Spatial Attention and Saccadic Camera Motion

James J. Clark
Centre for Intelligent Machines, McGill University

Abstract

An important aspect of computer-controlled camera motion systems is that of the generation of saccadic movements, which shift the camera gaze quickly from one fixation position to another. Recent psychophysical experiments suggest that there exists a causal connection between spatial shifts in visual attention and the production of saccadic eye movements in humans. Motivated by this experimental evidence, we propose a winner-take-all based model of exogenous spatial attention, involving both sustained and transient feature detection channels, and link it to the targetting and triggering of saccadic eye movements. We show that this model accounts for a range of oculomotor phenomena observed in human subjects. We describe the application of this model to a robotic camera gaze control system.

1 Introduction

There is an increasing number of computer-controlled camera positioning systems appearing in university and industrial machine vision research laboratories [6]. Much effort has been expended in developing motion control algorithms which allow these systems to perform smooth pursuit, gaze stabilization and tracking behaviours [4]. Less has been published, however, on the generation of rapid, or *saccadic* camera motions that direct gaze from one target to another. This is, perhaps, due to the perceived simplicity of saccade generation as compared with the control of tracking motions. In one respect, the generation of saccadic motions is more difficult than that of the generation of tracking motions. This difficulty lies in the determination of the target of the motion. In the case of tracking motion, the target is the object being tracked, which is presumably the object in the center of the visual field. There are many possible saccadic targets, however. The process of saccadic target selection, sometimes referred to as the "next-look" process, is known to be one of the critical aspects of

active vision systems [22]. In this paper, we provide a model of the process by which the "next-look" is determined. This model is motivated by a growing number of psychophysical experiments involving human subjects which support the view that there is a direct connection between saccadic eye movements and visual spatial attention (see, for example, Henderson [15]). In particular it appears that the parameters and the timing of saccadic eye movements are determined by visual attention. It is thought that the targets of saccadic eye movements are defined by the spatial locus of attentional enhancement immediately prior to the motion.

Our approach to saccade generation combines a winner-take-all based attentional subsystem, similar to that proposed by Koch and Ullman [16], a distributed population coding of the saccade target location, and a simple hysteretic mechanism for triggering the saccade motion. Existing computational models of spatial attention, such as that of Tsotsos [23], do not provide a direct connection between attention and eye movements. Our approach considers this connection as fundamental. In this paper we describe how our model can account for a wide range of human oculomotor phenomena, as detailed in [7]. In addition, we will demonstrate that the model can serve as the basis for a system for generating saccadic camera movements in robotic systems.

2 A Model of Visual Attention and Saccadic Eye Movements

In this section we present the details of our model of visual attention and its link to saccadic eye movements. It brings together a number of ideas found in the neuroscience literature. The critical aspects of the model are:

- Winner-take-all interaction between elements of a spatial saliency map (Koch and Ullman [16]).

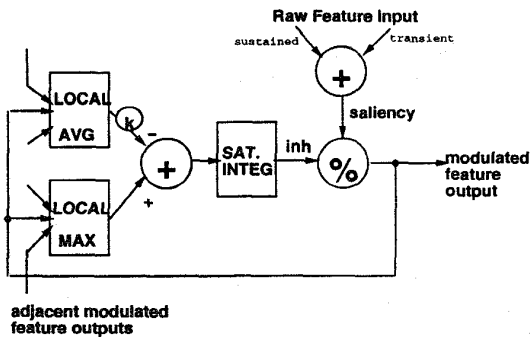- Spatio-temporal integration for target specification [10].

Figure 1: A schematic view of our model of spatial visual attention. One cell in a 2-D array of such cells is shown.

- Distributed representation of target location [17].

- Triggering of saccades based on transitions of the winner-take-all network.

2.1 Model of Attention

Our model of attention is depicted in schematic form in figure 1. In this model feature maps of various kinds are computed and combined into a "saliency" map. Different features can be weighted by different amounts in producing this saliency map (as in the Koch and Ullman model [16]). This is a commonly used technique for determining the target of an attentional shift (see, for example, [18]). There are two types of feature detectors posited in this model, transient and sustained, as found in mammalian visual systems [5]. The transient feature detectors are fast responding but have relatively low spatial resolution. The sustained feature detectors are slower to respond but have higher spatial resolution. The outputs of these feature detectors are modulated via a division by an inhibitory signal. The inhibition signal at each location in the feature map is produced by a saturating, leaky, integrator. This integrator temporally accumulates the difference between the local spatial average of the modulated feature output at that location (weighted by a factor $k > 1$) and a local estimate of the maximum modulated feature output. If the average modulated feature output at the location (weighted by k) is greater than the local maximum then the integrator will discharge, reducing the inhibition. If, on the other hand, the modulated feature output is less than the local maximum value, the integrator will be-

gin to charge, increasing the inhibition. This positive feedback results in a winner-take-all process, wherein the location with the locally greatest feature detector activity will inhibit its neighbors, reducing their activity even further and therefore strengthening its hold. When the input feature activity changes, the winner-take-all network will take on a new equilibrium, with new locations being inhibited and a new winning location established.

