EXPLICIT UNCERTAINTY FOR EYE MOVEMENT SELECTION

F. Colas,[1*]   F. Flacher,[1]   P. Bessière,[2]   B. Girard[1]

1 – Laboratoire de Physiologie de la Perception et de l'Action   2 – Laboratoire d'Informatique de Grenoble
CNRS - Collège de France   CNRS - Grenoble Universités
11 place Marcelin Berthelot   655 avenue de l'Europe
75231 Paris Cedex 05   38334 Montbonnot

ABSTRACT

In this paper, we consider the issue of the selection of eye movements in an eye-free Multiple Object Tracking task. We propose a Bayesian model of retinotopic maps with a complex logarithmic mapping. This model is structured in two parts: a representation of the visual scene, and a decision model based on the representation. We compare different decision models based on different features of the representation and we show that taking into account uncertainty helps predict the eye movements of subjects recorded in a psychophysics experiment.

KEY WORDS

Bayesian modelling, eye movements, retinotopic maps.

1   Introduction

In this study, we investigate the possible role of uncertainty evaluation in selection processes related to active perception. Uncertainty is the consequence of the inverse nature of perception, as well as incompleteness of the models. We choose to handle and reason with it using the Bayesian Programming framework [1]. We use an eye-free version of the standard Multiple Object Tracking (MOT) paradigm [2] as a basic selection task. In MOT, the subject is presented a number of moving objects, some of which are targets while the others are distractors. The targets are cued at the beginning of each trial, the subject has then to remember where the targets are, while all objects move, and to designate the targets at the end of the trial.

We design Bayesian models computing a sequence of probability distributions over the next eye movement to perform, based on a sequence of observations of objects in the visual field. They are inspired by the anatomy and electrophysiology of eye-movement selection related brain regions. These regions (fig. 1), the superior colliculus (SC), the frontal eye fields (FEF) and the lateral bank in the intraparietal sulcus (LIP) have a number of common points. They all receive information concerning the position of points of interest in the visual field (visual activity), memorize them (delay activity) and can generate movements towards them (motor activity) [4, 5, 6].
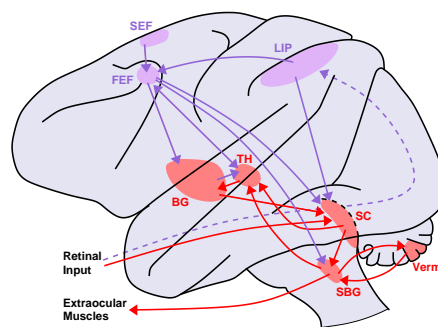
Figure 1: Premotor and motor circuitry shared by saccade and smooth pursuit movements (Macaque monkey). In red, short subcortical loop, in purple, long cortical loop. The dashed arrow stands for the cortical pathway including notably the lateral geniculate nucleus and the visual cortex. BG: basal ganglia, FEF: frontal eye fields, LIP: lateral bank of the intraparietal sulcus, SBG: saccade burst generators, SC: superior colliculus, SEF: supplementary eye fields, TH: thalamus, Verm: cerebellar vermis. Adapted from [3].

These positions are encoded by topographically organized cells, with receptive/motor fields defined in retinotopic reference frames. In the SC of primates, these maps have a complex logarithmic mapping [7, 8], which is represented on fig. 2 by the blue lines (plain lines: iso-eccentricities; dotted lines: iso-directions). Concerning the FEF, the eccentricity of the position vector is encoded logarithmically [9], however the encoding of direction is not well understood yet. Finally, the structure of the LIP maps is still to be deciphered, but a continuous topographical organization seems to exist, with an over representation of the central visual field [10]. We thus use the primate SC maps geometry in our models, with the assumptions that human SC and cortical maps probably have a similar geometry.

The spatial working memory-related neurons in SC [11], FEF [12] and LIP [13] are capable of dynamic remapping. They can be activated by a memory of the position of a target, even if the target was not in the cell's receptive field at the time of presentation. They behave as if they were part of a retinotopic memory map, where a remapping mechanism would allow the displacement of the memorized activity when an eye movement is performed. We include this remapping capability in the representation part of our models.

