Inhibition of Return Using an Adaptive Threshold Map

Ahmed Sheikh Deeb
School of Engineering and Computer Science
Victoria University of Wellington
Wellington, New Zealand
ahmed.sheik.deeb@ecs.vuw.ac.nz

Christopher Hollitt
School of Engineering and Computer Science
Victoria University of Wellington
Wellington, New Zealand
christopher.hollitt@ecs.vuw.ac.nz

***Abstract*—Active vision systems inspired by human or animal vision often employ a data driven approach, where the gaze of the machine is directed to regions of high saliency in the visual field. One common problem with such an approach is that the system can become fixated on a stimulus of high saliency and fail to detect the presence of other important features that have lower saliency. In animals the mechanism of inhibition-of-return operates to prevent prolonged fixation. We discuss an artificial inhibition-of-return system for use in a machine active vision system. Our system relies on a adaptive threshold map that changes the salience required for a stimulus to be examined. Recent fixation points raise the saliency threshold, so that the vision system is driven to examine less salient parts of the visual field. We examine the changes in the visual foraging behaviour caused by varying the rate at which the system forgets about previous fixation locations.**

***Keywords*—Inhibition of return; visual attention; foraging; saccadic eye movements; saliency; superior colliculus.**

I. Introduction

The acquisition of visual data is now a commonplace part of many research and industry applications. However, processing of high resolution images data in real time remains a computationally challenging problem. Active vision provides one approach that can mitigate the data processing challenge. An active vision system can employ intelligence to the collection of visual data. That is, an active vision system can direct the vision system to preferentially collect data that is expected to be of use for a particular application. This should be contrasted with passive vision systems that typically collect and process large amounts of data that is of low value for the task.

A necessary aspect of an active vision system is capacity to direct the gaze towards different locations in the visual scene. The analogous behaviour in animals and humans has been long studied and a generally accepted model for the observed behaviour has been proposed [1]. This model posits that objects of high salience tend to attract visual attention, leading to the generation of saccades that orient the gaze with the appropriate stimulus [2].

A representation of the surrounding environment is stored in a saliency map. This map encodes how interesting or striking the various areas of the visual field appear. Many aspects of an object or area can lead to it gaining high saliency, including colour, shape, orientation or movement [3]. In animals the superior colliculus is thought the be the brain structure responsible for combining various feature maps to form the overall saliency map that indicates where the animal should direct its gaze.

One problem encountered when implementing a saccade generation system is that the system may become fixated on the most salient stimulus in the visual field. For some machine vision tasks (such as tracking a single target) this causes no difficulties. However, in tasks requiring a broader sense of the environment this prolonged fixation behaviour is not desirable. For many such tasks the desired behaviour is that of visual foraging, where the environment is continually scanned for pertinent information [4].

An important mechanism underpinning successful visual foraging systems is known as inhibition of return (IOR) [5]. This mechanism was first discussed by Posner and Cohen [6], who suggested an algorithm that guarantees releasing the recently attended location and moving on to the next [7], [8]. Since the late 1990s, active vision systems have typically included some form of IOR [9], [10]. However, an implementation with the richness of behavior seen in biological systems is still lacking, in part because the neural mechanism describing biological IOR is still not well understood [11].

In this paper, the overall architecture of a saccade generation system including IOR is presented. This technique requires the computation of an adaptive threshold map that modulates the salience necessary to generate a saccade. Biological systems are believed to contain an analogue of this map in the superior colliculus [2]. The particular focus of our work is to produce an adaptive threshold map so that the systems visual foraging behaviour can be changed depending on the visual environment. Our intent is not to replicate the details of the animal model, but rather to provide the same basic IOR functions that are performed by the human superior colliculus.

In section II we will define inhibition of return and review some previous machine implementations of IOR. In section III we review our architecture used for building the threshold map and discuss its function. This is followed by discussion of our proposed implementation of an adaptive threshold map and of simulated results in sections IV and V respectively. In section VI our proposed system is discussed and a summary is provided.

