Directing Attention to Onset and Offset of Image Events for Eye-Head Movement Control

Winky Y. K. Wai and John K. Tsotsos
Department of Computer Science, University of Toronto
Toronto, Ontario, Canada M5S 1A4,

Abstract

This paper proposes a model that investigates a new avenue for attention control based on dynamic scenes. We have derived a computational model to detect abrupt changes and have examined how the most prominent change can be determined. With such a model, we explore the possibility of an attentional mechanism, in part guided by abrupt changes, for gaze control.

The computational model is derived from the difference of Gaussian (DOG) model and it examines the change in the response of the DOG operator over time to determine if changes have occurred. On and off-DOG operators are used to detect "on" and "off" events respectively. The response of these operators is examined over various temporal window sizes so that changes at different rates can be found. The most salient "on" and "off" events are determined from the corresponding winner-take-all (WTA) network. The model has been tested with image sequences which have changes caused by brightness or motion and the results are satisfactory.

1 Introduction

As solving computer vision problems always involve a huge amount of computation, an attentional mechanism is necessary in any computer vision systems that hope to have real-time performance [8]. Recently, Yantis and several co-authors revealed from some psychological experiments that the abrupt appearance of an object in the visual field draws visual attention, e.g., [11], [12], [13], [14], [15]. Inspired by this novel idea of attentional capture, we derive a computational model to find abrupt changes from a sequence of images. The output of such a model will be useful for a vision system in which attention is guided by abrupt changes.

2 The DOG model

In order to find changes in a sequence of images, we consider the response of applying a difference of Gaussian (DOG) operator to the raw image. This is to allow our computational model to have as much biological resemblance as possible, and to reduce the effect of noise. The DOG model is composed of the difference of two impulse response functions that model the centre and surround mechanisms of retinal cells. Mathematically, the DOG operator is defined as

$$DOG(\vec{x}) = \alpha_c G(\vec{x};\sigma_c) - \alpha_s G(\vec{x};\sigma_s)$$

where G is a two-dimensional Gaussian operator at $\vec{x}$:

$$G(\vec{x};\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{|\vec{x}|^2}{2\sigma^2}} \quad (1)$$

Parameters σ_c and σ_s are standard deviations for centre and surround Gaussian functions respectively. These Gaussian functions are weighted by integrated sensitivities α_c (for the centre) and α_s (for the surround). A detailed analysis of how the shape of the DOG operator is affected by varying the ratio $\frac{\sigma_c}{\sigma_s}$ and $\frac{\alpha_c}{\alpha_s}$ is given in [3]. The response, $R(\vec{x},t)$, of the DOG filter to an input signal $s(\vec{x},t)$ at $\vec{x}$ during time t is given by:

$$R(\vec{x},t) = \int\int_{|\vec{w}|<\infty} DOG(\vec{w}) s(\vec{x}-\vec{w},t) dx dy \quad (2)$$

where the convolution is applied over a circular region centred at $\vec{x}$ (i.e., $|\vec{w}|$ denotes the distance from $\vec{x}$, the centre of the operator).

3 Extending the DOG operator to detect changes

The receptive field[1] in human retinal cells is subdivided into the centre and surround regions, and is classified to be on centre or off centre. The common feature of the two types of receptive fields is that the centre and surround regions are antagonistic. Therefore, if the regions are simulated simultaneously, they cancel each other's contribution. The DOG operator introduced in Section 2 uses simple linear differences to model this centre-surround interaction, and the response $R(\vec{x},t)$ is analogous to a measure of contrast of events happening between centre and surround regions of the operator. When $\sigma_c < \sigma_s$, the centre of the operator has a positive contribution while the surround has a negative contribution. This type of DOG operator responds well when an event is at its centre region. This behaviour is similar to the excitation of on receptive fields in the retina when the stimulus is in the central region of the receptive field. Therefore, this type of operator is termed the on-DOG operator. Conversely, when $\sigma_c > \sigma_s$, the signs of the contribution of the centre and the surround are reversed. This type of DOG operator responds well when events are at its surround region. This is similar to the situation that off centre receptive fields in the retina fire a response when the stimulus is at its surround region. As a result, it will be

[1] The receptive field is an area on the retina in which stimulation of that area influences the firing rate of the associated neuron.

