Towards a Biologically Plausible Active Visual Search Model

Andrei Zaharescu, Albert L. Rothenstein, and John K. Tsotsos

Dept. of Computer Science and Centre for Vision Research,
York University, Toronto, Canada
{andreiz, albertlr, tsotsos}@cs.yorku.ca

Abstract. This paper proposes a neuronal-based solution to active visual search, that is, visual search for a given target in displays that are too large in spatial extent to be inspected covertly. Recent experimental data from behaving, fixating monkeys is used as a guide and this is the first model to incorporate such data. The strategy presented here includes novel components such as a representation of saccade history and of peripheral targets that is computed in an entirely separate stream from foveal attention. Although this presentation describes the prototype of this model and much work remains, preliminary results obtained from its implementation seem consistent with the behaviour exhibited in humans and macaque monkeys.

1 Motivation

Have you ever tried to imagine how would life be without moving our eyes? Most of the tasks that we perform in our everyday routine are greatly facilitated by our capacity to perform the fast eye movements known as saccades. Most simple actions, like finding the exit door, or locating the right button in the elevator, require a search of the visual environment and can be categorized as active visual search tasks, a search for given visual target that requires more than one eye fixation [1]. Several models ([2], [3], [4], [5]) [6, 7] have addressed aspects of attentive eye movements or active visual search with some success. At the present time, we are unaware of the existence of a computational biologically plausible model that compares favourably with human or primate active visual search performance. This contribution attempts to make significant inroads towards this goal. Although a general solution will be sketched, the demonstration is only for a limited case; the implementation of the full solution is in progress.

2 Background

Eye movements have been studied both psychophysically and neurophysiologically [8]. Several attempts have been made ([9], [10]) to formalize the performance characteristics of active search and to relate them to stimulus properties

L. Paletta et al. (Eds.): WAPCV 2004, LNCS 3368, pp. 133–147, 2005.

and cortical physiology. An outstanding accomplishment is the work of Motter et al. ([11][11], [12], [13]). They addressed the problem of active search for a target in a display of simple stimuli. The performance of monkeys on this task is documented and characterized. These characterizations form part of the basis for the model presented here. A second goal is to explore the use of the Selective Tuning Model (STM) [14] framework for this task, basing the work on the solution for one type of active fixation that was introduced in that paper.

2.1 Motters Experiments

Imagine the following experiment, shown in Fig. 1 below: the task is to locate a red bar oriented on a 135° angle (assume the trigonometric direction of the angle counterclockwise) from among a set of distracters oriented at 45° or 135°, coloured red or green. This experiment represents a conjunction search problem, where distracters and the target share common features [15]. The marked cross represents the initial point of fixation. The white rectangle denotes the portion of the overall scene that is visible within a single fixation. The remainder of the scene is depicted by the black rectangle that is not within the visual field of the subject. It can only be inspected by changes in fixation. Motter et al. ([11], [12], [13]) sought to discover the algorithm used by primates (macaque monkeys) to perform this task. As a result, they were able to characterize the solution used by macaques when performing active visual search tasks, the main components being summarized below: The probability of target detection falls off as a function of target eccentricity from fixation. The area within which targets are discovered is a function of the density of items around the target itself. In order for objects to be identified, some minimal distance must separate their representations in visual cortex. Therefore, because cortical magnification falls off as a function of eccentricity, the density of items around the target determines how close to the point of fixation the target must be before it can be identified. Color can be used to 'label' item relevancy. By knowing the probability of target detection as a function of eccentricity and assuming a random walk through relevant stimuli, they were able to account completely for search rate performance. No previous model of visual attention can account for this performance. The main goal of the research reported here is the attempt to cast such search characteristics into STM.