II. Background

The neuronal process of that prevents recently attended locations being re-examined is known as inhibition of return [11]. This mechanism has been widely studied mainly in humans an in primates [5], [8], [12].

The purpose of this mechanism is to ensure an adequate distribution of the attention across the visual scene. Such a strategy is vital for animals which require a general sense of their surroundings in addition to focused attention on a particular task. Unfortunately, inhibition of return is not yet well understood in biological systems [8], as multiple areas of the brain are known to have effects on saccadic and other eye control movements [13]. The interactions between the various structures results in a biological system of considerable complexity that remains the subject of active research.

Biological studies have shown that IOR can be associated with either objects or locations [7]. That is, IOR can prevent the return of visual attention to previously examined objects, or to previously examined locations. One advantage of a space-based model is that it allows the study of IOR for more than one object at the same time [14]. A location based approach is also more in keeping with the low-level processing thought to occur in the superior colliculus. An object based technique would require considerably more sophisticated image processing to manage the objects present in a scene.

From a biological point of view the short term memory associated with IOR in primates is limited to five seconds or objects [7] depending on the details of the situation. That is, the IOR system typically discourages saccades to the five most recently revisited objects or the areas examined in the previous five seconds.

When implementing artificial IOR systems, some researchers have chosen to store locations of all previous fixation points. This requires a reasonable amount of memory dedicated to the saccade generation system [11].

Others have preferred editing the visual scene itself by suspending the previous attended locations from competing for attention. This is typically done by reducing the saliency of previously attended objects. This approach requires installing a habituation filter [13], [15] to modify the saliency map. Although this technique has many applications, the concept of manipulating the saliency map can be suboptimal as it changes the properties of represented objects. This may result in faulty calculations of how interesting different objects are, particularly if the objects are changing.

III. System Architecture

The overall architecture of our system is inspired by the neural structure of primates and is shown in Fig.1. Our artificial superior colliculus contains two important maps, the saliency map and the threshold map. The diagram indicates the main interactions within the superior colliculus, but also shows the flow of data between associated areas of the system.

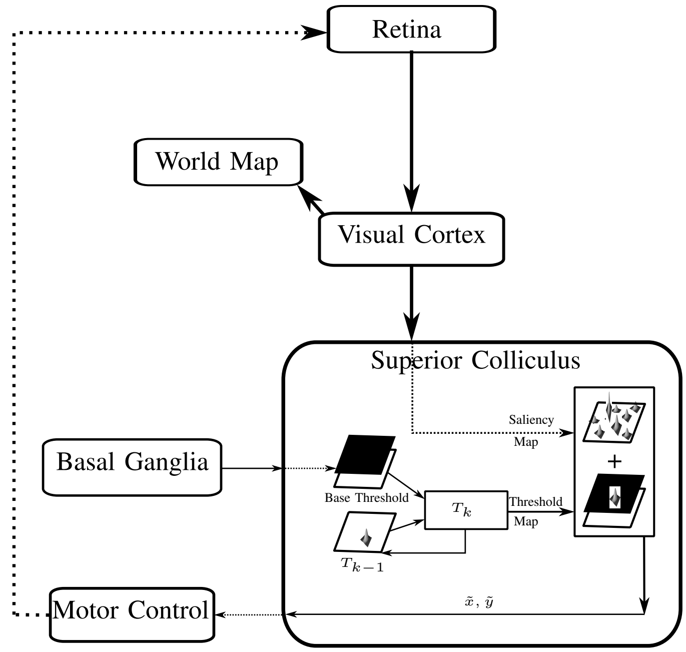

Fig. 1. Proposed mechanism for inhibition of return.

A. Retina

The retina acts as the input device of the biological system [14] and attention can be regarded as being directed to the objects lying in its central region (the fovea). Such objects can be regarded as being within a "spotlight of attention" [14]. In our implementation the retina functions are performed by a low cost CCD camera that passes its output image stream to the visual cortex.