1051-4651/94 $04.00 © 1994 IEEE

termed the off-DOG operator. We will denote the response from on and off-DOG operators at location $\vec{x}$ and time t by $N(\vec{x},t)$ and $F(\vec{x},t)$ respectively. In order to decide if there are any changes happening at location $\vec{x}$, we would have to look at how functions $N(\vec{x},t)$ and $F(\vec{x},t)$ change over the current temporal window. We will first develop a model that detects changes over two consecutive images (i.e. a temporal window of size two).

3.1 Detecting "on" events

Define an "on" event as the situation in which the pixel intensity at the centre region increases. This may be caused by an increase of illumination or the appearance of objects in some previously blank positions. With an "on" event at location $\vec{x}$, the response, as denoted by the function $N(\vec{x},t)$, should increase over time. Typically, in order to reduce the effect of noise and to ensure that the change is significant, we require the rate of increase to exceed a certain threshold, say, θ_1. This rate can be measured by the temporal derivative, $\frac{\partial N}{\partial t}$, of the function N. With a temporal window of size two, $\frac{\partial N}{\partial t}$ can be approximated by taking the first difference of response from two successive image frames. Therefore, the condition

$$\frac{\partial N}{\partial t} \approx \Delta N = N(\vec{x},t) - N(\vec{x},t-1) \geq \theta_1 \qquad (3)$$

has to be satisfied. We also require that

$$|N(\vec{x},t)| \geq \theta_2 \qquad (4)$$

to ensure the contrast between the centre and the surround is large enough to indicate that there is something interesting (e.g., an object) under the spatial extent of the operator.

However, even though Equations (3) and (4) are satisfied, it is not sufficient to make certain that there is an "on" event at the centre region. An increase in the function $N(\vec{x},t)$ may be caused by a reduced intensity level at the surround region while the intensity at the centre region remains unchanged. Therefore, to assure that there is change at the centre, we have one more condition:

$$\frac{\partial N_c}{\partial t} \approx \Delta N_c = N_c(\vec{x},t) - N_c(\vec{x},t-1) \geq \theta_3 \qquad (5)$$

where $N_c(\vec{x},t)$ is the response from the centre region of the operator only[2].

3.2 Detecting "off" events

"Off" events are defined to be events in which the intensity at the centre region decreases. This may be caused by a decrease of illumination or objects disappearing from some previously occupied positions. Under these circumstances, the function $F(\vec{x},t)$ increases over time. With similar reasoning as in looking for "on" events, we require

$$\frac{\partial F}{\partial t} \approx \Delta F = F(\vec{x},t) - F(\vec{x},t-1) \geq \theta_4 \qquad (6)$$

Again, $\frac{\partial F}{\partial t}$ is the temporal derivative of the function F and it is approximated by first difference. To ensure that the change in response is because of a change in the centre region, we require

$$\frac{\partial F_c}{\partial t} \approx \Delta F_c = F_c(\vec{x},t) - F_c(\vec{x},t-1) \geq \theta_5 \qquad (7)$$

[2]The response at the centre can be computed by $N_c(\vec{x},t) = \int\int_{\vec{w}\in centre} DOG(\vec{w})s(\vec{x}-\vec{w})dxdy$

with $F_c(\vec{x},t)$ being the measure of response from the centre region only[3]. But different from the case of finding "on" events, we do not have a condition for minimum contrast between the two regions. The reason for this will be discussed in Section 7.

4 Competition among scales

In order to find events of different scales, on and off-DOG operators of various scales are used. A decision process, which is performed by initiating a winner-take-all (WTA) process for each type of event, is to pick the location and scale of the most salient "on" and "off" events. Units in each WTA network represent the response from all locations and spatial scales of the corresponding event. A winner is determined from each WTA network using the updating rule described by Tsotsos [10]. This WTA updating rule is an enhanced version from the Koch and Ullman scheme [5], and has been proven to converge quickly [1], [6], [9]. But before the WTA processes are initiated, the response from different operator sizes are normalized. A normalization process is necessary because an operator of a larger scale may have a greater response over one with a smaller scale simply because of an increase in area. Therefore, a mechanism has to be found to take a balance between the difference in size and response.

The normalization function we use in our model is the one suggested by Culhane and Tsotsos [2]. Culhane and Tsotsos originally suggest a normalization function to select receptive fields (RF) among different sizes. Their normalization function is a function of the area of the RF, which is measured by the number of pixels in the RF. Such a function can be used by our model because the area over which a DOG operator is applied can be treated as an RF. Therefore, in choosing the operator of appropriate scale, it is analogous to choosing a winning RF among all possible sizes. In our model, we can express the area of RF in terms of the radius of the operator, since a DOG operator of radius r is approximated by a template of size $(2r+1)\times(2r+1)$ [4]. Therefore, the normalization function we use can be expressed as a function of σ:

$$W(\sigma) = \frac{\rho+1}{\rho+\beta^{-(2m\sigma+1)}} \qquad (8)$$

The parameter ρ will affect the asymptote of the function, while β will affect the steepness of the first part of the function. Empirically, setting $\rho = 10$ and $\beta = 1.3$ (as suggested by Culhane [1]) seems to yield satisfactory results for our experiments.