2.2 Targetting and Triggering of Saccades

One of the key aspects of our model is that shifts in attentional activity give triggers saccadic movements and that the locus of attentional activity at the time of triggering specifies the target of these motions. This tight connection between the targetting and triggering of saccadic camera movements is not present in previous theories of attention, such as that of Tsotsos [23].

The process of specifying the position of the target of a saccadic eye movement has often been referred to in the eye movement literature as "saccadic programming" (see, for example, [1, 2, 12, 14]). This programming process has typically been viewed as consisting of two components, amplitude programming and direction programming. There are a number of experiments that appear, on the surface, to support the saccadic programming viewpoint, such as the amplitude or direction precuing experiments of Abrams [1] and the "double-step" experiments of Becker and Jurgens [3].

It is our view that explicit saccadic programming of either saccade amplitude or direction is unneccessary. Rather, the pattern of the low level feature detector activity, as modulated by visual attention, is used to determine both the timing and target of the saccadic eye movement. Following the proposal of Lee, Rohrer, and Sparks [17], our model assumes that the command for the saccadic eye movement is coded in a distributed fashion by a population of neurons, whose activity is attentionally modulated. Thus the target will be specified by the *center of mass*, or centroid, of this pattern of activity. This motor target is continually available. No saccadic "programming" need take place. The amplitude and direction of saccades are implicit in the target locations. We have shown, in simulations of our model, that our view is sufficient to explain the results of the Abrams and the Becker and Jurgens experiments. Details can be found in [7].

In our model, a saccade is *triggered* when the level of inhibition at any location drops from its maximum value to its minimum value. The dynamics of the winner-take-all are such that the values of the inhibi-

tion signal will, in steady state, be either exactly the minimum value or exactly the maximum value. The requirement that the inhibition signal drop from maximum to minimum at a given location before a saccade is made to that location, provides a hysteresis. That is, a location that was previously fixated must become inhibited before a saccade can be made to it. Note that the inhibition dropping to its minimum value at a location only *triggers* the saccade. Depending on the feature activity elsewhere, the *target* of the saccade may not be that particular location. In single target cases the saccade will usually be made to the location which triggers the saccade, however. Using the terminology of Posner [20], our model states that a saccadic eye movement is triggered when attention is "engaged" at a new location. This is in contrast to models, such as that proposed by Fischer [13], in which eye movements are triggered when attention is "disengaged" from its current location.

3 Comparison with Human Oculomotor Behaviour

In this section we discuss the behaviour of our model in terms of saccadic latency and spatial accuracy of the motion relative to the target, and compare these behaviours to the observed in human subjects. Saccadic movements can be characterised by their timing relative to the appearance of a target, and by the location of the endpoint of the movement trajectory relative to the target. In studies of saccadic motions the timing is often expressed in terms of the *saccadic latency*, which is the length of time between the appearance of the target and the time at which the eye begins to move.

3.1 The Gap Effect

Saslow [21] observed that saccadic latencies were reduced when the temporal gap between the offset of the fixation stimulus and the onset of the target stimulus was increased. This phenomenon has come to be known as the *gap effect*. Furthermore, it was observed that saccadic latencies are increased when there exists a temporal overlap between the fixation offset and the target onset.

