges [22] with higher levels of the hierarchy displaying wider receptive fields as well as an insensitivity to orientation and scale.

The path of this stream starts in the primary visual cortex and then proceeds, mainly sequentially [23], from V1 through V2 up to V5 in a retinotopic manner [24, 25]. Finally, it passes through the Posterior Inferior Temporal Cortex (PIT) followed by the Anterior Inferior Temporal Cortex (AIT) which is where this visual stream ends. From here on, it branches into the medial temporal lobe and PFC cortex whose feature biasing role has been described initially. Two interesting features of the Inferior Temporal Cortex (IT) include highly specialised cells that respond to very specific stimulus along with a response that is invariant to the number of objects present called Cardinality Blindness which is an effect of the trend for wider receptive fields and rotation invariance displayed by the ventral stream.

As such, instead of having a retinotopic map like the dorsal stream, the ventral stream at the level of the IT is described as having a sparse representation of all current recognisable and behaviourally relevant objects in the visual field with more of an emphasis on the classification rather than the location of said object.

In view of this basic introduction to some of the cortical regions relevant to the HVS, the next section presents a computational framework that encompasses the functionality of all the different regions.

3 Human Visual System Computational Framework

The consolidation of facts from the previous section points towards the 6 low-level features of importance to the HVS: Intensity, Red-Green Antagonist and Blue-Yellow Antagonist attributed to the LGN, Flicker and Motion attributed to the SC and Orientation of edges attributed to the Primary Visual Cortex V1. These 6 features form the basis upon which the HVS carries out the rest of its functions. As such the proposed Computational Framework looks at the extraction of these features followed by their possible application specifically to visual saliency, object learning and object recognition from a robotic viewpoint.

3.1 Visual Saliency

According to [16], visual attention is defined as the "process of enhancing the responses of neurons that relate a subset of the visual field with the purpose of overcoming the computational limitations of the visual system" and its importance as mentioned before lies in the reduction of the computational burden on the human brain.

Building on this, the book *Selective Visual Attention : Computational Models and Applications* [1], through the use of various experiments and observations, identifies two important aspects of visual attention. The first aspect is that visual attention is divided temporally into pre-attention followed by attention. The second aspect is that the attention phase has two operating modes which are Top-Down and Bottom-Up.

Pre-Attention and Attention

This aspect of visual attention consisting of two sequential processes is described by [26], [27] and [28]. The first process, pre-attention, performs the extraction of the low level features. This is performed very quickly, in parallel and on the whole visual field imitating the functionality of the LGN, V1 and part of the SC. Subsequently, once all the features have been explicitly extracted, the second process, attention, takes over. In this stage, the low level features are combined together through a process called feature integration which is proposed in [28].

However according to [28], feature integration in the HVS is only carried out through visual attention. Therefore, this means that although the visual system could perceive the presence of multiple objects with a particular combination of features, it does not explicitly know their location due to the cardinality of the ventral stream. Thus the HVS has to resort to a serial search over the visual field in order to locate objects with a combination of features.

Top-Down and Bottom-Up Drives

Secondly, according to [29–31], a subject's gaze is drawn to a particular point due to a combination of bottom-up stimuli that arise solely from changes within the scene and top-down stimuli that modulate the attention depending on the task at hand. This separation of function provides a powerful scheme for object recognition mimicking the feedback loop of the FEF and PFC. Furthermore it has led to the development of two classes of saliency algorithms in literature which are either unguided (bottom-up) or guided (top-down effects).

In the unguided scenario, also known as 'free-roaming', the weights for all feature maps are identical and the final saliency map is generated through the summation of all equally weighted feature maps like Itti et al's Baseline Saliency Model [32]. On the other hand the guided scenario modifies the feature weightings independently depending on the task at hand, examples of which are [33, 34] which demonstrate limited top-down effects due to the complexity of factors contributing to this mode.

Of important note is that humans start out life with unguided saliency [35] and later start paying guided attention to specific objects signifying an accumulation of knowledge before the application of top-down effects: knowledge which arises from object learning.