5 Extending the temporal window

When the temporal window is expended to a size T ($T > 2$), our model would be looking for changes that occur at different rates. The change of functions $N(\vec{x},t)$ and $F(\vec{x},t)$ over the T frames has to be monotonic increasing and the magnitude of the overall change (when compared between the first and the last image frame) has to be significant. The ideal situation would be that the T sampled

[3]$F_c(\vec{x},t)$ can be computed from the equation $\int\int_{\vec{w}\in centre} DOG(\vec{w})s(\vec{x}-\vec{w})dxdy$.

[4]In this expression, $r = m\sigma$ where $\sigma = \max(\sigma_c,\sigma_s)$ and m is the number of standard deviations for the cutoff for Equation (1). The number 1 is added to make the centre of the operator fall on a pixel instead of between pixels.

responses all fall onto a straight line with a slope that satisfies the minimum change requirement. However, this ideal case is rarely satisfied due to noise. Instead, we must look at the change in response over time to see if it can be approximated by a straight line. For a particular location $\vec{x}$, this straight line will be joining the response at time t and $t-(T-1)$. Empirically, all responses in between should be within a distance of $0.2\theta_1$ (for "on" events) or $0.2\theta_4$ (for "off" events) from this line to give a monotonic increasing response. The algorithm to detect "on" and "off" events for a general temporal window size ($T \geq 2$) is given in Algorithm 1 (Figure 1).

6 Determining parameter values

The main parameters for the DOG operator are σ_c, σ_s, α_c and α_s. As Fleet[3] pointed out, it is sufficient to consider the ratio $\frac{\sigma_s}{\sigma_c}$ and $\frac{\alpha_c}{\alpha_s}$. By keeping the ratio $\frac{\sigma_s}{\sigma_c}$ fixed and varying σ_c (when $\sigma_s > \sigma_c$), the peak spatial frequency that the operator can detect is shifted. Moreover, the actual size of σ_c should be set according to the size of object or spatial frequencies that the system is required to discern. Therefore, for the on-DOG operators, our model will look at a series of operators with $\frac{\sigma_s}{\sigma_c}$ fixed at k, with $k > 1$, while σ_c varies. For off-DOG operators, the value of σ_s will vary while the ratio $\frac{\sigma_s}{\sigma_c}$ will be $\frac{1}{k}$ (since $\sigma_s < \sigma_c$).

The parameters α_c and α_s affect the sensitivity of the operator. We want to choose α_c and α_s such that the response of the operator will be low when it is applied to a region with uniform intensity (i.e., no contrast between the centre and the surround). That is, when the operator is applied across a region with constant intensity I, we want

$$R(\vec{x},t) = \int\int_{|\vec{w}|<\infty} (\alpha_c G(\vec{w};\sigma_c) - \alpha_s G(\vec{w};\sigma_s)) I dxdy = 0$$

This requires α_c and α_s to satisfy the relation

$$\frac{\alpha_c}{\alpha_s} = 1 \tag{9}$$

This relation implies that the function $DOG(\vec{x})$ will integrate to zero, and this further implies that the operator cannot be of one sign. Therefore, this relation also assures that the on-DOG operator has positive values at the centre and negative values at the surround, and vice verse for the off-DOG operator.

7 Determining threshold values

As the Gaussian function (Equation (1)) falls off quickly with the increase of magnitude of its input parameter, it is sufficient to apply the convolution in Equation (2) over $\vec{w} = \vec{x} - \vec{x}'$, where $\vec{x}'$ is within a distance of $m\sigma$ from $\vec{x}$. Therefore, we compute $R(\vec{x},t)$ as

$$R(\vec{x},t) = \int\int_{|\vec{w}|\leq m\sigma} DOG(\vec{w}) s(\vec{x}-\vec{w},t) dxdy$$