The gap effect is readily apparent in simulations of our model. The time units for the simulation are arbitrary, and so no absolute comparisons of the simulation results with observed data can be made. In figure 2 we show the saccadic latencies produced by our
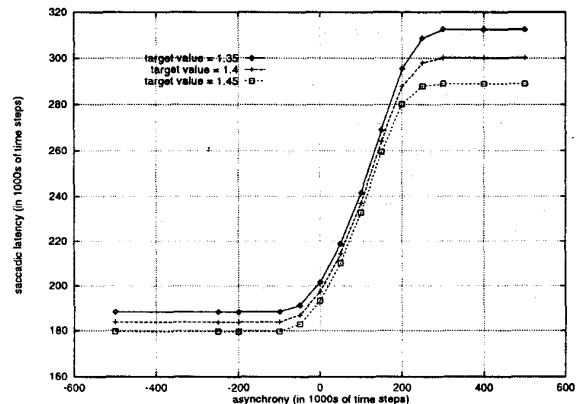

Figure 2: Simulations of the model exhibiting the gap/overlap effect for various values of the saliency of the target stimulus.

model for a range of stimulus asynchronies and relative target-fixation stimulus saliencies. The shape of the curves closely follow the form observed by Saslow [21] and Reulen [19] for human subjects.

Note that, in both the gap and overlap conditions, in our model the relative saliency of the target and fixation stimuli affects the saccadic latency. This effect has been observed in human subjects in many studies [25, 24]. These studies show that saccadic latencies are reduced when the salience (e.g. luminance) of the target stimulus is increased.

3.2 Modulation of the Global Effect

Coren and Hoenig [9] observed that the amplitudes of saccades to point targets can be systematically affected by the presence of distractors. Saccades tend to bring the eye to the "center-of-gravity" of the target+distractor complex. This phenomenon, called the *global effect* by Findlay [11], has been observed in many other experiments and with various stimulus configurations. Coeffe and O'Regan [8] performed a set of experiments which showed that the global effect can be modulated by varying saccadic latency (by voluntary control, for example) and by controlling the predictability of the target location. In particular, long latency saccades were found to be more accurate than short latency saccades.

We replicated (in format, if not in detail) the experiment of Coeffe and O'Regan using our model of saccade generation. In their experiment, subjects would fixate on a cross presented in the middle of the vi-
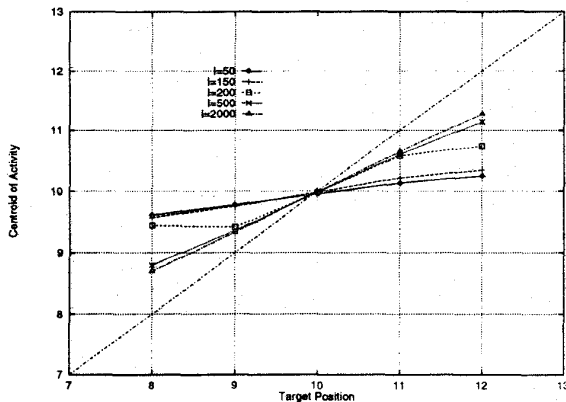
Figure 3: Simulations of the model showing that increasing saccadic latency reduces the global effect. Each curve represents a different saccadic latency. Latency values are given in 1000's of time steps in the legend. The vertical axis is the centroid of the modulated feature activity, and is taken as the saccade command position.

sual field. Then a brief cue would be flashed on at a location in the periphery, followed by a string of 10 letters. The subjects were required to report the letter at the cue location. The subjects needed to move their eyes to do this task successfully. Coeffe and O'Regan measured the landing position of the eye and plotted it against the position of the cue. They found that when saccades had long latencies the eye landed closer to the cue location than when the latencies were short. For short latencies the eye would undershoot targets towards the end of the letter string and overshoot targets towards the beginning of the letter string.

As can be seen in figure 3, our simulations show the same behavior as observed by Coeffe and O'Regan. The overshoot of near targets and the undershoot of far targets is seen to decrease as latencies increase. This is due to two effects in our model. The first is the transient and sustained components of the feature detectors. For short latencies the feature detector response is dominated by the transient component which has a low spatial frequency cutoff, effectively blurring the target and distractors together. At longer latencies the sustained component dominates, which has a higher spatial frequency cutoff, and hence creates less blurring of the target and distractors. The second factor contributing to the dependence of feature centroid on latency is the action of the winner-take-

all network. Increasing the latency allows more time for the ultimate winning feature location to suppress its neighboring distractors, thus reducing the effect of the spatial blurring. In our simulation, the first factor dominates at longer latencies (after the peak of the transient response, which occurs at around 500×10^3 time steps in our simulation), and the second factor dominates at short latencies.

4   Robot Psychophysics

We have implemented our model of attention driven saccade generation on a real-time robot vision system. This system consists of a Panasonic WVCP410 colour video camera, mounted on a Directed Perception PTU-46 pan-tilt unit. The video data is digitized with a Matrox Meteor digitizer board, and the computations are done on a 200 MHz Pentium based computer. The attention and motion control algorithms were implemented in Visual C++.