2.2 Selective Tuning Model

The Selective Tuning Model (STM) is a proposal for the explanation at the computational and behavioural levels of visual attention in humans and primates. Key characteristics of the model include:

1. A top-down coarse-to-fine Winner-Take-All selection process;
2. A unique Winner-Take-All (WTA) formulation with provable convergence properties;
3. A Winner-Take-All that is based on region rather than point selection;
4. A task-relevant inhibitory bias mechanism;

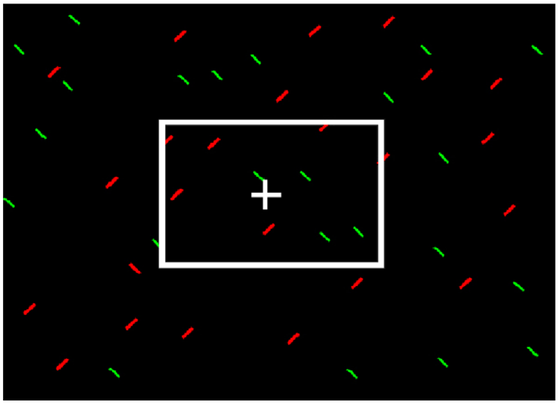

Fig. 1. Experimental setup

5. Selective inhibition in spatial and feature dimensions reducing signal interference leading to suppressive surrounds for attended items;
6. A task-specific executive controller.

The processing steps of the executive controller relevant for visual search tasks follow:

1. Acquire target as appropriate for the task, store in working memory;
2. Apply top-down biases, inhibiting units that compute task irrelevant quantities;
3. See the stimulus, activating feature pyramids in a feed-forward manner;
4. Activate top-down WTA process at top layers of feature pyramids;
5. Implement a layer-by-layer top-down search through the hierarchical WTA based on the winners in the top layer;
6. After completion, permit time for refined stimulus computation to complete a second feedforward pass. Note that this feedforward refinement does not begin with the completion of the lowermost WTA process; rather, it occurs simultaneously with completing WTA processes (step 5) as they proceed downwards in the hierarchy. On completion of the lowermost WTA, some additional time is required for the completion of the feedforward refinement.
7. Extract output of top layers and place in working memory for task verification;
8. Inhibit pass zone connections to permit next most salient item to be processed;
9. Cycle through steps 4 - 8 as many times as required to satisfy the task.

Greater detail on STM and its mechanisms may be found in ([14], [16], [17]).

3 The Active Visual Search Model

We have developed a general-purpose active visual search model that tries to incorporate most of the observations noted by Motter et al. within the STM framework. It will be presented next, beginning with a general overview followed by a detailed description. The basic aspects of the problem to be solved include: it is a search problem; it involves eye movements; the probability of target detection is directly related to the target eccentricity; saccadic eye movement is guided by item density; the cortical magnification factor needs to be taken into account; target characteristics can bias the search process; for the Motter task, saccadic eye movements are guided by colour, but not orientation.

3.1 Overview

Active visual search denotes a target location task that involves eye movements. In STM, feature pyramids are constructed through pyramidal abstraction and, as argued in [14], this abstraction leads to natural separation between a central region where stimuli are veridically analyzed and a peripheral one where they are not. The obvious solution to this is to include an active strategy for determining whether to attend to a stimulus in this central region or to a peripheral one, and if the latter, to initiate a fixation shift in order to bring that peripheral item into the veridical central region. Thus, STM incorporates two main search sub-processes: a central search and a peripheral search. It is noted that this addresses only one aspect of attentive fixation shifts. Each search sub-process is guided by a corresponding instance of the attention mechanism. Covert attention allows inspection of stimuli in the foveal region, trying to locate the target, while the overt counterpart guides saccadic eye movements towards the next fixation once the covert search strategy fails to locate the target. It does not mean that covert attention does not have access to the peripheral information, but rather that there is an independent overt mechanism that utilizes the peripheral information only. Is there a neural correlate to this separation? There is strong biological evidence for the existence of an area in the parieto-occipital (PO) of macaque that receives retinotopical input from V1, V2, V3, V4 and MT and that is sensitive to the regions outside the central 10-degrees of visual angle ([18]). The receptive fields are large an order of magnitude larger than the ones in V2. It is a hypothesis of this solution that area PO may have the function of detecting peripheral targets in this context. Due to the boundary problem resulting from limited size receptive fields and the resulting non-veridical analysis in the periphery [14], area PO must have access to features that have not been processed through many layers of the hierarchy (they would not be valid otherwise). Thus, PO receives input only from early features in this model.