When the operator is applied to some ideal image with constant intensity I_c across the centre and constant intensity I_s for the surround, the response $R(\vec{x},t)$ can be written as a sum of two products. That is,

$$\begin{aligned} R(\vec{x},t) &= I_c \int\int_{\vec{w}\in centre} DOG(\vec{w}) dxdy \\ &\quad + I_s \int\int_{\vec{w}\in surround} DOG(\vec{w}) dxdy \end{aligned} \tag{10}$$

in which the double integral can be computed independent of the pixel values. The centre region of the on-DOG operator refers to the part of the operator where $DOG(\vec{w}) > 0$ and the surround region is the part where $DOG(\vec{w}) < 0$. For off-DOG operators, the signs of the inequalities are reversed. By separating the operator into two regions and assuming uniform intensity for each of them, we will be able to find a close form for the maximum possible contrast, which is also the maximum response, for both types of operators. Empirically, setting the thresholds (θ_1, θ_2, θ_3, θ_4 and θ_5) to a certain percentage of the maximum response seems to yield a good decision criterion for choosing the appropriate values.

Define *ImageResolution* to be the set of pixels composing the image (i.e. the size of the image). Scale $\mathcal{S}$ represents a pair of values for σ_c and σ_s. For an on-DOG operator of scale $\mathcal{S}$, $\sigma_c = \mathcal{S}$ while $\sigma_s = k\mathcal{S}$ ($k > 1$). For an off-DOG operator of scale $\mathcal{S}$, $\sigma_s = \mathcal{S}$ and $\sigma_c = k\mathcal{S}$. Let $\mathcal{R}_N(\vec{x},\mathcal{S})$ and $\mathcal{R}_F(\vec{x},\mathcal{S})$ be the response in detecting "on" events and "off" events respectively at location $\vec{x}$ (with respect to pixel intensity values) with operator scale $\mathcal{S}$. Let T be the maximum temporal window size being considered.

1. For all possible temporal window size T', $2 \leq T' \leq T$, do step 2 to step 12 for all possible subsequences of T consecutive images.
2. For each pixel $\vec{x} \in$ *ImageResolution*, do step 3 to step 9.
3. For each scale $\mathcal{S}$, do step 4 to step 9.
4. Compute $N(\vec{x},t')$, $F(\vec{x},t')$, $N_c(\vec{x},t')$, $F_c(\vec{x},t')$ for $t-(T'-1) \leq t' \leq t$.
5. Compute $\Delta N = N(\vec{x},t) - N(\vec{x},t-(T'-1))$, $\Delta F = F(\vec{x},t) - F(\vec{x},t-(T'-1))$, $\Delta N_c = N_c(\vec{x},t) - N_c(\vec{x},t-(T'-1))$, $\Delta F_c = F_c(\vec{x},t) - F_c(\vec{x},t-(T'-1))$.
6. Check if the functions $N(\vec{x},t)$ and $F(\vec{x},t)$ are monotonic increasing, and if $N(\vec{x},t')$ and $F(\vec{x},t')$ are of a reasonable fit to a straight line. That is, $N(\vec{x},t')$ and $F(\vec{x},t)$ can be approximated by a straight line with maximum error $0.2\theta_1$ and $0.2\theta_4$ respectively.
7. If $N(\vec{x},t)$ satisfies step 6 and $\Delta N \geq \theta_1$, $|N(\vec{x},t)| \geq \theta_2$, $\Delta N_c \geq \theta_3$, then something goes on at $\vec{x}$
   Set $\mathcal{R}_N(\vec{x},\mathcal{S}) = \Delta N$;
   else
   Set $\mathcal{R}_N(\vec{x},\mathcal{S}) = 0$
8. If $F(\vec{x},t)$ satisfies step 6 and $\Delta F \geq \theta_4$, $\Delta F_c \geq \theta_5$, then something goes off at $\vec{x}$
   Set $\mathcal{R}_F(\vec{x},\mathcal{S}) = \Delta F$;
   else
   Set $\mathcal{R}_F(\vec{x},\mathcal{S}) = 0$
9. Return to step 4 if have not computed all scales.
10. Normailze response from all scales using Equation (8).
11. Two WTA networks, one for "on" events and the other for "off" events, are initiated to find the location and scale of the most salient "on" and "off" events.
12. Return to step 2 if have not worked on all necessary temporal window sizes.
13. Pick the final winner for on and off-DOG operators by running WTA over the winner for each temporal window.
14. Return to step 1 for the next set of images.