After having presented the structure of our models, we compare their movement predictions with recorded human movements and show that the explicit use of uncertainty improves the quality of the prediction.

2 Model

Our model has two stages: a *representation* of the visual field and the *decision* process of the next eye movement.

2.1 Representation

The representation model is a dynamic retinotopic map of the objects in the visual field. This representation is structured in two successive layers: the *occupancy* of the visual field, and a memory of the *position* of each target.

Occupancy of the visual field   The first part is structured like an occupancy grid, a recursive Bayesian filter introduced for obstacle representation in mobile robotics [14]. The environment is discretized into a regular grid $\mathcal{G}$ (with the logcomplex mapping) and we define a binary variable Occ_c^t in each cell c and for each time t that states whether or not there is an object in the corresponding location in the visual field. The input is introduced as a set of binary variables Obs_c^t. The observation and occupancy of each cell are linked by a probabilistic relation $P(Obs_c^t \mid Occ_c^t)$ that states it is likely to observe the assumed occupancy of the cell. As for all subsequent probability distributions that appear in our models, we give this probability distribution a parametrical form whose parameters we learn for part of the experimental data.

The remapping capability of this model relies on the current displacement Mvt^t and the distribution $P(Occ_c^t \mid Occ^{t-1} \, Mvt^t)$ that transfers the occupancy associated to antecedent cells to the corresponding present cell with an additional uncertainty factor.

Due to the high dimensionality of this representation space, we approximate the inference over the whole grid by a set of inferences for each cell c that depend only on a subset $\mathcal{A}(c)$ of antecedent cells c' for the current eye movement. Thus the update of the knowledge on occupancy in our model is recursively computed as follows:

$$P(Occ_c^t \mid Obs^{1:t} \, Mvt^{1:t}) \tag{1}$$
$$\propto \quad P(Obs_c^t \mid Occ_c^t)$$
$$\times \sum_{Occ_{\mathcal{A}(c)}^{t-1}} \left[\begin{array}{l} P(Occ_c^t \mid Mvt^t \, Occ_{\mathcal{A}(c)}^{t-1}) \\ \prod_{c'} P(Occ_{c'}^{t-1} \mid Obs^{1:t-1} \, Mvt^{1:t-1}) \end{array} \right]$$

Position of the targets   To introduce the discrimination between targets and distractors, we add a set of variables Tgt_i^t that represent the location of each target i at each time t. We also include remapping capability for the targets so that an eye movement Mvt^t updates the distribution on Tgt_i^t. This is done in a dynamic model

$P(Tgt_i^t \mid Tgt_i^{t-1} \, Occ^t \, Mvt^t)$ similar to the dynamic model of occupancy.

In addition to question 1, the knowledge over the targets is computed at each time step as follows:

$$P(Tgt_i^t \mid Obs^{1:t} \, Mvt^{1:t}) \tag{2}$$
$$\propto \sum_{Tgt_i^{t-1}} \left[\begin{array}{l} P(Tgt_i^{t-1} \mid Obs^{1:t-1} \, Mvt^{1:t-1}) \\ \times \sum_{Occ^t} \left[\begin{array}{l} P(Occ^t \mid Obs^{1:t} \, Mvt^{1:t}) \\ \times P(Tgt_i^t \mid Tgt_i^{t-1} \, Occ^t \, Mvt^t) \end{array} \right] \end{array} \right]$$

where the summation over the whole grid can be approximated as above, by separating the cells.

Both questions 1 and 2 are the current knowledge about the visual scene that can be inferred from the past observations and movements and the hypotheses of our model.

2.2 Decision

Based on this knowledge, we propose models that determine where to look next. We make the hypothesis that the representation model exposed above is useful for producing eye movements. To test this hypothesis, we compare one model that does not use this representation, *constant* model, with one that does, *target* model.