The covert search part of the system has been discussed in detail by in previous work [14] and has been briefly mentioned in the background section (the executive controller). The novelty is introduced by the addition of a saccadic eye movement control system. From this point onward we will refer to it as the Overt Control System. Also, we will call a (Feature) Plane a population of neurons with similar function and behaviour. The Overt Control System interacts with the rest of the system in the following ways:

– It receives an Eye Movement Bias from the Task Controller;
– It receives input from both the Foveal Target Plane and the Peripheral Target Plane, which operate in a STM Winner-Take-All network;
– It outputs a request to the Eye Movement Controller for a saccade.

A system block diagram is shown in Fig. 2, placing the new component, the Overt Control System, in the context of the full model and its implementation. The components relevant to the active visual search task are elaborated in the remainder of the paper.

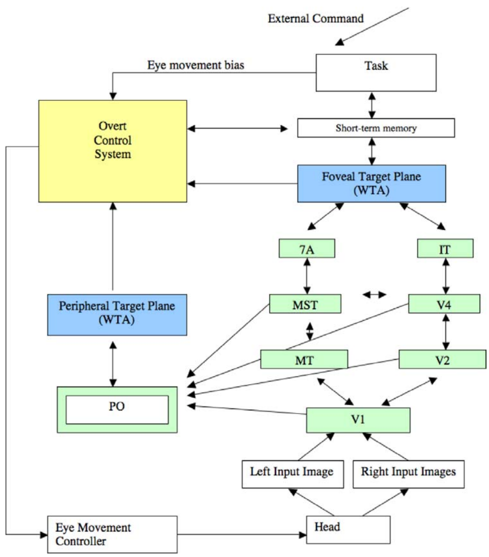

Fig. 2. General Overview of the Active Visual Search Model

3.2 The Overt Control System

In the current section we will describe the Overt Control System at a very general level, in terms of the interactions that take place among its components. Later we will provide a detailed description of the system tailored to address the limited task that is exemplified by Fig. 1. The Task Controller via the Eye Movement Bias triggers the Overt Control System, which is responsible for the choice of the most salient next saccade from the peripheral target plane. In doing so, it is important that it keeps track of previously inspected positions in the scene, since a return inspection would not yield to new information. This may be addressed by an inhibition of return mechanism (IOR; see [19]). IOR has been a staple of attention mechanisms since Koch & Ullman [20]. There, it was applied to a saliency map in order to inhibit a winner in a winner-take-all

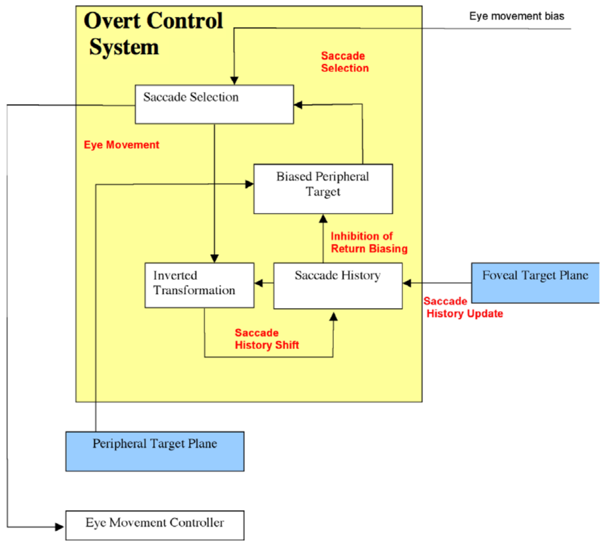

Fig. 3. Components of the Overt Control System

scheme, thus enabling the next strongest stimulus to emerge. Here, it must be in part a memory of what has been inspected and then used to inhibit a set of locations in order to avoid wasted fixations or oscillatory behaviour. The exact mechanism of IOR for saccadic eye movements is not currently known [19]. In the current model, we store the inhibition of return information in an entity named Saccade History. The Saccade History is constructed using a pool of neurons that encode position in the world; neurons will be on if they represent locations that have been previously visited for the task, and off otherwise. There are all reset to off at the end of the task. The saccade selection task from the peripheral target plane will be a basic winner-take all process selecting a winning location with all features competing equally. The feature locations are inhibitively biased by the locations in the Saccade History and indirectly by the concentration of distracters, due to RF size at the PO level. The Task Controller manages the overall behaviour (saccade selection) of the Overt Control System via an Eye Movement Bias, which influences equally in a multiplicative fashion all the participating neurons in the next saccade Winner Take All process. Thus, a high value of the bias would facilitate a new saccade, whereas a low value would prevent a new saccade from taking place. In case of a saccade winner, the command will be sent to the Head controller. Upon the selection of a winner, a new saccade will be issued. The Saccade History information is updated so that it integrates the current fixation point and also so that it maintains the information in a world system of coordinates that will be consistent with the new eye shift. It means that for each eye shift, there will be an associated saccade history shift. These basic components are depicted in Fig. 3. Each of the Overt Control System components will be discussed in greater detail in the next section, using the example task proposed in the opening section.