We have not yet carried out experiments replicating the "gap effect" or the "global effect", but we have performed a basic experiment that demonstrates the effect of target saliency on saccadic latency. In this experiment the camera viewed a scene consisting of a board on which is placed four patches of colour, one yellow, one red, one green and one blue patch. An example of such a view is shown in figure 4. The camera image was first converted into a foveal format, where resolution is highest in the centre of the frame and decreases towards the periphery of the image frame. This was done to reduce computational requirements. For the purposes of this experiment, saliency was defined as a combination of colour saturation and proximity to a selected hue.

Given a particular target hue, the attentional algorithm will select the image location that maximizes the proximity to this hue and the color saturation level. A saccadic movement of the camera to that location will be triggered once the winner-take-all process switches. We then redefine the desired hue via an external input from the computer keyboard. This alters the attentional landscape, resulting in a switching of the winner-take-all, and causing the camera to move to direct its gaze to another location.

The time elapsed between the changing of the saliency and the triggering of the camera motion, i.e. the saccadic latency, depends on the saliency of the new target. This is shown in figure 4, which plots saccadic latency versus saliency values, averaged over 10 runs for each of the four types of desired hue change
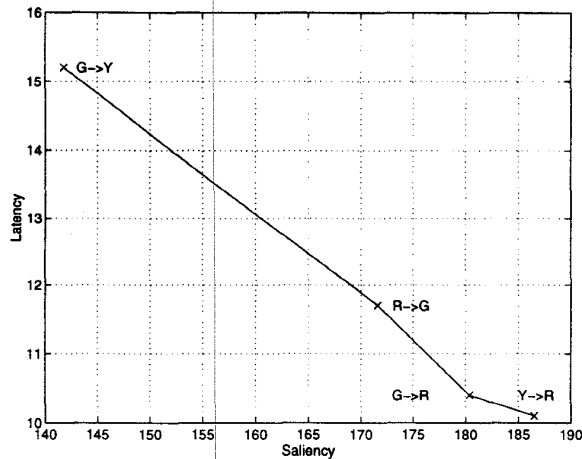
Figure 4: Saccadic latencies of the robot camera positioning system, as a function of target saliency. Latency values are measured in video frame periods (30 msec).

tested in the experiment. The time scale for the latency values is in terms of video frame times (i.e. 30 msec). The latency is a function of the time constant of the integrator in the winner-take-all network. In figure 4, the label $R \to G$ indicates a situation where the desired hue was changed from Red to Green. Similar interpretations can be applied to the other labels. Of the four colour patches, the red patch had the highest saliency (when its hue was selected), followed by green, then yellow. The blue patch had the lowest saliency, and in fact, saccadic movements could never be triggered to the blue patch. Note that, in figure 4, the longest latency was produced in moving to the yellow patch, while the shortest latency was produced in moving to the red patch. There is an obvious correlation between the saccadic latency and the saliency of the target, just as is observed in human subjects.

5 Summary and Conclusions

In this paper we have presented a simple model for the generation of saccadic camera movements. The viewpoint espoused in our model differs fundamentally from most existing models of saccadic motion generation. Its principal aspects are that

- The targets of saccades are not "programmed" by any modular process, but are continuously de-

fined by the pattern of activity of the attentionally modulated feature values.

- Saccades are triggered when attention is "engaged" at a new location (as indicated by the attentional inhibition reducing to a minimum at the new location).

We showed, via computer simulation of a simplified instantiation of our model, that our model can replicate a wide range of oculomotor behaviors observed in human subjects, such as the gap effect, the global effect and the variation of saccadic latency with target saliency. In replicating these phenomena with our model, we are able to provide insights as to their underlying mechanisms.

We also applied our model to the generation of saccadic camera motions in a robotic system. This system exhibited a correlation between saccadic latency and target saliency similar to that observed in human subjects engaged in similar tasks.

The model presented here is computationally simple, and the sub-systems required by the model, namely the local averaging, local maximum, temporal integration, and division processes, can be implemented in analog VLSI technology. This implies that our model is ideally suited to implementation in a real-time robotic active vision system.

Acknowledgements

References

[1] Abrams, R.A. (1992), "Planning and producing saccadic eye movements", in **Eye Movements and Visual Cognition**, Rayner, K. ed., Springer-Verlag, pp 66-88

[2] Abrams, R.A. and Jonides, J. (1988), "Programming saccadic eye movements", *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 14, pp 428-443

[3] Becker, W. and Jurgens, R. (1979) "An analysis of the saccadic system by means of double step stimuli", *Vision Research*, Vol. 19, pp 967-983

[4] Blake, A. and Yuille, A. (eds.) (1992), **An Introduction to Active Vision**, MIT Press, Cambridge, Massachusetts.

