A COMPUTATIONAL SALIENCY MODEL INTEGRATING SACCADE PROGRAMMING

Tien Ho-Phuoc, Anne Guérin-Dugué and Nathalie Guyader
GIPSA-lab/Department of Images and Signals, 961 rue de la Houille Blanche, F - 38402 Saint Martin d'Hères cedex, France

Keywords: Saccade programming, Saliency map, Spatially variant retinal resolution.

Abstract: Saliency models have showed the ability of predicting where human eyes fixate when looking at images. However, few models are interested in saccade programming strategies. We proposed a biologically-inspired model to compute image saliency maps. Based on these saliency maps, we compared three different saccade programming models depending on the number of programmed saccades. The results showed that the strategy of programming one saccade at a time from the foveated point best matches the experimental data from free viewing of natural images. Because saccade programming models depend on the foveated point where the image is viewed at the highest resolution, we took into account the spatially variant retinal resolution. We showed that the predicted eye fixations were more effective when this retinal resolution was combined with the saccade programming strategies.

1 INTRODUCTION

Eye movement is a fundamental part of human vision for scene perception. People do not look at all objects at the same time in the visual field but sequentially concentrate on attractive regions. Visual information is acquired from these regions when the eyes are stabilized (Egeth and Yantis, 1997; Henderson, 2003). Psychophysical experiments with eye-trackers provide experimental data for both behavioral and computational models to predict the attractive regions.

Computational models are in general divided into two groups: task-independent models (bottom-up) and task-dependent models (top-down). Most models describe bottom-up influences to create a saliency map for gaze prediction. They are inspired by the concept of the Feature Integration Theory of Treisman and Gelade (Treisman and Gelade, 1980) and by the first model proposed by Koch and Ullman (Koch and Ullman, 1985). The most popular bottom-up saliency model was proposed by Itti (Itti et al., 1998).

Eye movement experimental data consists, in general, of fixations and saccades. Most models were usually evaluated with distribution of fixations rather than distribution of saccade amplitudes (in this paper, *saccade distribution* will be used to refer to distribution of saccade amplitudes). In order to evaluate more precisely human saccades, saliency prediction must be included inside a spatio-temporal process, even if we consider only a bottom-up visual saliency map from low-level features based on still images. A question can be asked: is visual saliency evaluated again at each fixation, or not?

Some studies showed that the saccadic system can simultaneously program two saccades to two different spatial locations (McPeek et al., 2000). This means that from one foveated point, the next two saccades are programmed in parallel; this is called the *concurrent processing* of saccades. Moreover, another study (McPeek et al., 1998) showed that when subjects are explicitly instructed to make a saccade only after the current information at the fovea has been analyzed, they have difficulty using this strategy. In this paper, we consider different saccade programming strategies during free viewing of natural images and we answer the questions: Do subjects make one saccade at a time, programming the next saccade from the current foveated point? Or do subjects program the next two saccades from the current foveated point? To answer these questions, we tested two different models depending on the corresponding number of programmed saccades from one foveated point (cf. section 2.4). These two models were compared with the baseline model in which all fixations were predicted from the same foveated point. The mechanism of saccade programming will be discussed through our experimental results, as it is still an open question.

Saccade programming models are greatly linked

Ho-Phuoc T., Guérin-Dugué A. and Guyader N. (2009).
A COMPUTATIONAL SALIENCY MODEL INTEGRATING SACCADE PROGRAMMING.
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, pages 57-64
DOI: 10.5220/0001535500570064