4 Simple Conjunction Search

This section will describe how the above framework can account for the active visual search performance observed by Motter and colleagues for a simple conjunction search as described in the Background. All the details of the framework will not be included here; only the elements directly involved in the solution of such tasks (Fig. 1) will be shown. Several other experimental results are currently being tested within the framework. The section begins with a brief description of the basic neural simulator that forms the substrate for the model implementation. The active visual search model has been implemented using TarzaNN, a General Purpose Neuronal Network Simulator [21] as a foundation. In TarzaNN, feature computations are organized in feature planes, interconnected by filters that define the receptive field properties of the neurons in the planes.

4.1 Eye Movement Controller

The current simulation implements a virtual eye movement controller, which only presents a part of the virtual 2D world to the system. In our case, referring to the setup of Fig. 1, the virtual world would represent the entire image, and the current input the area within the rectangle. An eye movement would correspond to a shift of the rectangle in the desired direction. This will eventually be replaced by a controller to a robotic stereo camera system.

4.2 The Input

The Foveal and Peripheral Target Planes are computed through a hierarchy of filters as shown in Fig. 2. The input image is first processed in order to incorporate the cortical magnification factor, the effect of large receptive fields and decreasing concentration of neural processing. There are mathematical descriptions to fit the biological data [22] that result in an increasing blur applied to the input image. In our model we have used the following two parameter complex log cortical magnification function (CMF):

$$CMF(z) = k\frac{log(z+a)}{log(z+b)}, a \cong 0.3, b \cong 50 \tag{1}$$

where z, a and b represent degrees of visual angle, and k a normalization factor (the maximum in our case). In the very center of the fovea there is no need to blur the original pixels, but as eccentricity (z) increases, the level of blurring will increase.

4.3 Feature Target Planes

In the implementation of the current example, we focused mainly on testing the validity of the Overt Control System, and not to implement a full-featured STM model, including all the areas described in the opening diagram (i.e. V1, V2, V4, IT, MT, MST, 7a; however see [14], [17] for several of these). We have modeled only a skeleton of feature planes sufficient to provide the Overt Control System

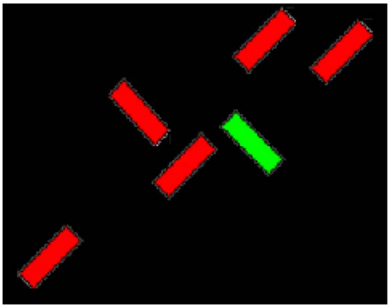

Fig. 4. Input Image - the fixated subset of a typical full scene

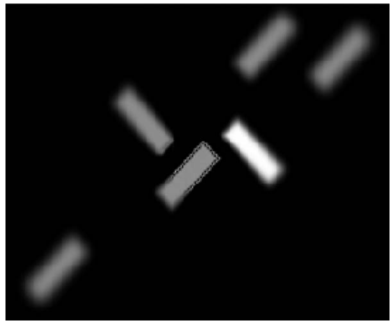

Fig. 5. Transformed input image (with the equivalent cortical magnification operator)

with the required information, namely the Foveal and Peripheral Target Planes, together with the eye movement bias. A port of the full featured STM model within TarzaNN is in progress. In order to focus this description on the Motter task, the only features discussed here will be colour and orientation. Hence, a sensible feature plane organization is the following:

- The foveal region hierarchy will have colour selective feature planes (R=r (g+b)/2, G=g-(r+b)/2, B=b-(r+g)/2), and for each colour, orientation selective feature planes (for 45° and 135°).
- The peripheral region will contain only colour selective feature planes.