Figure 1: Algorithm 1 – Algorithm for a general temporal window size T ($T \geq 2$)

7.1 Determining thresholds for the on-DOG operator

For on-DOG operators, the set of pixels $\vec{x}'$ that belongs to the centre region of the on-DOG operator are those that

satisfy the relation

$$|\vec{w}| < \sigma_s \sqrt{\frac{2\log(\frac{\alpha_c}{\alpha_s}k^2)}{k^2-1}}$$

Let

$$z_n = \sqrt{\frac{2\log(\frac{\alpha_c}{\alpha_s}k^2)}{k^2-1}} \text{ and } w_n = \sigma_s z_n$$

Therefore, any pixels that are within a distance of w_n from the centre pixel $\vec{x}$ falls into the centre region of the operator and other pixels will fall into the surround region. In terms of w_n, Equation (10) becomes

$$\begin{aligned} R(\vec{x},t) &= I_c \int\int_{|\vec{w}|<w_n} DOG(\vec{w})dxdy \\ &\quad + I_s \int\int_{w_n<|\vec{w}|\leq m\sigma} DOG(\vec{w})dxdy \end{aligned}$$

We can derive a close form for the two double intergals in the above equation in terms of z_n:

$$\int\int_{|\vec{w}|<w_n} DOG(\vec{w})dxdy =$$
$$-\alpha_c(e^{-\frac{(kz_n)^2}{2}} - 1) + \alpha_s(e^{-\frac{z_n^2}{2}} - 1) = s_{nc}$$
$$\int\int_{w_n<|\vec{w}|\leq m\sigma} DOG(\vec{w})dxdy =$$
$$\alpha_c(e^{-\frac{(kz_n)^2}{2}} - e^{-\frac{(km)^2}{2}}) - \alpha_s(e^{-\frac{z_n^2}{2}} - e^{-\frac{m^2}{2}}) = s_{ns}$$

The value of $R(\vec{x},t)$ will be maximum for an on-DOG operator when the centre has uniform maximum intensity I_{max} and the surround has minimum intensity I_{min}. These maximum and minimum values can be found by looking at a histogram formed by intensity values of pixels from current images. Therefore, the maximum response R_n^{max} for an on-DOG operator is

$$R_n^{max} = s_{nc}I_{max} + s_{ns}I_{min}$$

With the value of R_n^{max} computed, values of thresholds θ_1, θ_2 and θ_3 are set to

$$\theta_1 = |p_1 R_n^{max}|, \quad \theta_2 = |p_2 R_n^{max}|, \quad \theta_3 = |p_3\theta_1|$$

where p_1, p_2 and p_3 are values between 0 and 1. Note that since s_{nc} and s_{ns} are independent of σ_c and σ_s, θ_1, θ_2 and θ_3 are the same for all spatial scales. Empirically, the values of p_1, p_2 and p_3 are set to 0.1, 0.15 and 0.2 respectively.

7.2 Determining thresholds for the off-DOG operator

The set of pixels, $\vec{x}'$, that fall into the centre region of the off-DOG operator has to satisfy the relation

$$|\vec{w}| < \sigma_c \sqrt{\frac{2\log(\frac{\alpha_s}{\alpha_c}k^2)}{k^2-1}}$$

Let

$$z_f = \sqrt{\frac{2\log(\frac{\alpha_s}{\alpha_c}k^2)}{k^2-1}} \text{ and } w_f = \sigma_c z_f$$

Any pixels that are within a distance of w_f from the centre of the operator fall into the centre region. Otherwise, they belong to the surround. And in terms of z_f, we evaluate

$$\int\int_{|\vec{w}|<w_f} DOG(\vec{w})dxdy =$$
$$-\alpha_c(e^{-\frac{z_f^2}{2}} - 1) + \alpha_s(e^{-\frac{(kz_f)^2}{2}} - 1) = s_{fc}$$
$$\int\int_{w_f<|\vec{w}|\leq m\sigma} DOG(\vec{w})dxdy =$$
$$\alpha_c(e^{-\frac{z_f^2}{2}} - e^{-\frac{m^2}{2}}) - \alpha_s(e^{-\frac{(kz_f)^2}{2}} - e^{-\frac{(km)^2}{2}}) = s_{fs}$$

The maximum response, R_f^{max}, of the off-DOG operator is achieved when the centre has minimum intensity I_{min} and the surround has maximum intensity I_{max}. So,

$$R_f^{max} = s_{fc}I_{min} + s_{fs}I_{max}$$

and we set

$$\theta_4 = |p_4 R_f^{max}|, \qquad \theta_5 = |p_5 R_f^{max}|$$

where p_4, and p_5 are values between 0 and 1. Once again θ_4 and θ_5 are independent of scale, σ_c and σ_s.