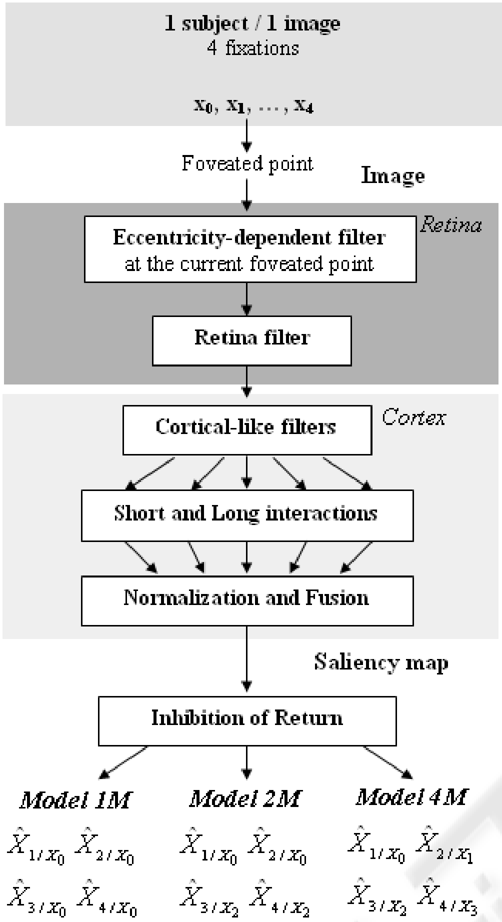


Figure 1: Functional description of the three saccade programming models proposed.

to image resolution when it is projected on the retina. The image on the fovea is viewed at higher resolution compared to the peripheral regions. This enhances the saliency of the region around the foveated point. The spatial evolution of visual resolution is one of the consequences of the non-uniform density of photoreceptors at retina level (Wandell, 1995). When comparing experimental and predicted scanpaths, Parkhurst (Parkhurst et al., 2002) noticed an important bias on saccade distribution. The predicted saccade distribution is quite uniform, contrary to the experimental one where fixations are indeed located on average near the foveated point. The decrease of visual resolution on the peripheral regions contributes to an explanation of this effect. In (Parkhurst et al., 2002), this property was integrated in the model at the final stage (multiplication of the saliency map by a gaussian function).

In our model, we implement the decreasing density of photoreceptors as the first cause of this inhomegeneous spatial representation. The output of photoreceptors encodes spatial resolution depending on eccentricity, given a parameter of resolution decay (Geisler and Perry, 1998). While this parameter influents low level retina processing, we intend to go further in varying its value to mesure its impact on high level visual processing, i.e. including the first steps of the visual cortex. Similarly, in (Itti, 2006), they noticed that the implementation of the spatially variant retinal resolution improved the ability of fixation prediction for dynamic stimuli (video games), but the question of saccade programming was not addressed. Here, we show that this positive effect of fixation prediction is even significant with static visual stimuli in combination with saccade programming models.

Consequently, the proposed models are dynamic ones as the high resolution regions, corresponding to the fovea, change temporally according to the scanpath. Three saccade programming strategies are analyzed. Experimental results show the interest of this approach with spatially variant retinal resolution. The models are described in section 2. Section 3 presents the experiment of free viewing to record eye fixations and saccades, and the evaluation of the proposed models. Discussion is drawn in section 4.

2 DESCRIPTION OF THE PROPOSED MODELS

Our biologically-inspired models (Fig. 1) consist of one or more saliency maps in combination with saccade programming strategies. The model integrates the bottom-up pathway from low-level image properties through the retina and primary visual cortex to predict salient areas. The retina filter plays an important role: implementing non-linear response of photoreceptors and then, spatially variant resolution. The retinal image is then projected into a bank of cortical-like filters (Gabor filters) which compete and interact, to finally produce a saliency map. This map is then used with a mechanism of "Inhibition of Return" to create the saccade programming model. All these elements are described below.

2.1 Retina Filter

In (Beaudot et al., 1993) the retina is modeled quite completely according to its biological functions: spatially variant retinal resolution, luminance adaptation and contrast enhancement. These three characteristics are taken into account while usually a simple model is used such as a function of difference of Gaussians (Itti et al., 1998).

Spatially variant retinal resolution: Since the density of photoreceptors decreases as eccentricity from the fovea increases (Wandell, 1995), the bluring effect increases from the fovea to the periphery and is reflected by an eccentricity-dependent filter. The cut-off frequency f_{co} of this filter decreases with eccentricity ecc expressed in degree (Fig. 2a). The variation of the cut-off frequency (Eq. 1) is adapted from (Geisler and Perry, 1998; Perry, 2002):

$$f_{co}(ecc) \quad = \quad fmax_{co} \cdot \frac{\alpha}{\alpha + |ecc|}, \qquad (1)$$

where α is the paramater controlling the resolution decay, and $fmax_{co}$ the maximal cut-off frequency in the fovea. Biological studies showed that α is close to $2.3°$ (Perry, 2002). Figure 2b shows an example of an image filtered spatially at the center with $\alpha = 2.3°$.