3.2 Proposed Model and its Implementation

Figure 2 provides a visualisation of the information flow within the model incorporating all the processes which have been mentioned so far while extending the application to object recognition and including a path for top-down modulation of saliency. The main contribution of this model is the use of a memory model that is very similar in functionality to that of the a human memory system and the use of this memory model for the purpose of object learning, recognition and searching. The process starts with the eyes where image capture is followed by the extraction of the corresponding parvo and magno images using the algorithm developed in [6].

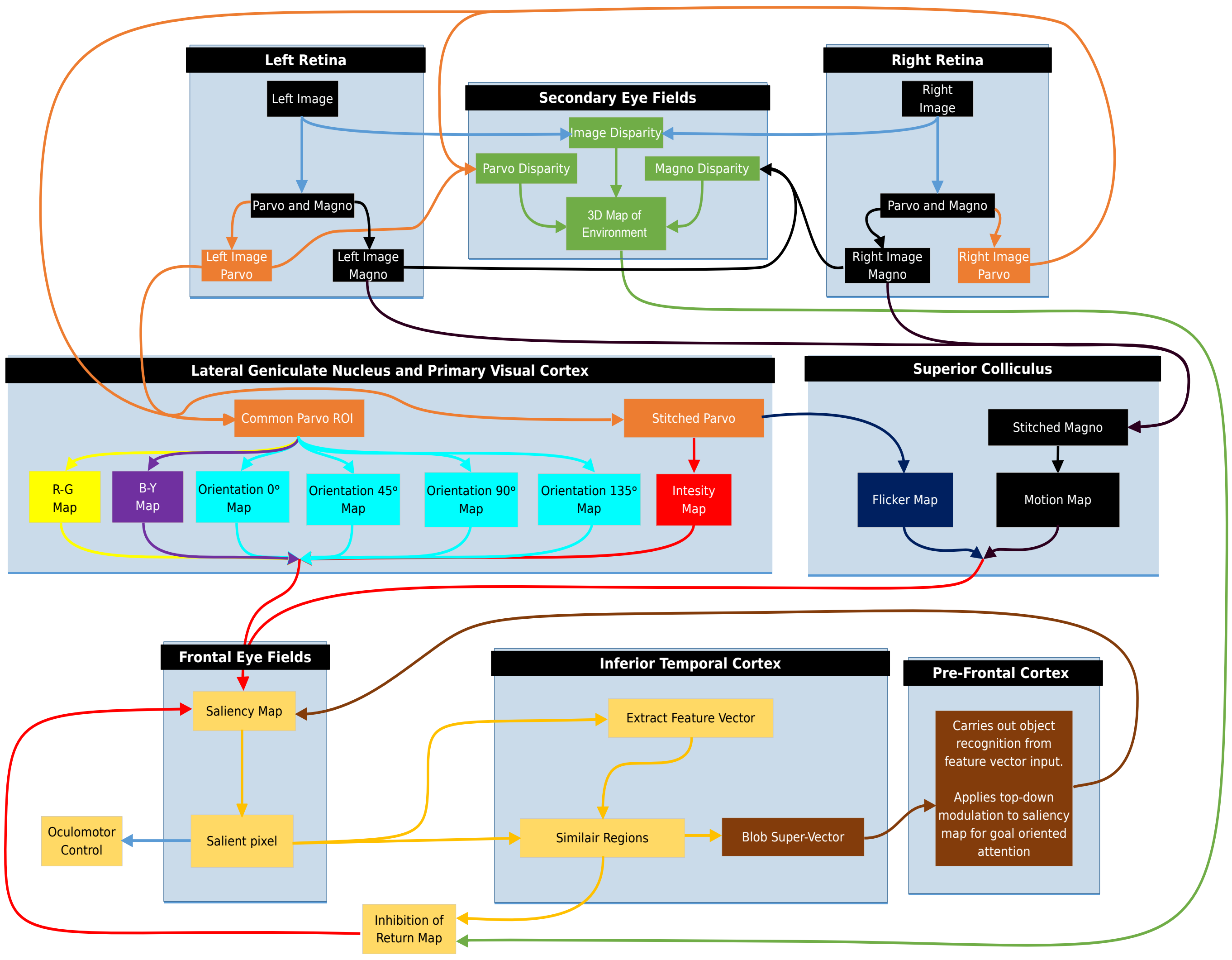


Fig. 2: Proposed Block Diagram for the Computational Framework

Subsequently, the overlapping area of both parvo images is extracted in order to be processed for R-G, B-Y and Orientation while both images are stitched together for the extraction of Intensity and Flicker. This is analogous to human vision where the central region of focus is processed to identify details and has a high presence of cones attributed to colour perception while changes in light intensity are computed over the whole field of view due to the uniform density of rods in the retina. [36]. Similairly, motion is described as the temporal change of light intensity on an array of detectors by [37] thus the extraction of motion is also carried out over the whole image which is why magno images are stitched together.

From this point, base feature map extraction at the capture resolution for the LGN and V1 block follows the process of visual saliency extraction by Itti et al [32]. In the case of the SC block, flicker base map is computed as described in [38] while the extraction of motion requires no further computation to extract the base feature map. Subsequently a Gaussian pyramid with 8 levels is created for each base feature image imitating the effect of center-on surround-off effects of the retina. The 8 levels are then normalised and collapsed to a level of choice into a conspicuity map. The lower the level, the higher the spatial resolution of the saliency map but the lower the noise to signal ratio.

From here on, the conspicuity maps are combined and averaged depending on the weights from the PFC block which are initially set as equal. The rest of the blocks then carry out object learning and recognition based on the extracted features.