In order to simulate the density distribution, which would occur normally at the top layers of a STM pyramid, a 2D Gaussian filter is applied. Please note that the current feature plane layout is tailored specifically for the above-mentioned experiment and it represents a very stripped-down version of realistic model (i.e. the peripheral region only receives the red-coloured filtered input image in reality, colour is coded in both fovea and para-fovea). Figures 4 through 10 show a feature plane walk-through starting with the sample input image. The captions provide the processing stage description in each case.

4.4 Saccade History

Earlier we introduced the idea of a Saccade History, to incorporate the task memory in the system. We will associate a feature plane with the concept. In the current simulation we are only considering a 2D version of the world, thus restricting the 3D space to a plane. However, at least at the hypothetical level, a generalization of the concept is immediately available for 3 dimensions.

The most efficient implementation of such a saccade history would be to only remember a list of the previously visited locations. However, it is not clear what a neural correlate of a list might be. Perhaps a more straightforward solution would be to model it as a scaled down version of the external world. This represents another hypothesis of the model, that a neural correlate of such a representation

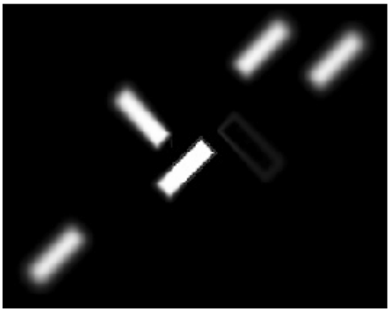

Fig. 6. R=r-(g+b)/2 colour filter plane

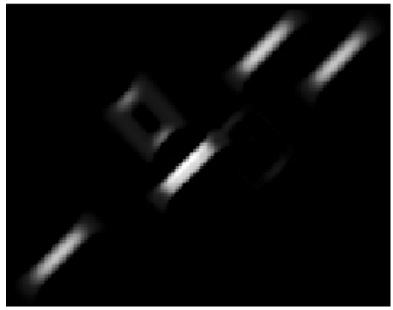

Fig. 7. 45° oriented region plane with R=r-(g+b)/2 plane as input

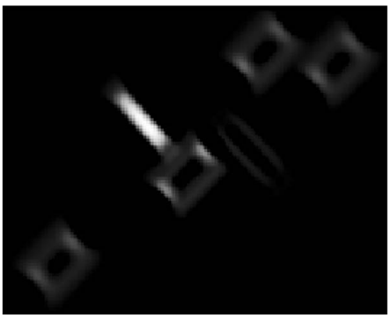

Fig. 8. 135° oriented region plane with R=r-(g+b)/2 plane as input

Fig. 9. Peripheral Region plane from R=r-(g+b)/2 colour filter plane

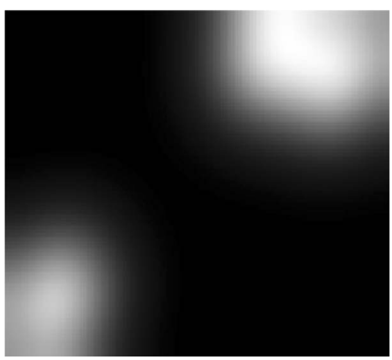

Fig. 10. Salient peripheral locations based on input as in figure 9 computing a stimulus density distribution using a Gaussian filter

exists with this functionality. The Saccade History encodes regions visited overtly by x/y location.

Assume the following Saccade History of previously executed saccades as shown in Figure 11. By convention, the centre of the map is at current eye fixation. A contour showing the currently fixation falls is depicted by a rectangle in Fig. 12. If a previous saccade location overlaps with a target, it is inhibited. For the initial Peripheral Target Plane, the Biased Peripheral Target will appear as shown in Fig. 13. The inhibition produced by the Saccade History plane is never total. Sometimes, distracters could be visited more than once in accomplishing a task [13].