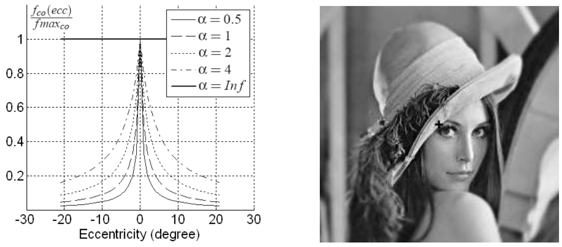
Luminance adaptation: Photoreceptors adapt to a varying range of luminance and increase luminance in dark regions without saturation of luminance in bright ones. They carry out a compression function (Eq. 2) :

$$y = y_{max} \cdot \frac{x}{x + x_o}, \qquad (2)$$

where x is the luminance of the initial image, x_o represents its average local luminance, y_{max} a normalization factor and y the photoreceptor output (Beaudot et al., 1993).

Contrast enhancement: The output of horizontal cells, the low-pass response of photoreceptors, passes through bipolar cells and then, through different types of ganglion cells: parvocellular, magnocellular and koniocellular cells. We only consider the two principal cells: parvocellular and magnocellular. Parvocellular cells are sensitive to high spatial frequency and can be modeled by the difference between photoreceptors and horizontal cells. Therefore, they enhance the initial image contrast and whiten its energy spectrum (Fig. 5b). Magnocellular cells respond to lower spatial frequency and are modeled by a low-pass filter like horizontal cells.

In human visual perception, we know that low frequencies precede high frequencies (Navon, 1977). In our model, we do not take into account this temporal aspect. However, as both low and high frequency components are necessary for a saliency map, we compute the retina output as a linear combination of the parvocellular and magnocellular outputs (Fig. 5b).



(a) (b)

Figure 2: (a) Normalized cut-off frequency as function of eccentricity for different values of α controlling the resolution decay. (b) An example of the output of the eccentricity-dependent filter with $\alpha = 2.3°$, the foveated point is at the image center (marked with the cross).

2.2 Cortical-Like Filters

2.2.1 Gabor Filters

Retinal image is transmitted to V1 which processes signals in different frequencies, orientations, colors and motion. We only consider the frequency and orientation decomposition carried out by complex cells. Among several works modeling responses of these cells, Gabor filters are evaluated as good candidates. A set of Gabor filters is implemented to cover all orientations and spatial frequencies in the frequential domain. A filter $G_{i,j}$ (Eq. 3) is tuned to its central radial frequency f_j at the orientation θ_i. There are N_θ orientations and N_f radial frequencies ($N_\theta = 8$ and $N_f = 4$). The radial frequency is such as $f_{N_f} = 0.25$ and $f_{j-1} = \frac{f_j}{2}, j = N_f, .., 2$. The standard deviations of $G_{i,j}$ are $\sigma^f_{i,j}$ and $\sigma^\theta_{i,j}$ in the radial direction and its perpendicular one, respectively. $\sigma^f_{i,j}$ is chosen in such a way that filters $G_{i,j}$ and $G_{i,j-1}$ are tangent at level of 0.5. We notice that the choice of the standard deviations influences the predicted saliency map. We choose $\sigma^f_{i,j} = \sigma^\theta_{i,j}$, which is justified in the next section.

$$G_{i,j}(u,v) = exp\left\{ -\left(\frac{(u' - f_j)^2}{2(\sigma^f_{i,j})^2} + \frac{v'^2}{2(\sigma^\theta_{i,j})^2} \right) \right\} \quad (3)$$

with:

$$\begin{cases} u' = u\cos(\theta_i) + v\sin(\theta_i) \\ v' = v\cos(\theta_i) - u\sin(\theta_i) \end{cases}$$

where u (respectively v) is the horizontal (respectively vertical) spatial frequency. Then, for each channel, complex cells are implemented as the square amplitude of the Gabor filter output, providing the energy maps $e_{i,j}$.

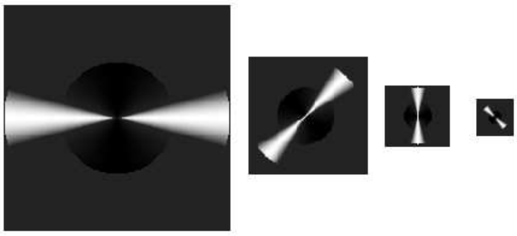


Figure 3: Example of butterfly masks with different orientations and scales.