In this framework, object learning is focused and directed by saliency much like infants. [39]. Thus once a salient point has been identified, a feature vector describing the point is created which, in turn, is used to find all areas within the image that correspond to this vector thus creating blobs of common regions. The blob which contains the salient point is subsequently chosen and a super-vector is created from the feature vectors of all the pixels in the blob. This feature vector is then passed to the PFC which carries out statistical learning on the super-vector with the use of SAM [40]. Thus given multiple inputs over the course of multiple frames, SAM creates a latent feature model of the input data which condenses the data into its most important features and also allows for the labelling of blobs. This then allows for one of two processes to take place.

The first, given an input blob, one could carry out object recognition and return the label of the currently salient object. The second, given a label to look for, SAM returns a feature vector best describing the object with a mean and variance, which is applied as weighting to the saliency map. This results in the extraction of multiple blobs which can then be sequentially tested through the object recognition route thus imitating the effect of cardinality blindness.

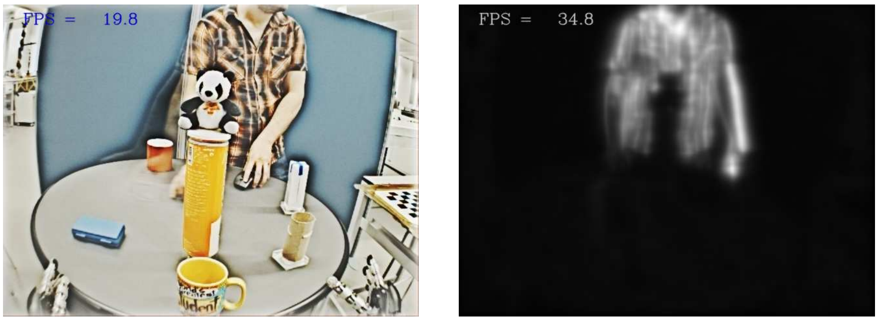
Furthermore, given the embodiment of such a system within a robot such as the iCub [41], would allow for a great refinement in the process through interaction just like infants [42]. Thus when an object becomes salient, the robot can interact with the object causing its motion. This would then provide an excellent trigger for the extraction of precise object boundaries by weighting the motion saliency more heavily and thus a better and more accurate super-vector is created given this boundary.

Finally, after the current salient point has been processed, an inhibition is required to be applied to the saliency map in order for the observer to choose a new location for processing to which there are two possible approaches. The first approach, which is depicted within the block diagram, requires the tracking of the current salient point within an internal 3D map that is accumulated through disparity maps obtained from multiple sources for reasons of data density.