B. Visual Cortex

The visual cortex is the area of the brain that is responsible for converting raw input image data into meaningful information. Our present system implements only a tiny part of the functions of an animal's visual cortex, by receiving the input signal from the retina and translating it into basic feature maps. The presence of features in these maps is represented by impulse functions [16]. For the purpose of this paper we have simplified the visual cortex to produce only a single feature map corresponding to the size of object.

C. World map

The awareness of the surrounded area plays an important role in managing the state of awareness in animals. We use a world map to store the machine's total knowledge of the world. It is an accumulative map that is generated by combining the feature maps from the virtual cortex as attention is shifted around the visual field. This map is designed to grow until the system gets full knowledge of the surrounding environment.

D. Basal Ganglia

The basal ganglia (BG) is associated with management of the excitation level of an organism [17]. This influences visual attention, as a highly excited organism tends to saccade more frequently. The Basal Ganglia can be considered as setting the minimum saliency that the system regards as being worth examination. In other words, it tunes the sensitivity of the saccade generation system. In animals the activity level in the BG is dependent on many factors such as the age and experience and the situation. In our model the BG manages the default level of the minimum interest.

E. Superior Colliculus

The superior colliculus (SC) is the main controller of saccadic eye movements and of gaze control in general [2] [18]. In animals the SC combines feature maps from multiple parts of the visual cortex [17], as well as spatial maps from other modalities, such as the audio and somato-sensory systems [19]. When a stimulus of sufficient intensity is detected the SC generates a signal to direct the sensory system at the stimulus.

Our implementation of the superior colliculus combines the various feature maps generated by the visual cortex after prescaling, smoothing, weighting and then being convolved with a two-dimensional Gaussian function. This convolution ensures that the system attends to general areas of interest, rather than being attracted to isolated pixels that appear highly salient due to noise [20]. In addition the smoothing of the feature maps ensures that the system is insensitive to minor relative misalignment in the feature maps.

Avoidance of prolonged fixation on the most salient object in the visual field requires the superior colliculus to implement IOR. The role of IOR is to guarantee a fair visual distribution around the environment. This role is implemented using a spatially varying threshold map inside the superior colliculus [21], [22]. Only when the saliency map exceeds the threshold could the SC change the location that is to be attended. If the saliency map is globally below the threshold then the fixation point remains unaltered. The two maps computed in the superior colliculus are:

1) Saliency map

   The saliency map is the result of processing the raw data obtained from the visual cortex. It is a two-dimensional map that encodes the saliency of objects in the visual environment [20]. This map is then compared with the threshold map and whatever object's salience is more than the relative location in the threshold map is considered to be the next attended spot.

2) Threshold map

   Both the world map and the basal ganglia are involved in creating this map. The threshold map takes the coordinates axis of the saliency map. It records the attended objects or locations in the visual scene. It remembers whether the sight was directed to particular objects/locations or not. According to this map if the object is remembered, it is more unlikely that the visual system will orient towards it again [4].

IV. Proposed Solution

We have developed an inhibition of return generation system for deployment on a saccadic attention system. Two Unibrain Fire-I CCD cameras acquire 320×240 pixel images at 30 frames per second using a YUV4:2:2 format. To implement a simple foveal system the input is then down-sampled to 64×64 pixel grayscale images for use in the saccade generation system described subsequently.

The input image is then processed to generate a set of feature maps, each of which indicates the presence of a particular target in the visual field. The various feature maps are combined to form an overall saliency map. However; in this paper, we have simplified the role of the feature maps to only one feature map. A single feature is sufficient to explore the inhibition of return mechanism described in this paper. Each location in the saliency map will compete to gain attention under a winner-take-all technique. Only the location that most stands out from the surrounded. When the most salient feature exceeds the threshold then a signal is generated to initiate a saccade movement intended to move the gaze towards the greatest stimulus.