As opposed to detecting "on" events, we do not require a contrast between the centre and the surround when detecting "off" events. When some object moves away from its original position (with the object centred at $\vec{x}$), the region with $\vec{x}$ as the centre under the current image is the background. Therefore, there is no need to require a contrast at the current frame for some region to be regarded as with an "off" event [5].

8 Integration with an attentional model

The computational model described above can be used to direct attention to abrupt changes for eye-head movement control. The vision system will acquire T images (where T is the maximum temporal window size), and use Algorithm 1 to find locations where "on" and "off" events have occurred. The output from Algorithm 1 is treated as a saliency measure for the input level of the processing hierarchy in Tsotsos' inhibitory attentional beam model [9]. The most conspicuous "on" and "off" events will compete with other attention attracting image events (e.g., events that deserve attention according to some task-driven guidance) and a higher order decision process is assumed to decide which event deserves attention. If necessary, the robot will move to fixate centre on the selected area for attention. The appropriate area will be inhibited after this shift of attention, and the input to the lowest level of the hierarchy that directs attention according to abrupt changes will be refreshed. The reason that the input has to be refreshed in our model is because we are interested in abrupt changes. Therefore, changes that have occurred over a certain period of time will lose their priority for attention. Furthermore, when the attention model is directing the motion of a robot head, we have to acquire a new set of images every time the head has moved.

One aspect for an attention model that we have not addressed is the issue of inhibition of return [7]. It was found from psychological experiments that there is a temporary inhibition after attention shifts away from a position. However, there have been no experiments to study how receptive fields are inhibited when attention is guided by abrupt changes. More experiments in the psychology area to study how this inhibition is achieved in the human visual system due to abrupt changes may help to suggest a way to deal with this problem in our model. Nevertheless, the proposed model contributes towards forming an attentional model in which attention can be guided by various aspects, including abrupt changes.

9 Implementation results

The computational model, as described by Algorithm 1 (Figure 1), has been implemented in software on Sili-

[5] We do not need to consider the contrast on the previous frame in order to classify there is an "off" event. The reason is because an "off" event may occur at a location which is a bright uniform background in the previous frame.

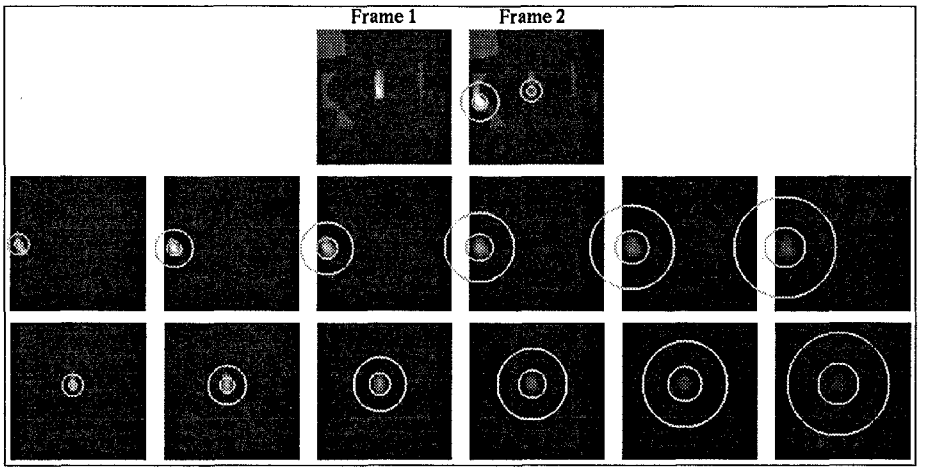

Figure 2: Example of moving a spotlight among some blocks and cylinders: The winning "on" event is the increase in luminance of the left most block; the winning "off" event is the decrease in luminance of the cylinder in the centre.