The main hypothesis is that uncertainty explicitly taken into account can help in the decision of eye movement. Thus we compare one model that does not take explicitly into account the uncertainty, target model, with one that does, *uncertainty* model.

Constant model   This model is a baseline for the other two. It is defined as the best static probabilistic distribution $P(Mot)$ that can account for the experimental eye movement. In this distribution, the probability for a given eye movement is equal to its experimental frequency. Thus we learned this distribution from our experimental data.

Target model   This second model determines its eye movements based on the location of the targets. It is a Bayesian fusion model with each target considered as the location where to look. It uses an inverse model $P(Tgt_i^t \mid Mot^t)$ that states that at time t the location of the target Tgt_i^t is probably near the eye movement Mot^t with a Gaussian distribution. Moreover, the prior distribution on the eye movement is taken from the constant model. Therefore, this target model refines the eye movement distribution with the influence of each targets.

As the exact locations of the targets are not known, this model takes into account the estimation from question 2 in the fusion. The actual eye movement distribution can be computed using the following expression:

$$P(Mot^t \mid Obs^{1:t} \, Mvt^{1:t})$$
$$\propto \quad P(Mot) \prod_{i=1}^{N} \sum_{Tgt_i^t} P(Tgt_i^t \mid Obs^{1:t} \, Mvt^{1:t}) P(Tgt_i^t \mid Mot^t)$$

Uncertainty model The behaviour of the previous model is influenced by uncertainty insofar as the incentive to look near a given target is higher for a more certain location of this target. As for any Bayesian model, uncertainty is handled as part of the inference mechanism: as a mean to describe knowledge.

In this third model, we propose to include uncertainty as a variable to reason about: as the knowledge to be described. The rationale is simply that it is more efficient to gather information when and where it lacks that is when and where there is more uncertainty.

Therefore, we introduce a new set of variables I_c^t representing an uncertainty index at cell c at time t. For this implementation, we choose to specify this uncertainty index as the probability distribution of occupancy in this cell. The nearer this probability is from $\frac{1}{2}$ the higher the uncertainty and the higher the probability to look there. In the end, this model computes the posterior probability distribution on next eye movement using the following expression:

$$P(Mot^t \mid Obs^{1:t}\ Mvt^{1:t}\ I^{1:t})$$
$$\propto\quad P(Mot^t \mid Obs^{1:t}\ Mvt^{1:t})P(I_{Mot^t}^t \mid Mot^t)$$

with $I_c^t = P(Occ_c^t \mid Obs^{1:t}\ Mvt^{1:t})$ (equation 1).

This model filters the eye movement distribution computed by the second model, in order to enhance the probability distribution in the locations of high uncertainty.

3 Results

As shown in figure 2, these models produce a probability distribution at each time step which is, except for the constant model, heavily dependent on the past observations and movements in the retinocentered reference frame. Therefore we first defined an appropriate tool to compare these model. Then we present the results of our models according to this evaluation method.

3.1 Comparison method

The generic Bayesian method to compare models (or parameters, that is formally the same issue) is to assess a prior probability distribution over the models, compute the likelihood of each model in view of the data, and use Bayes rule to obtain a probability distribution over the models: $P(Model \mid Data) \propto P(Model) \times P(Data \mid Model)$.

As deciding on priors is sometimes an arbitrary matter and this prior may have a negligible influence with a growing number of data points, a common approximation is simply comparing the likelihood of the models. Choosing the model with the highest likelihood is dubbed as *maximum likelihood estimation*.