Inhibition of return is generated within a system that is normally responsible for orienting the gaze direction. The purpose of IOR is to bias the visual attention away from recently inspected items so that the search of the environment becomes more efficient [8]. The system is done by an adaptive threshold map that needs continuously updated global information. The proposed system's flow chart is shown in Fig. 2. The figure illustrates that the world map is being updated by the input image. The saliency map is generated and compared with the threshold map. For any location, if the saliency was more than the amplitude in the threshold map, then the system will shift the gaze towards it. Notice that the threshold map is being updated every iteration.

The saliency is computed for any object in the visual scene by a 2D Gaussian according to Eq. 1. This is consistent with the behavior in biological system [20], such that:

$$S(x,y) = \sum_j A_j e^{-a(r-r_j)^2} \tag{1}$$

where A_j is the saliency of the j-th object, $r_j = \sqrt{x_j^2 + y_j^2}$ is the distance between the object at (x_j, y_j) and the origin. a is the constant that determines the Gaussian spread. For this paper a was set to be 0.014. The smoothed map S forms the saliency map in the superior colliculus.

The next step is to compute the threshold map using Eq. 2. Threshold map represents a record of attended objects or locations in the visual scene. This map is responsible for remembering where the visual sight has previously been directed. The current threshold map is computed based on the previous threshold map, the current attended location and the base threshold T_{BG}. The BG is responsible of the default minimum base threshold and is neglected in order not to disturb the effect of α on the system behavior.

$$T_k(x,y) = \alpha T_{k-1}(x,y) + G(\tilde{x}, \tilde{y}) \tag{2}$$

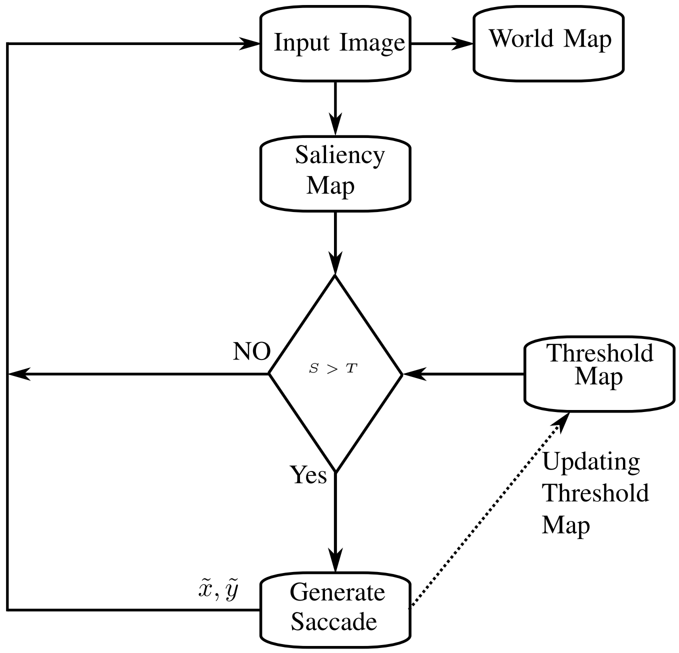

Fig. 2. System flow chart.

Where $\tilde{x}, \tilde{y}$ is the fixation point, while G is a 2D Gaussian. $T_k(x, y) \leq 1$ is assumed, since it is relative to the assumption that $A_j \leq 1$ in the saliency map. Another reason for this assumption is to prevent unattended locations' amplitude from growing beyond the maximum possible saliency. The modified threshold map is computed:

$$\tilde{T}_k(x, y) = \max(1, T_k(x, y)) \tag{3}$$

Finally, the saliency map is compared with the threshold map to determine whether a new location is to be attended. The threshold comparison is performed by calculating $U(x, y) = S(x, y) - \tilde{T}_k(x, y)$ so that only locations that have a saliency above the threshold will have positive U values. We then find a potential saccade target by locating the largest peak in U at the point $(x^\star, y^\star)$.

$$(x^\star, y^\star) = \arg\max_{x,y} U(x, y) \tag{4}$$

If $U(x^\star, y^\star) > 0$ then we generate a saccade to $(x^\star, y^\star)$, otherwise we maintain the previous fixation location. That is, only when $U(x^\star, y^\star) > 0$ has the system detected a sufficiently interesting feature that the current fixation point should be changed.