Another possible approach would be to keep the past feature vector, invert its direction and apply a modified weighting to the saliency map thus encouraging a new location. In this manner, multiple past vectors could be subsequently accumulated into a single vector with a temporal diminishing factor applied recursively such that inhibition fades with time.

4 Preliminary Results

The implementation of the proposed visual model is currently being carried out on the iCub robotic platform. Due to the complexity of the model, a networking platform is required in order to be able to split the computation between several interconnected computers and for this purpose YARP [43] is used since it is also the networking interface used by the iCub. For the image processing aspect, OpenCV 3.0.0 [44] is being used compiled with Compute Unified Device Architecture (CUDA) 6.5 [45] to leverage the acceleration of NVIDIA GPUs with the aim of delivering a realtime system (30fps).



(a) Enhanced Texure and Contours

(b) Motion map

Fig. 3: Parvo and Magno output for the Retina Model with their corresponding frame rates for CPU computation at an input resolution of 640x480

The realisation of this model has so far achieved execution of five out of eight blocks. The first two being the retina models which utilise the bioinspired module that is available as part of the contrib modules for OpenCV3 submitted by the same authors of [6], the output of which can be seen in Figure 3

The resulting images are then passed to the third and fourth blocks currently implemented that are the LGN and V1 block and the SC block which carry out Itti's Baseline Saliency Model [32] for the generation of a saliency map as shown in Figure 4 at a resolution of 40x32 for input images at 640x480 resolution.

From this Saliency Map, the most salient pixel is chosen and passed to the oculomotor controller that directs the gaze of the iCub towards that location. Furthermore, a primitive inhibition of return has also been applied as a substitute for the inhibition of return that will be implemented after object learning has been achieved. The current method retains a 40x32 map in memory with saliency inhibition values applied to the respective pixel that is currently the focus of attention. This primitive implementation does not take into consideration that the iCub head changes orientation as it looks towards a salient point but still, the result, which can be seen in [46], displays promising behaviour.

Furthermore there are two main avenues for future improvement. The first is the implementation of a computationally fast stitching process that takes advantage of the known camera orientations which would allow for the application of separate regions for colour processing and motion processing because so far, all stages beyond the retina use the overlapped region thus mimicking a retina that has a constant distribution of both cones and rods.

The second improvement deals with the effect of camera motion on the magno and parvo images. Currently, the embodiment of the system results in blurry magno and parvo images whenever a saccade or head motion is executed. It just happens that the HVS also encounters this problem which it solves through the

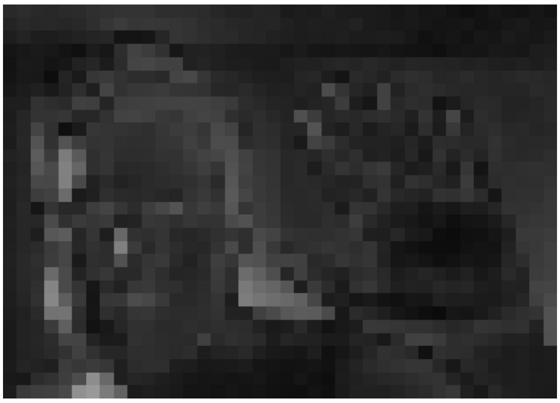


Fig. 4: Saliency Map at a resolution of 40x32

application of a process called saccadic masking [47]. Thus blur will be mitigated in the future through a communication loop that blocks image input to the model whenever an oculomotor command is sent.

The model's computation is divided between 2 computers. The first carries out both retina models on a Xenon 6 Core CPU at 2.8GHz which is a purely CPU based implementation running at an average of 25fps. The second carries out the saliency computation and robot control on a Core i7-3630QM CPU running at 2.4GHz with NVIDIA GTX 675MX GPU having 4GB of VRAM at an average rate of 15fps. Thus assuming negligible transfer delays due to low resolution images being transferred on high bandwidth local networks, the whole system runs at a throughput of 15fps with an end-to-end delay of approximately a 100ms.