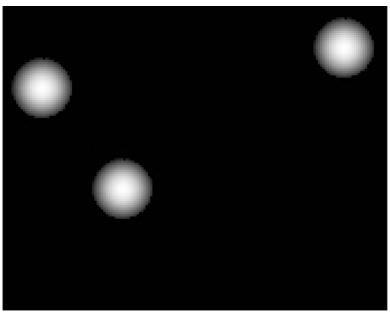

Fig. 11. Assumed Past Saccade History

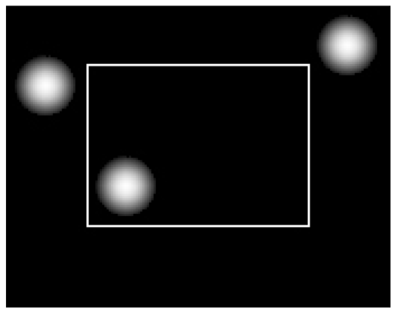

Fig. 12. Assumed Past Saccade History

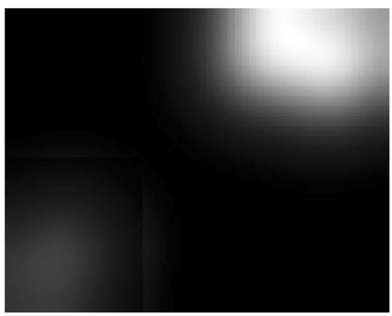

Fig. 13. Inhibition of return Biased Peripheral Target

4.5 Saccade History Feature Updating

The Saccade History Plane is updated to include the current fixation for each shift, as shown at the top of Fig. 14 . In this particular implementation of the system we have chosen to encode the location of fixation only. The properties of the attended stimulus (provided by the Foveal Feature Plane) are not taken into account. Whether location and feature qualities are both represented is unclear; there is little neurophysiological guidance on this currently. Further studies and experiments are needed in order provide an answer.

4.6 Saccade History Transformation

The center of the Saccade History Plane always encodes the current fixation. When a saccade is performed, the point of fixation moves in the world. In order to properly represent all previously fixated regions and maintain the world coordinates, a corresponding shift has to take place of all represented fixation locations (this may be related to the attentional shifts reported in [23] but it is too early at this stage to know). If we consider the saccade as a transformation given by the parameters (Tx,Ty), where these represent Euclidean distances in pixels for this example, then we need to perform the inverse of that transformation (Tx, Ty) to the Saccade History Plane. This is due to the fact that the world and the observer have systems of reference of opposite polarities. If the initial Saccade History Plane is filtered via a pool of neurons with receptive field properties described by a 2D Gaussian spread function shifted in the opposite direction of the saccade, we obtain the desired result. The winning saccade triggers the shifted corresponding inverted set of neurons, while inhibiting the others, thus ensuring that the appropriate transformation takes place. Biological evidence suggests that such a correspondence between the saccade target and a world transformation exists [24]. As more factors are taken into account, such as the 3rd dimension, head movements, and world movements, additional transformations will be needed to maintain the saccade history, using essentially the same mechanism. Pouget et al. [25] provide a detailed biological implementation of coordinate transformations that we are considering including in a future version of the system.

4.7 Saccade Selection

The most salient feature in the biased peripheral feature plane encodes the coordinate of the next saccade. This will be established via a Winner Take All process operating on the biased peripheral features selecting the strongest responding one.

As the Saccade History Plane is shifted, some information is discarded, since data is shifted outside the boundaries of the space, as illustrated in Fig. 14. The coordinates of the next saccade are sent to the Eye Movement Controller.

5 Results

The overall system generated very satisfying results, in accordance with the goals. Fig. 15 presents the result of an average trial (in terms of number of fixations). We present the results of the same search task when the there is no colour discrimination and the inhibition of return mechanism is turned off. As it can be observed in Fig.16, the inhibition or return mechanism is crucial for the task, so that the process will not get into an infinite loop. We show another run where the next saccade fixation point is chosen randomly. As it can be seen in Fig. 17, it took over 200 trials to obtain the same result. The performance of the system is dictated by some of the system control variables. They are listed below,

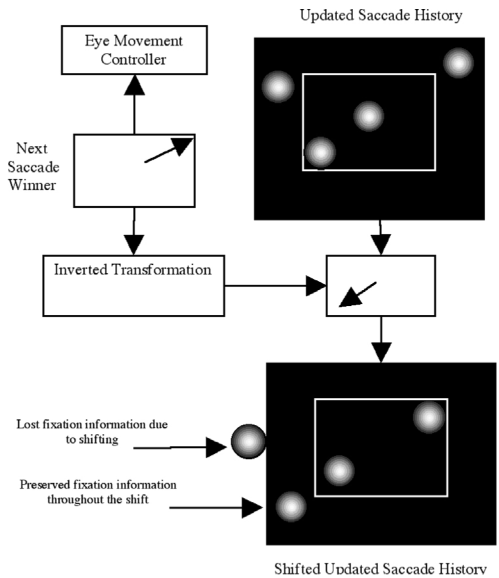

Fig. 14. Saccade History Shifting. This example illustrates how, upon a Saccade History shift, certain previously visited locations information can get lost

together with the relationship in which they influence the overall performance of the system:

- The size of the window into the virtual world (directly proportional);
- The accuracy of the next fixation in terms of selecting the region with the most distracters (inversely related to the size and spread of the 2D Gaussian filter as illustrated in Fig.10);
- The size of the inhibited region in the Saccade History Plane (directly proportional);
- The amount of inhibition in the Saccade History Plane (directly proportional).
- The ratio between dimensions of The Saccade History Plane and the Input Image

The probability of target detection is indeed directly related to the target eccentricity, and that follows as a direct consequence of the fact that cortical magnification is taken into account, and the solution uses neuronal response thresholding for detection. The current system represents our first approach to the problem, and at the present time we cannot provide a detailed analysis of its performance with regards to real test data. Relevant statistical data needs to be generated with a large set of example cases in order to be able to plot it against the graphs provided by Motter et al. for direct comparison.

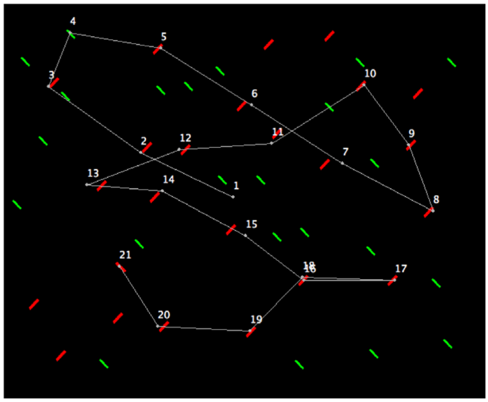

Fig. 15. Sample Run

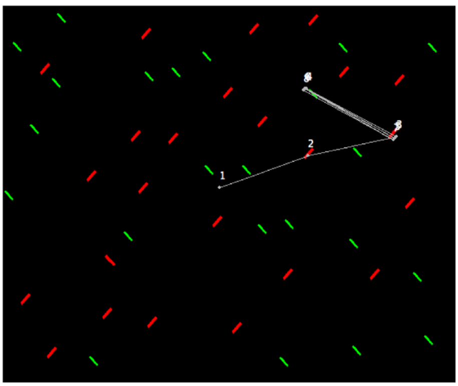

Fig. 16. Example without Colour Discrimination and Inhibition of Return

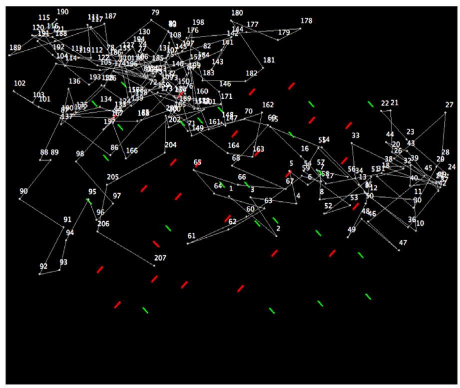

Fig. 17. Sample Run using a random saccade fixation

6 Conclusions

The system represents our first approach to solve the active visual search problem in a biologically plausible fashion. It integrates findings about the specifics of the task and about biological constraints. The proposed solution to active visual search contains several novel components. The newly added Overt Control System is capable of guiding the saccadic eye movements taking into account peripheral distracter characteristics and by incorporating inhibition of return information. The general system can solve a whole class of active visual search problems as long as the appropriate Peripheral Target and a Foveal Target Plane are provided (see Fig.2). The Overt Control System represents and important milestone in the evolution of the Selective Tuning Model, exposing it to a whole new set of problems and challenges From the biological standpoint, the current implementation of the system exhibits a certain weakness in the assumption of the 2D structure of the Saccade History Plane. We are not claiming that there exists a one-to-one correspondence between the model proposed and a biological hierarchy of neuronal constructs. However, from a functional standpoint, a similar mechanism needs to encode the previously visited locations of the world and store them in some type of memory (short term, or more specialized location based memory/world map). Further research will need to address this outstanding issue.