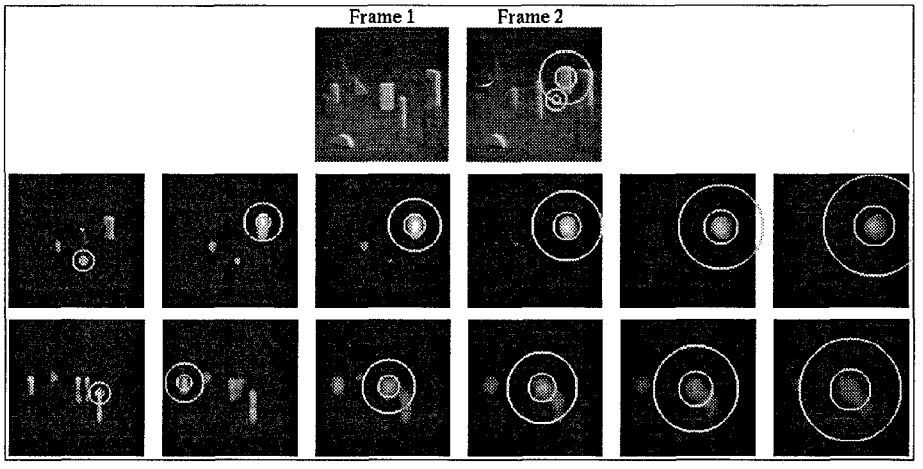

Figure 3: Example of several blocks moving among other stationary blocks: The most prominent "on" event is at the new position of the rectangular block at the upper right; the winning "off" event is at the old position of the cylindrical block.

con Graphics 4D/380 VGX. The WTA updating process is implemented as a simple sequential search because of hardware limitations. Simulations are to run on sequences of digitized 128 × 128 images in which changes are caused by brightness and motion. In the present implementation, each sequence of images is acquired in advance through a stationary monochrome CCD camera. The parameters of the model are set to $\alpha_c = \alpha_s = 2$, $k = 2.5$ and $m = 2$. The actual size of operators used may vary between experiments. The thresholds θ_1, θ_2, θ_3, θ_4 and θ_5 are set as described in Section 7. The parameters we used for the normalization function (Equation (8)) are $\rho = 10$ and $\beta = 1.03$.

When the lighting condition in a visual scene changes, an "on" event will refer to a case where the amount of light falling on an object increases. On the other hand, an "off" event refers to a case where the amount of light falling on an object decreases. Figure 2 shows the result from a sequence of images in which a spotlight moves from the cylinder in the centre (an "off" event) to the block on its left (an "on" event). Six different scales for each type of operator are used to detect "on" and "off" events in this sequence under a temporal window of size two. The sizes of these operators are chosen by an intuitive approximation of the scale of events. The second and third rows in Figure 2 show which pixels are marked as candidates for "on" and "off" events respectively. The grey level of each pixel shows the magnitude of change measured at that point. The brighter the pixel, the greater the magnitude of change. Pixels at which no changes have been detected are indicated by intensity value 0 (black). From left to right, each column shows the magnitude of change as measured by operators of increasing scale. Blue circles indicate the winning area when detecting "on" and "off" events using operators of a particular scale. Red and yellow circles indicate the scale, as well as the location, of the operator that gives the most salient "on" and "off" event respectively. In the last row, the areas for most conspicuous "on" and "off" events are superimposed in Frame 2 of the original image. From this example, we can see that the magnitude of change as measured by operators decreases when the operator size increases beyond the scale of the event. Note that the scales for the winning "on" event and the winning "off" event are not necessarily the same. Therefore, this example shows that besides detecting the change of brightness of objects, the model can

also detect events with the appropriate scale.

Figure 3 is an example in which changes are caused by motions of objects. The representation used for displaying the results is the same as in Figure 2 and we also use a temporal window of size two. By assuming that objects have pixel intensity values greater than the background, locations where objects have moved to will be regarded as "on" events. Previous positions of objects are indicated by "off" events. Note that if this assumption is violated, events that register the old and new positions of objects may be reversed. We further assume that there is some mechanism outside the model to decide if the new position of an object is the "on" or "off" event.

There are four different events happening in Figure 3. First, the triangular block at the upper left of Frame 1 disappears in Frame 2. Second, there is a small rectangular block on the left moving towards the centre. Third, the rectangular block moves from the centre to the upper right. Fourth, the cylindrical block from the lower right moves towards the centre. The "on" events for this example are, therefore, the moving of the two rectangular blocks and the cylindrical block to their new positions. The "off" events are the disappearance of the triangular block and the three blocks moving away from their original positions. Note that not all pixels at the new position of the cylindrical block are classified as having an "on" event. The new position of the cylindrical block partly overlaps with the former position of the rectangular block. Since the intensity levels of the two objects are roughly the same, there is no change in response at the overlap part. The same effect is observed when detecting "off" events. We can also observe from this example that different events are chosen as winner from operators of different sizes.

Simulations using a temporal window other than 2 were also tried with the expected results. For example, suppose the luminance of blocks is decreased gradually and events are detected over a longer time course instead of a shorter one. As the amount of light is decreased continuously for the whole scene, we can expect that there will be "off" events only.