2.2.2 Interaction between Filters

The responses of neurons in the primary visual cortex are influenced in the manner of excitation or inhibition by other neurons. As these interactions between neurons are quite complex, we only consider two types of interactions based on the range of receptive fields (Hansen et al., 2001).

Short interactions are the interactions among neurons having overlapping receptive fields. They occur with the same pixel in different energy maps (Eq. 4):

$$m^s_{i,j} = 0,5.e_{i,j-1} + e_{i,j} + 0,5.e_{i,j+1} - 0,5.e_{i+1,j} - 0,5.e_{i-1,j}. \quad (4)$$

These interactions introduce inhibition between neurons of neighboring orientations on the same scale, and excitation between neurons of the same orientation on neighboring scales. For the standard deviations of the cortical-like filters, if $\sigma^f_{i,j} > \sigma^\theta_{i,j}$ the filters are more orientation-selective. However, this choice reduces the inhibitive interaction. So, we choose $\sigma^f_{i,j} = \sigma^\theta_{i,j}$.

The second interactions are long interactions which occur among colinear neurons of non-overlapping receptive fields and are often used for contour facilitation (Hansen et al., 2001). This interaction type is directly a convolution product on each map $m^s_{i,j}$ to produce an intermediate map $m_{i,j}$. The convolution kernel is a "butterfly" mask $b_{i,j}$ (Fig. 3). The orientation of the mask $b_{i,j}$ for the map $m^s_{i,j}$ is the orientation θ_i, and the mask size is inversely proportional to the central radial frequency f_j. The "butterfly" mask $b_{i,j}$ has two parts: an excitatory part $b^+_{i,j}$ in the preferential direction θ_i and an inhibitive one $b^-_{i,j}$ in all other directions. It is normalized in such a way that its summation is set to 1. Figure 4 (first row) shows the interaction effect for contour facilitation (Fig. 4c).

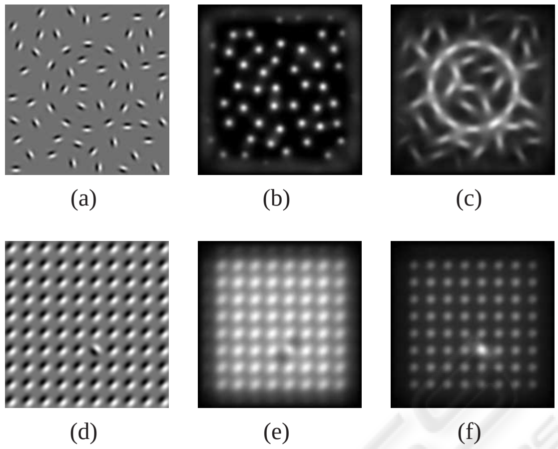


Figure 4: Tests showing the influence of interaction and normalization steps. The initial images are on the left column. The corresponding saliency maps are on the center and right column. First row, interaction effect: (a) Initial image; (b) without interaction; (c) with interaction. Second row, normalization effect: (d) Initial image; (e) without normalization; (f) with normalization.

2.3 Normalization and Fusion

Intermediate maps $m_{i,j}$ must be normalized before fusion. Moreover, an object is more salient if it is different in comparison to its neighbors. We use the method proposed by L. Itti (Itti et al., 1998) to strengthen intermediate maps. Then, the maps are thresholded. Figure 4 (second row) represents the role of the normalization step in reinforcing the filter output where a Gabor patch is different from the background (Fig. 4f).

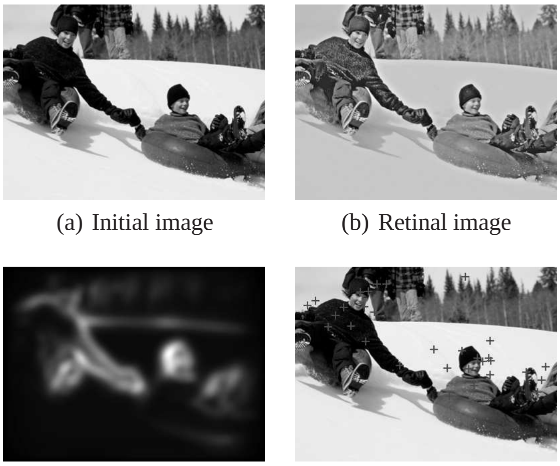
The normalization of each map is carried out as follows:

- Normalize the intermediate map in $[0,1]$
- Let us designate $m^*_{i,j}$ the maximal value of map $m_{i,j}$ and $\overline{m}_{i,j}$ its average. Then, the value at each pixel is multiplied by $(m^*_{i,j} - \overline{m}_{i,j})^2$.
- Set to zero all the values which are smaller than 20% of the maximal value.