[5] Breitmeyer, B. and Ganz, L. (1976), "Implications of sustained and transient channel for theories of visual pattern masking, saccadic suppresion and information processing", *Psychology Review*, Vol. 83, pp 1-36

[6] Christensen, H.I., Bowyer, K.W., and Bunke, H. (1993), **Active Robot Vision: Camera Heads, Model Based Navigation and Reactive Control**, World Scientific, Singapore

[7] Clark, J.J. (1998), "Dynamics of spatial attention and saccadic eye movements", accepted for publication in *Vision Research*

[8] Coeffe, C. and O'Regan, J.K. (1987), "Reducing the influence of non-target stimuli on saccade accuracy: predictability and latency effects", *Vision Research*, Vol. 27, No. 2, pp 227-240

[9] Coren, S. and Hoenig, P. (1972), "Effect of non-target stimuli upon length of voluntary saccades", *Perceptual Motor Skills*, Vol. 34, pp 499-508

[10] Deubel, H., Wolf, W., and Hauske, G. (1984), "The evaluation of the oculomotor error signal", in **Theoretical and Applied Aspects of Eye Movement Research**, Gale, A.G. and Johnson, F. (eds.), Elsevier Science Publishers B.V., pp 55-62

[11] Findlay, J.M. (1982), "Global visual processing for saccadic eye movements", *Vision Research*, Vol. 22, pp 1033-1045

[12] Findlay, J.M. (1992), "Programming of stimulus-elicited saccadic eye movements", in **Eye Movements and Visual Cognition**, Rayner, K. ed., Springer-Verlag, pp 31-45

[13] Fischer, B. (1993), "Express saccades and visual attention", *Behavioral and Brain Sciences*, Vol. 16, pp 553-610

[14] He, P. and Kowler, E. (1989), "The role of location probability in the programming of saccades: Implications for Center-of-Gravity tendencies", *Vision Research*, Vol. 29, No. 9, pp 1165-1181

[15] Henderson, J.M. (1992), "Visual attention and eye movement control during reading and scene perception" in **Eye Movements and Visual Cognition**, Rayner, K. ed., Springer-Verlag, pp 260-283

[16] Koch, C. and Ullman, S. (1985), "Shifts in selective visual attention: Towards the underlying neural circuitry", *Human Neurobiology*, Vol. 4, pp 219-227

[17] Lee, C., Rohrer, W.H., and Sparks, D.L. (1988), "Population coding of saccadic eye movements by the superior colliculus", *Nature*, Vol. 332, pp 357-359

[18] Milanese, R. (1991), "Detection of salient features for focus of attention", *Proceedings of the 3rd Meeting of the Swiss Group for Artificial Intelligence and Cognitive Science*, World Scientific Publishing, October 1991

[19] Reulen, J.P.H. (1984), "Latency of visually evoked saccadic eye movements. I. Saccadic latency and the facilitation model", *Biological Cybernetics*, Vol. 50, pp 251-262

[20] Posner, M.I., Cohen, Y., and Rafal, R.D. (1982), "Neural systems control of spatial orienting", *Philosophical Transactions of the Royal Society, London B*, Vol. 298, pp 187-198

[21] Saslow, M.G. (1967), "Effects of components of displacement-step stimuli upon latency for saccadic eye movement", *Journal of the Optical Society of America*, Vol. 57, pp 1024-1029

[22] Swain, M. and Stricker, M. (eds.) (1991), "Promising Directions in Active Vision", University of Chicago Technical Report, CS 91-27.

[23] Tsotsos, J.K. (1990), "Analyzing vision at the complexity level: constraints on an architecture, an explanation for visual performance, and computational justification for attentive processes", *Behavioral and Brain Sciences*, Vol. 13, No. 3, pp 423-468

[24] Unema, P.J.A. (1995), **Eye Movements and Mental Effort**, Dr. Phil. Dissertation, Technischen Unversitat Berlin, Verlag Shaker Aachen

[25] Wheeless, L.L., Cohen, G.H., and Boynton, R.M. (1967), "Luminance as a parameter of the eye movement control system", *Journal of the Optical Society of America*, Vol. 52, pp 210-213