V. Results

We conducted a set of experiments to study the behaviour of our proposed system. As we wish to study α, we used a fixed single image as an input to our system. As the corresponding saliency maps that do not vary in time, we can focus on the effect of the threshold map. A typical example is shown in Fig. 3 using a series of MATLAB® simulation of the proposed system. The input image is shown in Fig. 3a, this image is processed, re-scaled and smoothed using a 2D Gaussian filter

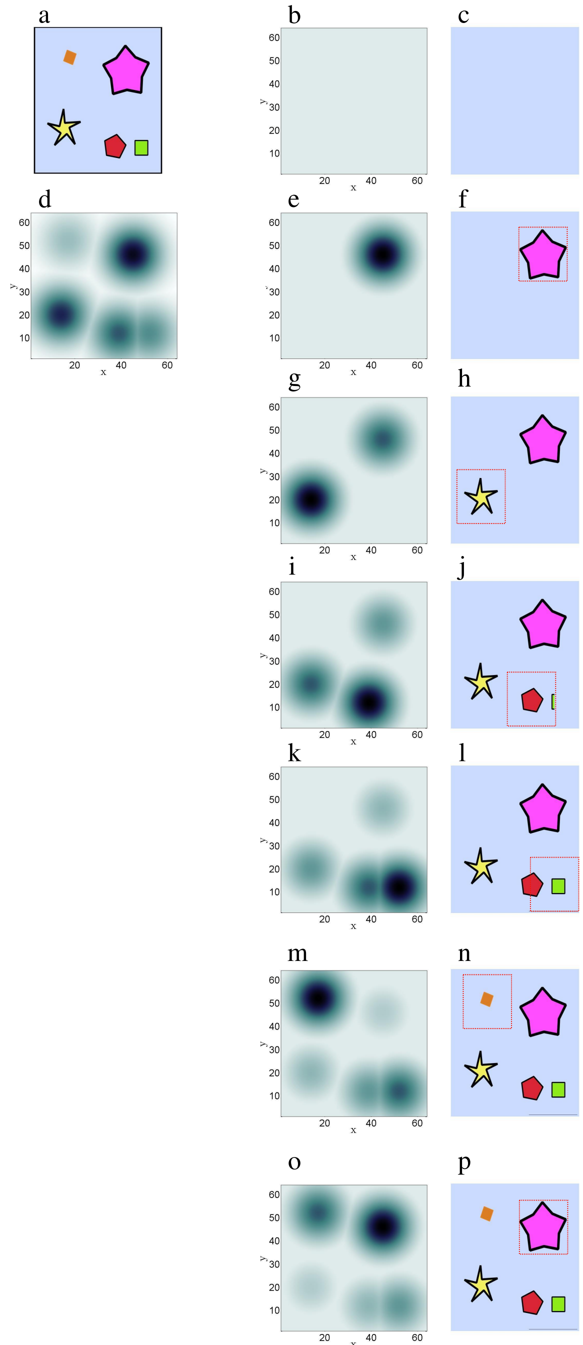

Fig. 3. Simulation results for inhibition of return threshold map and world map. Subfigure shows the input image of the available environment. Subfigers d is a 2D representation of the saliency map. Subfigures b, e, g, i, k, m and o show the proposed adaptive threshold map while subfigures c, f , h, j, l, n and p show the world map as it storers the previously attended objects while the current ones are indicated by a red square around them.