Finally, all intermediate maps are summed up in different orientations and frequencies to obtain a saliency map. Examples of retinal image and saliency map are given in Fig. 5.

2.4 Saccade Programming

Our models predict the first four fixations as an extension of the study in (McPeek et al., 2000) (two



(a) Initial image (b) Retinal image

(c) Saliency map (d) Human fixations

Figure 5: Examples of retinal output, saliency map and human fixations for a part of an image used in the experiment.

saccades or two fixations). The fixations are predicted (and consequently the saccades) from the output saliency map by implementing a simple mechanism of "Inhibition of Return" (IOR). We have defined three saccade programming models, called "1M", "2M" and "4M" (Fig. 1). The "1M" model has often been used with the hypothesis that all fixations can be predicted from the same foveated point at the image center. Here, we tested two other models where each foveated point allows the next or the next two fixations to be predicted ("4M" or "2M", respectively). Let us explain first the IOR mechanism with the "1M" model which is considered as the baseline model. The first fixation is chosen as the pixel which has the maximal value on the saliency map. Then, this fixation becomes the foveated point to predict the second saccade. Hence, the second fixation is chosen as the pixel of maximal value after inhibiting the surface around the foveated point (radius of 1°). This mechanism continues for the third and fourth fixations. For the "2M" and "4M" models, the IOR mechanism is applied in the same manner as for the "1M" model except that the four fixations are predicted from more than one saliency map (see below).

In the "1M" model, the eccentricity-dependent filter is applied only once to the center X_0 of images (because during the eye movement experiment, images appeared only if subjects were looking at the center of the screen). Then, a saliency map is computed from this foveated point to predict four fixations $\hat{X}_{1/X_0}$, $\hat{X}_{2/X_0}$, $\hat{X}_{3/X_0}$, $\hat{X}_{4/X_0}$ by using the IOR mechanism. For the "2M" model, the eccentricity-dependent filter is applied first to the center X_0, as for the "1M" model, to predict the first two fixations (on the first saliency map) $\hat{X}_{1/X_0}$, $\hat{X}_{2/X_0}$; then, the eccentricity-dependent filter is applied again to the second fixation for predicting the next two fixations (on the second saliency map) $\hat{X}_{3/X_2}$, $\hat{X}_{4/X_2}$. Because the second fixation is different from one subject to another we take into account the second fixation X_2 of each subject for each image. The "4M" model does the same by applying the eccentricity-dependent filter and calculating the saliency map for each fixation of each subject (four saliency maps are sequentially evaluated).

3 EXPERIMENTAL EVALUATION

3.1 Eye Movement Experiment

We ran an experiment to obtain the eye scanpaths of different subjects when they were looking freely at different images. Figure 5d shows the fixations of all subjects on a part of an image. The recording of eye movements served as a method to evaluate our proposed model of visual saliency and saccade programming.

Participants: Eleven human observers were asked to look at images without any particular task. All participants had normal or corrected to normal vision, and were not aware of the purpose of the experiment.

Apparatus: Eye tracking was performed by an Eyelink II (SR Research). We used a binocular recording of the pupils tracking at 500Hz. A 9-point calibration was made before each experiment. The velocity saccadic threshold is 30°/ s and the acceleration saccadic threshold is 8000°/ s^2.

Stimuli: We chose 37 gray level images (1024 × 768 pixels) with various contents (people, landscapes, objects or manufactural images).

Procedure: During the experiment, participants were seated with their chin supported in front of a 21" color monitor (75 Hz refresh rate) at a viewing distance of 57 cm (40° × 30° usable field of view). An experiment consisted in the succession of three items : a fixation cross in the center of the screen, followed by an image during 1.5 s and a mean grey level screen for 1 s. It is important to note that the image appeared only if the subject was looking at the fixation cross; we ensured the position of the eyes before the onset of images. Subjects saw the same 37 images in a random order.

We analyzed the fixations and saccades of the guiding eye for each subject and each image.