As the decision models compute a probability distribution, we can compute, for each model at each time step, the probability of the actual eye movements recorded from subjects, as well as the probability of the whole set of recordings. In order not to have a measure that tends to
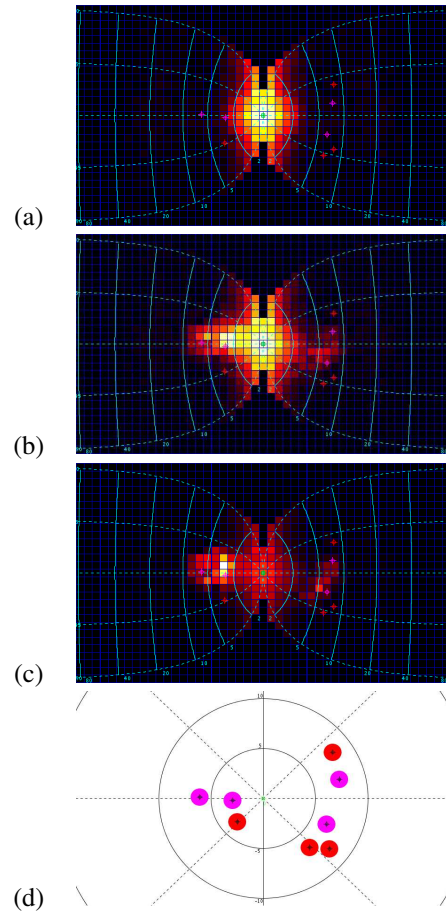
Figure 2: Example of probability distributions computed by each model in the same configuration. Panel (a) is the distribution of constant model. Panel (b) shows the probability distribution for the target model that shows a preference for the targets. Panel (c) shows the probability distribution for the uncertainty model that highlights some of the targets. Bottom panel shows the position of the targets (magenta) and objects (red) in the visual field.

zero as the number of trials increase, we choose the geometric mean of the likelihood across trials, as it tends to be independent on the number of trials. Thus we compare:

$$\sqrt[N]{\prod_{\tau=1}^{N}\prod_{t=1}^{T} P([Mot = mot_\tau^{t+1}])}$$

for the constant model,

$$\sqrt[N]{\prod_{\tau=1}^{N}\prod_{t=1}^{T} P([Mot = mot_\tau^{t+1}] \mid obs^{1:t}\ mot_\tau^{1:t})}$$

for the target model, and

$$\sqrt[N]{\prod_{\tau=1}^{N}\prod_{t=1}^{T} P([Mot = mot_\tau^{t+1}] \mid obs^{1:t}\ mot_\tau^{1:t}\ i^{1:t})}$$

for the uncertainty model, where mot_τ^t is the actual eye movement recorded in trial τ at time t.

3.2 Results

The data set is gathered from 11 subjects with 110 trials each for a total of 1210 trials (see [15] for details). Each trial was discretized in time in 24 observations for a grand total of 29040 data points. Part of the data set (124 random trials) was used to determine the 9 parameters of the model and the results are computed on the remaining 1089 trials.

Ratio	Constant	Target	Uncertainty
Constant	1	280	320
Target	3.5×10^{-3}	1	1.14
Uncertainty	3.1×10^{-3}	0.87	1

Table 1: Ratio of the measures for pair of models.

Table 1 presents the results of our three decision models for this data set. It shows that the model that generates motion with the empiric probability distribution but without the representation layer is far less probable than the other two (by respectively a factor 280 and 320). This shows that, as expected, the representation layer is useful in deciding the next eye movement.

Table 1 further shows that the model taking explicitly into account uncertainty is better than the model that does not by 14%. This is in favor of our hypothesis that taking explicitly into account uncertainty is helpful in deciding the next eye movement.

It should be noted that the choice of the geometric mean prevents the ratio of our models to raise exponentially as the number of trials grows. In our case, the likelihood ratio between the model with explicit uncertainty and the one without is 4.9×10^{63}. With half the trials, this likelihood ratio is the square root, that is only 7.0×10^{31}. We preferred presenting the results with a measure independent of the number of trials.

Moreover, this model is designed to study the criteria of eye movement selection and not to solve the task autonomously. Nevertheless, we can add simple algorithms to decide the eye movement at each time step and to point out the targets at the end of the trial, based on the state of the representation layer. A naive implementation yielded a chance level performance. We did not have the opportunity to delve deeper into this question, as this is not the aim of our model, but this result could be improved by a more accurate target layer.