(see Eq. 1) in order to build up the saliency map shown in Fig. 3d. In the same figure threshold maps are computed using Eq. 2. The intent of this work was to assess the viability of the general approach to compute the threshold map before attempting to implement the system. In particular the simulations were intended to gain an understanding of the interplay between the two maps in the superior colliculus, especially

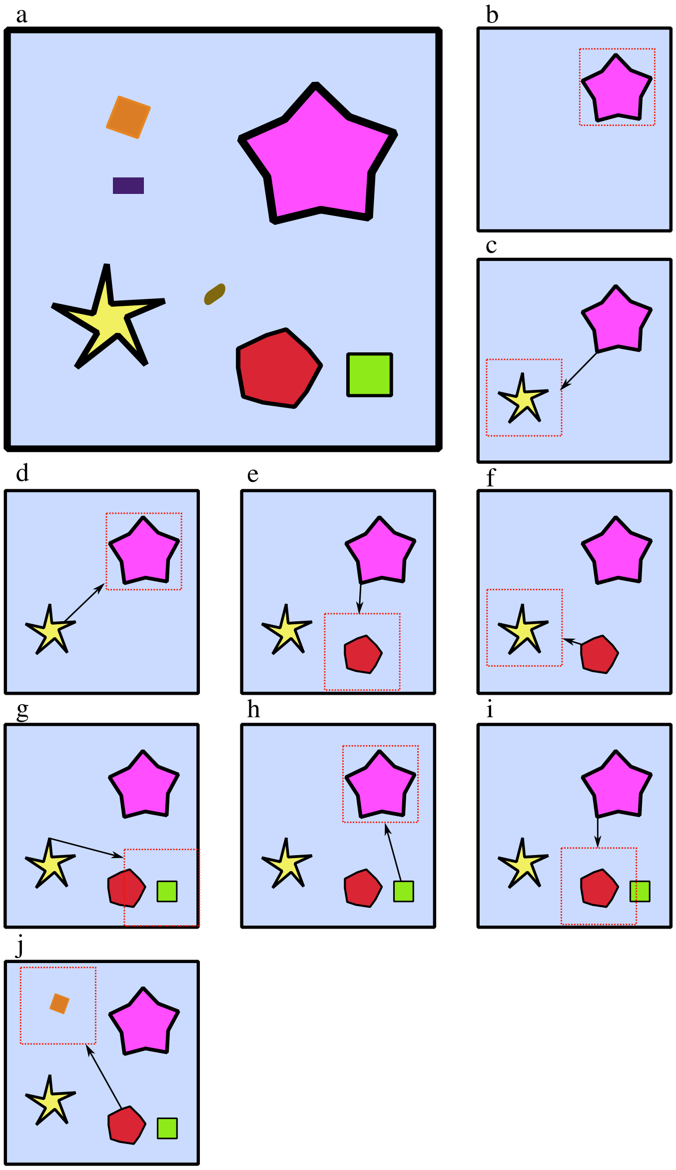

Fig. 4. The order of the attended objects showing the routing of the visual path. For subfigure a which shows an input image with seven objects in it. By assigning the reduction factor α = 0.17 the obtained visual foraging is shown by subfigures b, c, d, e, f, g, h, i and j. The red square shows the current attended object and arrow shows the path the visual awareness takes.

the comparison algorithm between the threshold map and the saliency map. Consequently, we limit our discussion only to the first nine iterations of the visual attention process.