3.2 Criterion Choice for Evaluation

The qualities of the model, more precisely the saccade programming strategy and the eccentricity-dependent filter, are evaluated with the experimental data (fixations and saccades). This evaluation allows to test the predicted salient regions and the saccade distribution.

Firstly, the three models "1M", "2M" and "4M" are used to test saccade programming strategies. The evaluation protocol as in (Torralba et al., 2006) is to extract from a saliency map the most salient regions representing 20% of the map surface. Let us call each fixation "correct fixation" (respectively "incorrect fixation") if the fixation is inside (respectively outside) these predicted salient regions. The ratio $R_c(i,s)$ of "correct fixation" for an image i and a subject s is given below:

$$R_c(i,s) = \frac{N_{inside}}{N_{all}}.100, \quad (5)$$

where N_{inside} is the number of fixations inside the salient regions and N_{all} is the total number of fixations.

The average ratio of correct fixations of all couples (subject $\times$ image) is computed. Particularly for each couple, in the "1M" model, one saliency map is used to calculate correct fixation index for the first four fixations. In the "2M" model, the first saliency map is used for the first two fixations and the second map for the next two fixations (Fig. 1). Similarly for the "4M" model, each saliency map is used for the corresponding fixation. It is noticed that in the "1M" model, the saliency map of each image is identical for all subjects (whose foveated point is always at the image center). However, it is no longer the case for the "2M" or "4M" model where a saliency map depends on the foveated points of a subject.

Secondly, from the most suitable saccade programming strategy chosen above, the predicted saccade distribution for the first four saccades is computed and compared with the emperical one from our experimental data. It has been shown that the parameter α (Eq. 1) which fits the experimental data based on contrast threshold detection when presenting eccentred gratings is around 2.3° (Perry, 2002). Here, by varying the α values, the expected effects are: (i) this parameter must have a great influence on saccade distribution and (ii) the best value would be in the same order of magnitude as 2.3°.

3.3 Results

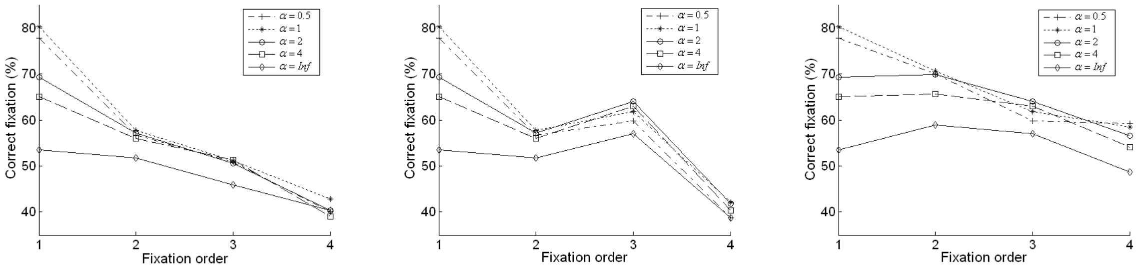
3.3.1 Evaluation of the Three Saccade Programming Models

For saccade programming, Fig. 6 shows the criterion of correct fixations R_c as a function of fixation order with five α values (0.5°, 1°, 2°, 4° and infinity) for the three models. R_c of the first fixation is identical for these three models because of the same starting foveated point. It is also the case for R_c of the second fixation in "1M" and "2M", and R_c of the third fixation in "2M" and "4M". R_c for all three models has the same global trend: decrease with fixation order. The R_c ratio is greater for the "4M" model in comparison to the "1M" model for all fixations. It results from the reinitialization of the foveated point. In the "2M" model, an intermediate of the two previous ones, the increase of R_c from the second to third fixation is also explained by this reinitialization. Moreover, the decrease at the second and fourth fixation (in "2M") presents necessity of the reinitialization, but at each fixation. The slower decrease of R_c in "4M" is also coherent with this interpretation.

The α parameter also influences the quality of the predicted regions. Saliency models including spatially variant retinal resolution give better results than the model with constant resolution for the first four fixations (t-test, $p < 0.005$, except the cases $\alpha = 0.5°$ and $\alpha = 1°$ for the third fixation). However, among $\alpha = 0.5°$, $\alpha = 1°$, $\alpha = 2°$ and $\alpha = 4°$ there is no significant difference except for the first fixation. At this fixation, there is no difference between R_c of cases 0.5° and 1° but they are significantly greater than those in cases 2° and 4° ($F(3,1612) = 10.45$, $p < 0.005$).