10 Discussion

We have derived a computational model that can detect abrupt changes such that it can be used in the early stages of processing in a vision system. In particular, its output can be used by an attentional model as a source of information to direct attention according to changes in an image sequence. The model we introduce operates on raw pixel intensities, and continuously refreshes its input in order to find the most recent change in the environment. However, the model has a number of assumptions and limitations. First of all, we take for granted that there is perfect registration from the device through which images are obtained and every change is caused by the change in the intensity value of a happening event. Second, we assume that all kinds of changes, including changes caused by shadows, are equally worthy of attention as other changes with respect to objects. If the effect of shadow is undesirable, some colour processing should be applied to filter out shadows [4]. Thirdly, the correspondence problem is not being addressed by the model. There is no mechanism to identify the changes on the same object at different times. Fourthly, there is no top-down component in our model in its most original form.

In spite of its limitations, our computational model has several significant contributions. It investigates a new avenue that directs the attention of a machine vision system with respect to abrupt changes. It is one of the few models that directs attention based on dynamic scenes. Furthermore, the model can detect changes in a purely bottom-up fashion and does not require prior knowledge of the scene. There is also no restriction on object types on which the model can work. Specifically, the computation model can be used to compute one of the inputs at the bottom layer of the processing hierarchy in Tsotsos' inhibitory attentional beam model, and forms one of the many pipelines that competes for attention. The work here also suggests a mechanism in which tracking can be done by a simple means. In terms of tracking, this model especially has the advantage that it is not restricted to work on motion only. With the input signal $s(\vec{x}, t)$ representing different features, the model can be used to find changes in a visual scene with respect to different features without alteration.

Acknowledgements

The second author is the CP-Unitel Fellow of the Canadian Institute for Advanced Research. This research was funded by the Information Technology Research Centre, one of the Province of Ontario Centres of Excellence, the Institute for Robotics and Intelligent Systems, a Network of Centres of Excellence of the Government of Canada. We also thank S. Culhane for help in typesetting this paper.

References

[1] Culhane, S. M., "Implementation of an Attentional Prototype for Early Vision", Master's Thesis, University of Toronto, 1992.

[2] Culhane, S. M. and Tsotsos, J. K., "An Attentional Prototype for Early Vision", *ECCV*, p. 551–560, 1992.

[3] Fleet, D. J., "The Early Processing of Spatio-Temporal Visual Information", Master's Thesis, University of Toronto, 1984.

[4] Gershon, R., "The Use of Colour in Computational Vision", PhD Thesis, University of Toronto, 1987.

[5] Koch, C. and Ullman, S., "Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry", *Human Neurobiology*, Vol. 4, p. 219–277, 1985.

[6] Lai, Y. Z., "A Prototype for Finding Motion Patterns in Optical Flow", Master's Thesis, University of Toronto, 1992.

[7] Posner, M. I. and Cohen, Y., "Components of Attention", in *Attention and Performance X*, p. 531–556, 1984.

[8] Tsotsos, J. K., "Analyzing Vision at the Complexity Level", *Behavioural and Brain Science*, No. 13, p. 423–469, 1990.

[9] Tsotsos, J. K., "Localizing Stimuli in a Sensory Field Using an Inhibitory Attentional Beam", Technical Report RBCV-TR-91-37, University of Toronto, 1991.

[10] Tsotsos, J. K., "An Inhibitory Beam for Attentional Selection", in *Spatial Vision for Humans and Robots*, Cambridge University Press, 1993.

[11] Yantis, S. and Hillstrom, A. P., "Stimulus-Driven Attentional Capture: Evidence From Equiluminant Visual Objects", *Journal of Experimental Psychology: Human Perception Performance*, in press.

[12] Yantis, S. and Johnson, D. N., "Mechanisms of Attention Priority ", *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 16, No. 4, p. 812–825, 1990.

[13] Yantis, S. and Jones, E., "Mechanisms of Attention Selection: Temporally Modulated Priority Tags", *Perception and Psychophysics*, 50(2), p.166–178, 1991.

[14] Yantis, S. and Jonides, J., "Abrupt Visual Onsets and Selective Attention: Evidence from Visual Search ", *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 10, No. 5, p. 601–621, 1984.

[15] Yantis, S. and Jonides, J., "Abrupt Visual Onsets and Selective Attention: Voluntary Versus Automatic Allocation", *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 16, No. 1, p. 121–134, 1990.