The main objective of our experiment was to study the effect of the reduction factor α on the proposed system's behavior. A fixed input set of nine images was generated. Each had between one and nine objects randomly located in the 64×64 pixel visual scene. The reduction factor was assigned $\alpha \in [0, 1]$ with a step size of 0.01, to explore which value of the reduction factor α provided the optimum foraging coverage of the current scene. That procedure would result in around one hundred values of α per image. Finally, this proposed visual system was run for each value of α and the attended objects

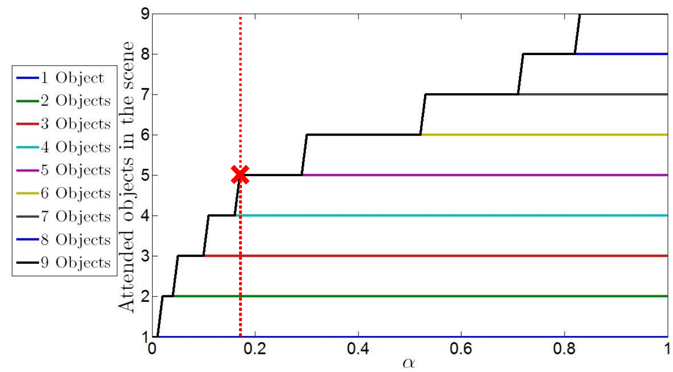

Fig. 5. Attended object coverage against α for between one and nine objects in the visual scene by setting $\alpha \in [0,1]$ in steps of 0.01. The red mark occurred when having seven objects in the visual scene and setting α to 0.17 as discussed in the results section.

are counted. This process was repeated for all nine images.

Fig. 4 illustrates an example of this experiment, with seven objects in the fixed visual scene. As our saliency is based on the size of the object only, the purple star on the upper right side of the Fig. 4a is the most interesting object. For this particular experiment, a reduction factor α of 0.17 was used. Our system successfully attended the purple star first, shown in Fig. 4b. This star was visited and hence is recorded as being attended at least once. The visual attention then moved on to the second most salient object, the yellow star as shown in Fig. 4c. The progression among the attended objects starting from the purple star and ended with the orange diamond, Fig 4b–i, was the route followed by the fixation point. The last step in the test is to count the number of the attended objects and compare it with the objects present in the scene. In the example shown, the system covered five objects out of seven.

Fig. 5 illustrates the relationship between the coverage of the attended objects with the reduction factor α. A red mark is shown in Fig. 5 as an indicator of the example discussed in relation to Fig. 4. The number of attended objects increases as α increases. When α reaches its maximum value (one), the system remembers every location it visited. Consequently, it will not revisit any locations and will spread the visual attention towards all objects. On the other hand the system will keep on staring at the most interesting object when α is zero. Under these conditions the system does not keep a record of past attended objects but remain fixated on the most salient object.

For the visual scene with the seven objects, the value of α that ensures full foraging coverage is 0.53 (see Fig. 5). Note that different values of α may result in different sight paths, even though the total coverage objects is the same.

VI. Discussion and Summary

This paper presented the inhibition of return architecture of our proposed system. This architecture is inspired by a biological system, with an artificial superior colliculus used to combine an overall saliency map and a threshold map. The superior colliculus is the main controller of saccade generation and is responsible for determining the target and timing of saccadic eye movements. The output of the superior colliculus is used by motor control units to control the fixation point of the visual system.

The architecture of a system for implementing an artificial inhibition of return mechanism with an adaptive threshold map was presented. The system depends on both saliency and threshold maps to attend novel areas. We have shown that it can alter the foraging in the available visual scene.

The effect of the reduction factor α was determined by studying the behavior for fixed artificial input maps. One hundred different values of α were examined in the range between zero and one. Low values of α correspond to a forgetful system that is willing to return to recently attended items, whereas high α leads to a system that is much more reluctant to return.

For each α value nine different maps were examined to determine the attention pattern of the system. As expected, a higher value of α led to higher coverage of the visual field. The simulations also revealed rich variations in the precise attention trajectory due to interactions between object saliency and threshold map decay.