3.3.2 Configuration of the Eccentricity-Dependent Filter

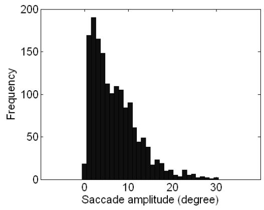
To evaluate the influence of the spatially variant retinal resolution on the saccade distribution, the same five α values are tested with the "4M" model which fits best the human fixations. We observe a great influence of α on saccade distribution. Figure 7 presents the experimental saccade distribution and those of the "4M" model according to the five α values. The distribution is very narrow with $\alpha = 0.5°$ (Fig. 7b), larger with greater α, and tends to a uniform distribution when α tends to infinity (Fig. 7f). The effect of the spatially variant retinal resolution can be shown through these distributions, concerning not only the dispersion but also the position of the maximum mode. Let us consider only two characteristics of the saccade distribution : form and position



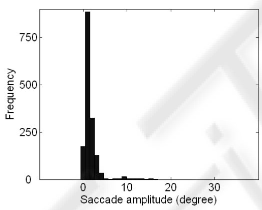
(a) "1M" (b) "2M" (c) "4M"

Figure 6: R_c as a function of fixation order of the three models with different α values.

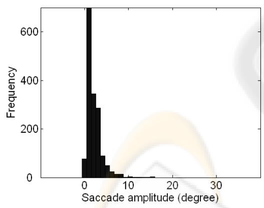
of the maximum mode. By varying α, we can adapt either distribution form or mode position. However, it is difficult to fit at the same time both form and mode position just only by adjusting one parameter (α here). Hence, the best parameter α is chosen as the best qualitative compromise between the distribution form and mode position. Among these α values, according to the mode position, the distribution with $\alpha = 2^\circ$ has the same mode position of about 2.1° as the experimental distribution. This distribution decreases progressively while eccentricity increases.



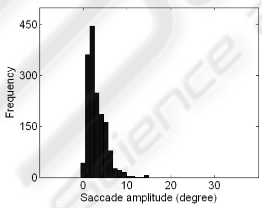
(a) Human



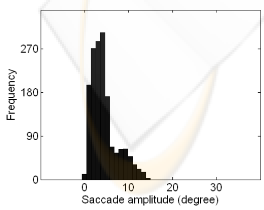
(b) "4M" with $\alpha = 0.5^\circ$



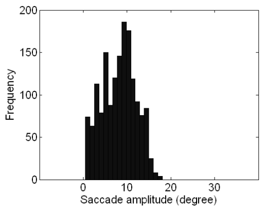
(c) "4M" with $\alpha = 1^\circ$



(d) "4M" with $\alpha = 2^\circ$



(e) "4M" with $\alpha = 4^\circ$



(f) "4M" with $\alpha = Inf$

Figure 7: Saccade distribution of the experimental data and the "4M" model according to different values of α.

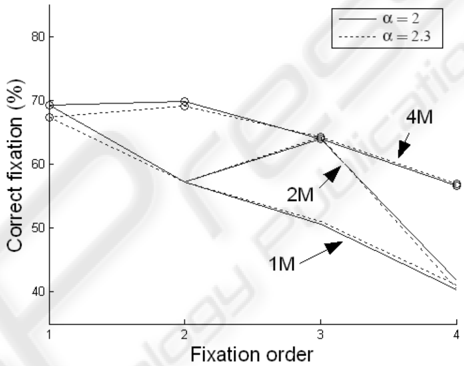


Figure 8: Comparison of R_c between $\alpha = 2^\circ$ and $\alpha = 2.3^\circ$ for the three models.

Indeed, when varying the α parameter, we notice a continuous effect both on the ratio R_c of "correct fixation" and saccade distribution. The results from the simulations with $\alpha = 2.3^\circ$ are very close to those obtained with $\alpha = 2^\circ$ (Fig. 8). By lack of space, the saccade distribution of the "4M" model for $\alpha = 2.3^\circ$, which is almost the same as for $\alpha = 2^\circ$, is not shown.