5 Conclusion

In conclusion, [28] states that "without attention, general purpose vision is not possible" thus this work has reviewed the different cortices in the human brain that process the visual input of the eyes and identified the key functionalities of each. Subsequently, these key functions were summed into an extraction of six important low-level features and their role within visual saliency.

Following this, bioinspired computational implementations for visual saliency were investigated to establish the computation of these features as well as that of the saliency map. A method for object learning, object recognition and possible interactive object segmentation are then proposed, together with two possibilities for the application of inhibition of return which leads to dynamic behaviour. Finally a description of the work that has been carried out is provided with performance results for the currently implemented blocks as well as some pitfalls that have been encountered along the way.

As one can see the implications of the model are very powerful and allow for a range of possibilities in behaviour. Currently the implementation of this model is in development and future work will look into documenting the efficacy of learning using the process state above as well as the level of fidelity with datasets of human visual saliency.

Acknowledgments I would like to thank all those who helped and inspired me especially Tony Prescott, Luke Boorman, Uriel Martinez and Andreas Damianou.

References

1. Zhang, L., Lin, W.: Selective visual attention: Computational models and applications. John Wiley & Sons (2013)
2. Subramanian, P.S.: Active vision: The psychology of looking and seeing (2006)
3. Schweigart, G., Mergner, T., Evdokimidis, I., Morand, S., Becker, W.: Gaze stabilization by optokinetic reflex (okr) and vestibulo-ocular reflex (vor) during active head rotation in man. Vision research **37**(12) (1997) 1643–1652
4. Robinson, D.A.: Eye movement control in primates. Science **161**(3847) (1968) 1219–1224
5. Henderson, J.M.: Human gaze control during real-world scene perception. Trends in cognitive sciences **7**(11) (2003) 498–504
6. Benoit, A., Caplier, A., Durette, B., Hérault, J.: Using human visual system modeling for bio-inspired low level image processing. Computer vision and Image understanding **114**(7) (2010) 758–773
7. Schiller, P.H.: Central connections of the retinal on and off pathways. (1982)
8. Gao, D., Mahadevan, V., Vasconcelos, N.: The discriminant center-surround hypothesis for bottom-up saliency. In: Advances in neural information processing systems. (2008) 497–504
9. Rodieck, R.: Which cells code for color? In: From pigments to perception. Springer (1991) 83–93
10. Mullen, K.T.: The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. The Journal of Physiology **359**(1) (1985) 381–400
11. Sabes, P.N., Breznen, B., Andersen, R.A.: Parietal representation of object-based saccades. Journal of Neurophysiology **88**(4) (2002) 1815–1829
12. Wurtz, R.H., Albano, J.E.: Visual-motor function of the primate superior colliculus. Annual review of neuroscience **3**(1) (1980) 189–226
13. Glimcher, P.W., Sparks, D.L.: Representation of averaging saccades in the superior colliculus of the monkey. Experimental Brain Research **95**(3) (1993) 429–435
14. Sommer, M.A., Wurtz, R.H.: Visual perception and corollary discharge. Perception **37**(3) (2008) 408
15. Mishkin, M., Ungerleider, L.G., Macko, K.A.: Object vision and spatial vision: two cortical pathways. Trends in neurosciences **6** (1983) 414–417
16. Cope, A.J.: The role of object recognition in active vision : a computational study. Ph.d (September 2011)
17. : Illustration from anatomy & physiology, connexions web site. https://commons.wikimedia.org/wiki/File:1424_Visual_Streams.jpg Anatomy & Physiology, Connexions Web site. http://cnx.org/content/col11496/1.6/, Jun 19, 2013. [Accessed 1-February-2016].
18. Hikosaka, O., Takikawa, Y., Kawagoe, R.: Role of the basal ganglia in the control of purposive saccadic eye movements. Physiological reviews **80**(3) (2000) 953–978
19. Bichot, N.P., Thompson, K.G., Rao, S.C., Schall, J.D.: Reliability of macaque frontal eye field neurons signaling saccade targets during visual search. The Journal of Neuroscience **21**(2) (2001) 713–725
20. Schall, J.D., Morel, A., Kaas, J.H.: Topography of supplementary eye field afferents to frontal eye field in macaque: implications for mapping between saccade coordinate systems. Visual neuroscience **10**(02) (1993) 385–393
21. Hubel, D.H., Wiesel, T.N.: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of physiology **160**(1) (1962) 106