4 Conclusion and discussion

As a conclusion, we propose a Bayesian model with two parts: a representation of the visual scene, and a decision model based on the state of the representation. The representation both tracks the occupancy of the visual scene as well as the locations of the targets.

Based on this representation, we tested several decision models and we have shown that the model that takes explicitly into account the uncertainty better fitted the eye movements recorded from subjects participating a psychophysics experiment.

Moreover, the eye movement frequency shows that, most of the times, the eye movements are of low amplitude, indicating either fixation or slow pursuit of an object. In these cases, the constant model has a likelihood comparable with or even sometimes greater than the other two. Thus the difference is due to the saccadic events, when the target and uncertainty model have a good likelihood contrary to the constant one that assign a lower probability as the eccentricity grows.

The difference between the target model and the uncertainty model, on the other hand is due to the filtering of the eye movements distribution from the target model by the uncertainty. The difference is less important than for the constant model as the uncertainty associated to the targets are often similar (isolated targets with comparable movement profiles). It could be interesting to enrich the stimulus in order to manipulate uncertainty more precisely.

Acknowledgements

References

[1] Olivier Lebeltel, Pierre Bessière, Julien Diard, and Emmanuel Mazer. Bayesian robots programming. *Autonomous Robots*, 16(1):49–79, 2004.

[2] Z.W. Pylyshyn and R.W. Storm. Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, 3(3):1–19, 1988.

[3] R.J. Krauzlis. Recasting the Smooth Pursuit Eye Movement System. *Journal of Neurophysiology*, 91(2):591–603, 2004.

[4] A.K. Moschovakis, C.A. Scudder, and S.M. Highstein. The microscopic anatomy and physiology of the mammalian saccadic system. *Prog Neurobiol*, 50:133–254, 1996.

[5] R.H. Wurtz, M.A. Sommer, M. Paré, and S. Ferraina. Signal transformation from cerebral cortex to superior colliculus for the generation of saccades. *Vision Res*, 41:3399–3412, 2001.

[6] C.A. Scudder, C.R.S. Kaneko, and A.F. Fuchs. The brainstem burst generator for saccadic eye movements. A modern synthesis. *Exp Brain Res*, 142:439–462, 2002.

[7] D.A. Robinson. Eye movements evoked by collicular stimulation in the alert monkey. *Vision Res*, 12:1795–1808, 1972.

[8] F.P. Ottes, van Gisbergen J.A., and J.J. Eggermont. Visuomotor fields of the superior colliculus: a quantitative model. *Vision Res*, 26(6):857–873, 1986.

[9] M.A. Sommer and R.H. Wurtz. Composition and topographic organization of signals sent from the frontal eye fields to the superior colliculus. *Journal of Neurophysiology*, 83:1979–2001, 2000.

[10] S. Ben Hamed, J.-R. Duhamel, F. Bremmer, and W. Graf. Representation of the visual field in the lateral intraparietal area of macaque monkeys: a quantitative receptive field analysis. *Experimental Brain Research*, 140:127–144, 2001.

[11] L.E. Mays and D.L. Sparks. Dissociation of visual and saccade-related responses in superior colliculus neurons. *J Neurophysiol*, 43(1):207–232, 1980.

[12] M.E. Goldberg and C.J. Bruce. Primate frontal eye fields. III. maintenance of a spatially accurate saccade signal. *Journal of Neurophysiology*, 64(2):489–508, 1990.

[13] J.W. Gnadt and R.A. Andersen. Memory related motor planning activity in the posterior arietal cortex of the macaque. *Experimental Brain Research*, 70(1):216–220, 1988.

[14] A. Elfes. *Occupancy grids: a probabilistic framework for robot perception and navigation*. PhD thesis, Pittsburgh, PA, USA, 1989.

[15] T.G. Tanner, L.H. Canto-Pereira, and H.H. Bülthoff. Free vs. constrained gaze in a multiple-object-tracking-paradigm. In *30th European Conference on Visual Perception*, Arezzo, Italy, August 2007.