Acknowledgements

This research was funded by the Institute for Robotics and Intelligent Systems, a Government of Canada Network of Centres of Excellence and by Communication and Information Technology Ontario, a Province of Ontario Centre of Excellence. JKT holds the Canada Research Chair in Computational Vision.

References

1. Tsotsos, J.K.: On the relative complexity of passive vs active visual search. International Journal of Computer Vision **7** (1992) 127 – 141
2. Wolfe, J.: Guided search 2.0. a revised model of visual search. Psychonomic Bulletin & Review **1** (1994) 202–238
3. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research **40** (2000) 1489–1506
4. Rao, R.P., Zelinsky, G.J., Hayhoe, M.M., Ballarad, D.H.: Eye movements in iconic visual search. Vision Research **42** (2002) 1447–1463
5. Hamker, F.M.: The reentry hypothesis: linking eye movements to visual perception. Journal of Vision **3** (2003) 808–816
6. Lanyon, L.J., Denham, S.L.: A model of active visual search with object-based attention guiding scan paths. Neural Networks **in press** (2004)
7. Lanyon, L.J., Denham, S.L.: A biased competition computational model of spatial and object-based attention mediating active visual search. Neurocomputing **in press** (2004)
8. Carpenter, R.: 8, Vision and Visual Dysfunction. In: Eye Movements. Volume 8. J. Cronly-Dillon (ed.), MacMillan Press (1991) 95–137
9. Triesman, A., Sato, S.: Conjunction search revisited. J. Experimental Psychology: Human perception and Performance **16** (1990) 459–478
10. Lennie, P., Trevarthen, C., van Essen, D., Wassle, H.: Parallel processing of visual information. San Diego: Academic Press (1990)
11. Motter, B., Belky, E.J.: The zone of focal attention during active visual search. Vision Research **38** (1998) 1007–1022
12. Motter, B.C., Belky, E.J.: The guidance of eye movements during active visual search. Vision Research **38** (1998) 1805–1815
13. Motter, B.C., Holsapple, J.W.: Cortical image density determines the probability of target discovery during active search. Vision Research **40** (2000) 1131–1322
14. Tsotsos, J.K., Culhane, S., Wai, W., Lai, Y., Davis, N., Nuflo, F.: Modeling visual attention via selective tuning. Artifical Intelligence **78** (1995) 507–547

15. Treisman, A., Gelade, G.: A feature integration theory of attention. Cognition Psychology **12** (1980) 97–136
16. Tsotsos, J.K.: Analyzing vision at the complexity level. Behavioural Sciences **13** (1990) 423–445
17. Tsotsos, J.K., Liu, Y., Martinez-Trujillo, J., Pomplun, M., Simine, E., Zhou, K.: Attending to visual motion. submitted (2004)
18. Galletti, C., Gamberini, M., Kutz, D.F., Fattori, P., Luppino, G., Matelli, M.: The cortical connections of area v6: an occipito-parietal network processing visual information. European Journal of Neuroscience **13** (2001) 1572–1588
19. Klein, R.M.: Inhibition of return. Trends in Cognitive Sciences **4** (2000) 138–147
20. Koch, C., Ullman, S.: Shifts in the selective visual attention: Towards the underlying neural circuitry. Human Neurobiology **5** (1985) 219–227
21. Rothenstein, A.L., Zaharescu, A., Tsotsos, J.K.: TarzaNN: A general purpose neural network simulator for visual attention modeling. In: Workshop on Attention and performance in computational vision. (2004)
22. Schwartz., E.L.: Topographic mapping in primate visual cortex: Anatomical and computational approaches. Visual Science and Engineering: Models and Applications **43** (1994)
23. Duhamel, J.R., Colby, C.L., Goldberg, M.E.: The updating of the representation of visual space in parietal cortex by intended eye-movements. Science **255** (1992) 90–92
24. Fukushima, K., Yamanobe, T., Shinmei, Y., Fukushima, J.: Predictive responses of periarcuate pursuit neurons to visual target motion. Exp. Brain Res (2002) 104–120.
25. Pouget, A., Sejnowski, T.J.: Spatial transformations in the parietal cortex using basis functions. Journal of Cognitive Neuroscience **9** (1997) 222–237