Our computational adaptive dynamic interaction model is loosely based on the biological model and appears to provide the same basic functionality. In particular its behaviour can be altered by choosing an appropriate α value. For example a system required to perform different tasks might have an appropriate value of α for each task. An object tracking task would have a very different optimal α than a pattern discrimination task for example. However, in most biological cases, IOR memory has a limit of four or five objects/locations [7]. The results presented here suggest that $\alpha \in [0.11, 0.29]$ is a reasonable default range for a machine implementation intended to have a similar base performance.

We consider our model to have a wide application in image analysis such as in an active visual robotic system. We believe that the resulting computational model has applications mainly in the domain of autonomous machine vision, as well as the rapid selection of regions of interest in complex, cluttered visual environments.

References

[1] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision research*, vol. 40, no. 10-12, pp. 1489–1506, 2000.

[2] A. King, "The superior colliculus," *Current Biology*, vol. 14, no. 9, p. 335, 2004.

[3] L. Itti, "Models of bottom-up attention and saliency," *Neurobiology of attention*, vol. 582, 2005.

[4] J. Snyder and A. Kingstone, "Inhibition of return at multiple locations and its impact on visual search," *Visual cognition*, vol. 15, no. 2, pp. 238–256, 2007.

[5] Z. Wang and R. Klein, "Searching for inhibition of return in visual search: A review," *Vision research*, vol. 50, no. 2, pp. 220–228, 2010.

[6] M. Posner and Y. Cohen, "Components of visual orienting," *Attention and performance X*, vol. 5319556, 1984.

[7] S. Tipper, S. Grison, and K. Kessler, "Long-term inhibition of return of attention," *Psychological Science*, vol. 14, no. 1, p. 19, 2003.

[8] R. Klein, "Inhibition of return," *Trends in Cognitive Sciences*, vol. 4, no. 4, pp. 138–147, 2000.

[9] J. Pratt and R. Abrams, "Inhibition of return to successively cued spatial locations.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 21, no. 6, p. 1343, 1995.

[10] S. Tipper, B. Weaver, and F. Watson, "Inhibition of return to successively cued spatial locations: Commentary on pratt and abrams (1995).," 1996.

[11] M. Huelse, S. McBride, and M. Lee, "Implementing inhibition of return: embodied visual memory for robotic systems," 2009.

[12] D. Souto and D. Kerzel, "Evidence for an attentional component in saccadic inhibition of return," *Experimental brain research*, vol. 195, no. 4, pp. 531–540, 2009.

[13] B. Scassellati, *Foundations for a Theory of Mind for a Humanoid Robot*. PhD thesis, Citeseer, 2001.

[14] S. Palmer, *Vision science: Photons to phenomenology*, vol. 1. MIT press Cambridge, MA., 1999.

[15] C. Breazeal, A. Edsinger, P. Fitzpatrick, and B. Scassellati, "Active vision for sociable robots," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 31, no. 5, pp. 443–453, 2001.

[16] S. Zeki, "A vision of the brain," *Cambridge, MA*, 1993.

[17] W. Hall and A. Moschovakis, *The superior colliculus: new approaches for studying sensorimotor integration*, vol. 21. CRC, 2004.

[18] N. Port and R. Wurtz, "Target selection and saccade generation in monkey superior colliculus," *Experimental Brain Research*, vol. 192, no. 3, pp. 465–477, 2009.

[19] N. Gandhi, "Motor functions of the superior colliculus," *Annual Review of Neuroscience*, vol. 34, no. 1, 2011.

[20] W. Wong, "Design of a saccadic active vision system," 2008.

[21] T. Ayabe, T. Ishizu, S. Kojima, T. Urakawa, N. Nishitani, Y. Kaneoke, and R. Kakigi, "Neural processes of attentional inhibition of return traced with magnetoencephalography," *Neuroscience*, vol. 156, no. 3, pp. 769–780, 2008.

[22] J. Pratt and B. Neggers, "Inhibition of return in single and dual tasks: Examining saccadic, keypress, and pointing responses," *Attention, Perception, & Psychophysics*, vol. 70, no. 2, pp. 257–265, 2008.