22. Lee, T.S., Mumford, D., Romero, R., Lamme, V.A.: The role of the primary visual cortex in higher level vision. Vision research **38**(15) (1998) 2429–2454
23. Felleman, D.J., Van Essen, D.C.: Distributed hierarchical processing in the primate cerebral cortex. Cerebral cortex **1**(1) (1991) 1–47
24. Tanaka, K.: Inferotemporal cortex and object vision. Annual review of neuroscience **19**(1) (1996) 109–139
25. Gross, C.G.: Single neuron studies of inferior temporal cortex. Neuropsychologia **46**(3) (2008) 841–852
26. Neisser, U.: Cognitive Psychology: Classic Edition. Psychology Press (2014)
27. Hoffman, J.E.: Hierarchical stages in the processing of visual information. Perception & Psychophysics **18**(5) (1975) 348–354
28. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. In: Matters of intelligence. Springer (1987) 115–141
29. Wolfe, J.M.: Guided search 2.0 a revised model of visual search. Psychonomic bulletin & review **1**(2) (1994) 202–238
30. Wolfe, J.M., Cave, K.R., Franzel, S.L.: Guided search: an alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human perception and performance **15**(3) (1989) 419
31. Navalpakkam, V., Itti, L.: Top–down attention selection is fine grained. Journal of Vision **6**(11) (2006) 4
32. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis & Machine Intelligence (11) (1998) 1254–1259
33. Wolfe, J.M., Cave, K.R., Franzel, S.L.: Guided search: an alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human perception and performance **15**(3) (1989) 419
34. Wolfe, J.M.: Guided search 2.0 a revised model of visual search. Psychonomic bulletin & review **1**(2) (1994) 202–238
35. Galazka, M., Nyström, P.: Visual attention to dynamic spatial relations in infants and adults. Infancy **21**(1) (2016) 90–103
36. Saleh, M., Debellemanière, G., Meillat, M., Tumahai, P., Garnier, M.B., Flores, M., Schwartz, C., Delbosc, B.: Quantification of cone loss after surgery for retinal detachment involving the macula using adaptive optics. British Journal of Ophthalmology (2014) bjophthalmol–2013
37. Borst, A., Egelhaaf, M.: Principles of visual motion detection. Trends in neurosciences **12**(8) (1989) 297–306
38. Itti, L., Baldi, P.: A principled approach to detecting surprising events in video. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Volume 1., IEEE (2005) 631–637
39. Pruden, S.M., Hirsh-Pasek, K., Golinkoff, R.M., Hennon, E.A.: The birth of words: Ten-month-olds learn words through perceptual salience. Child development **77**(2) (2006) 266–280
40. A. Damianou L. Boorman, N. Lawrence, T.P.: A top-down approach for a synthetic autobiographical memory system. In: Proceedings of the 4th International Conference on Biomimetic and Biohybrid Systems (Living Machines). (2015)
41. IIT: icub: an open source cognitive humanoid robotic platform. http://www.icub.org/ [Accessed 1-February-2016].
42. Kellman, P.J., Spelke, E.S., Short, K.R.: Infant perception of object unity from translatory motion in depth and vertical translation. Child development (1986) 72–86

43. YARP: Yet another robot platform. http://wiki.icub.org/yarpdoc/ [Accessed 10-February-2016].
44. Itseez: Opencv 3.0.0. http://opencv.org/opencv-3-0.html [Accessed 10-February-2016].
45. NVIDIA: Cuda developer website. https://developer.nvidia.com/cuda-toolkit [Accessed 10-February-2016].
46. Camilleri, D.: icub visual saliency. https://www.youtube.com/watch?v=_OgBuLZHCh8 [Accessed 10-February-2016].
47. Burr, D.C., Morrone, M.C., Ross, J., et al.: Selective suppression of the magnocellular visual pathway during saccadic eye movements. Nature **371**(6497) (1994) 511–513