4 DISCUSSION

Firstly, all three models have the ratio of correct fixation decreasing according to the fixation order. This fact can be explained by the influence of top-down mechanisms which arise late and reduce the role of bottom-up in visual attention. In reality, as time passes, fixations of different subjects are more dispersive and subject-dependent. However, the influence of bottom-up still persists and the percentage of correct fixations is much higher than by chance in all three models. This confirms the role of bottom-up in visual attention even with the increasing presence of top-down.

The result of this study seems not to support programming of several saccades in parallel. In fact, we tested three models: the "1M" model programming four saccades in parallel from the same foveated point, the "2M" model programming two saccades in parallel and the "4M" model programming only one saccade at a time. This study showed that the "1M" model is not realistic and the "4M" model seems to be the most realistic. The behaviour of the "2M" model illustrates that recomputing saliency in updating the foveated point is beneficial. The best performance comes however from the "4M" model presenting effectiveness of the reinitialization at each fixation. We can conclude that saccade programming in parallel seems not to be used by subjects when they have to look freely at natural images.

Secondly, this study shows the positive effect of the spatially variant retinal resolution on the prediction quality. Whatever the saccade programming strategy is, the models including the spatially variant retinal resolution greatly outperform the models with constant resolution in terms of the quality of fixation prediction and saccade distribution. The parameter α which controls the resolution decrease has an important impact on saccade distribution (dispersion and mode position). Moreover, we found the expected range of value for this parameter using our model to compute saliency maps. We also notice that if we have the same mode position, the dispersion of the saccade distribution remains smaller on predicted data than on experimental data, as we only consider a bottom-up model and we have only one parameter to adjust.

In our models, a foveated point for the next saccade is selected from subjects' fixations instead of being looked for in the present saliency map. While fixations are different from one subject to another, the model is a subject-dependent model. If we want to go further in creating a more general model of predicting eye movements automatically, the model would take into account human task, for example categorization or information search, and hence passes from a region-predicting model to a scanpath-predicting model.

ACKNOWLEDGEMENTS

This work is partially supported by grants from the Rhône-Alpes Region with the LIMA project. T. Ho-Phuoc's PhD is funded by the French MESR.

REFERENCES

Beaudot, W., Palagi, P., and Herault, J. (1993). Realistic simulation tool for early visual processing including space, time and colour data. In *International Workshop on Artificial Neural Networks, LNCS*, volume 686, pages 370–375, Barcelona. Springer-Verlag.

Egeth, H. E. and Yantis, S. (1997). Visual attention: Control, representation, and time course. In *Annual Review of Psychology*, volume 48, pages 269–297.

Geisler, W. S. and Perry, J. S. (1998). A real-time foveated multiresolution system for low-bandwidth video communication. In *Human Vision and Electronic Imaging, Proceedings of SPIE*, volume 3299, pages 294–305.

Hansen, T., Sepp, W., and Neumann, H. (2001). Recurrent long-range interactions in early vision. In S. Wermter, J. A. and Willshaw, D., editors, *Emergent Neural Computational Architectures Based on Neuroscience, LNCS/LNAI*, volume 2036, pages 139–153.

Henderson, J. M. (2003). Human gaze control in real-world scene perception. In *Trends in Cognitive Sciences*, volume 7, pages 498–504.

Itti, L. (2006). Quantitative modeling of perceptual salience at human eye position. In *Visual Cognition*, volume 14, pages 959–984.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 20, pages 1254–1259.

Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. In *Human Neurobiology*, volume 4, pages 219–227.

McPeek, R. M., Skavenski, A. A., and Nakayama, K. (1998). Adjustment of fixation duration in visual search. In *Vision Research*, volume 38, pages 1295–1302.

McPeek, R. M., Skavenski, A. A., and Nakayama, K. (2000). Concurrent processing of saccades in visual search. In *Vision Research*, volume 40, pages 2499–2516.

Navon, D. (1977). Forest before trees: the precedence of global features in visual perception. In *Cognitive Psychology*, volume 9, pages 353–383.

Parkhurst, D., Law, K., and Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. In *Vision Research*, volume 42, pages 107–123.

Perry, J. S. (2002). http://fi.cvis.psy.utexas.edu/software.shtml.

Torralba, A., Oliva, A., Castelhano, M. S., and Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. In *Psychological Review*, volume 113, pages 766–786.

Treisman, A. and Gelade, G. (1980). A feature integration theory of attention. In *Cognitive Psychology*, volume 12, pages 97–136.

Wandell, B. A. (1995). *Foundations of Vision*. Stanford University.