Biol Cybern (2009) 100:203–214
DOI 10.1007/s00422-009-0292-y

ORIGINAL PAPER

Biological Cybernetics

Bayesian models of eye movement selection with retinotopic maps

Francis Colas · Fabien Flacher · Thomas Tanner · Pierre Bessière · Benoît Girard

Received: 31 July 2008 / Accepted: 9 January 2009 / Published online: 11 February 2009

Abstract Among the various possible criteria guiding eye movement selection, we investigate the role of position uncertainty in the peripheral visual field. In particular, we suggest that, in everyday life situations of object tracking, eye movement selection probably includes a principle of reduction of uncertainty. To evaluate this hypothesis, we confront the movement predictions of computational models with human results from a psychophysical task. This task is a freely moving eye version of the multiple object tracking task, where the eye movements may be used to compensate for low peripheral resolution. We design several Bayesian models of eye movement selection with increasing complexity, whose layered structures are inspired by the neurobiology of the brain areas implied in this process. Finally, we compare the relative performances of these models with regard to the prediction of the recorded human movements, and show the advantage of taking explicitly into account uncertainty for the prediction of eye movements.

Keywords Bayesian modeling · Retinotopic maps · Eye movements selection · Multiple-object tracking

F. Colas (✉) · F. Flacher · B. Girard
Laboratoire de Physiologie de la Perception et de l'Action,
CNRS/Collège de France, 11 pl. Marcelin Berthelot,
75231 Paris Cedex 05, France
e-mail: colas.francis@gmail.com

F. Flacher
e-mail: fabien.flacher@gmail.com

B. Girard
e-mail: benoit.girard@college-de-france.fr

T. Tanner
Department of Cognitive and Computational Psychophysics,
MPI for Biological Cybernetics, Spemannstr. 38,
72076 Tübingen, Germany
e-mail: tanner@tuebingen.mpg.de

P. Bessière
Laboratoire d'Informatique de Grenoble, CNRS/Grenoble Universités,
655 av. de l'Europe, 38334 Montbonnot, France
e-mail: bessiere@imag.fr

1 Introduction

We usually make a few saccades per seconds. Saccades, and other eye movements, may result from a decision on where to look next, in order to gain information about the visual scene by driving the fovea towards regions of interest. Indeed, as the sensitivity and spatial resolution of the retina decays towards the periphery of the visual field, we are uncertain about the accuracy of what we perceive in the periphery and about what we can expected to learn from an eye movement towards a peripheral position. The uncertainty is a common issue for both perception—because we cannot be sure of what we perceive—and action—because we cannot be sure of the consequences of our actions. In this paper, we investigate the possible role of uncertainty evaluation in selection processes related to active perception. We build a Bayesian model inspired by the neurophysiology of eye movement selection related brain regions, in order to investigate eye movements selection during freely moving eye multiple object tracking task (MOT).

1.1 Bayesian methodology

In order to handle uncertainty and to explicitly reason about it, we use the Bayesian Programming framework (Lebeltel et al. 2004; Bessière et al. 2008). This framework provides a systematic procedure to build and use a Bayesian model. Such a model uses probability distributions to

represent knowledge with uncertainty. It then reasons about this knowledge by applying the rules of probability theory. More precisely, starting from a joint probability distribution, marginalization and Bayes' rules allow to compute any conditional or marginal probability distribution. As this joint probability is usually of very high dimensionality, we use conditional independence hypotheses to decompose the joint distribution in a simpler product of smaller distributions.

In the end, a Bayesian programmer specifies a set of variables, a *decomposition* of the joint probability distribution and a mathematical expression for each factor that appears in this decomposition. At that point, any distribution on the variables can be computed. The programmer is usually interested on one particular distribution, which is called a *question*. The inference can be automatically computed through the use of both marginalization and Bayes rules.

1.2 Eye movement circuitry

Even if we do not have the pretension to build a complete model of the neurophysiology of the brain regions related to eye movement selection, the structure of our model is inspired by their anatomy and electrophysiology. Saccadic and smooth pursuit circuitry share a large part of their functional architecture (Krauzlis 2004). Among those regions containing saccadic and smooth pursuit subcircuits (Fig. 1), the superior colliculus (SC), the frontal eye fields (FEF) and the lateral bank of the intraparietal sulcus (LIP) in the posterior parietal cortex have a number of common points. They all receive information concerning the position of points of interest in the visual field (visual activity), memorize these positions (delay activity) and are implied in the selection of the gaze targets among these points (presaccadic activity) (Moschovakis et al. 1996; Wurtz et al. 2001; Scudder et al. 2002). These positions are encoded by cells with receptive/motor fields defined in a retinotopic reference frame. Our model is based on retinotopic probability distributions encoding similar information (observations, memory of target positions, motor decision).

In the SC, these cells are clearly organized in topographic maps, in various species (Robinson 1972; McIlwain 1976, 1983; Siminoff et al. 1966; Herrero et al. 1998). In primates, these maps have a complex logarithmic mapping (Fig. 2) (Robinson 1972; Ottes et al 1986), similar to the mapping found in the striate cortex (Schwarz 1980). Concerning the FEF, mapping studies clearly show a logarithmic encoding of the eccentricity of the position vector (Sommer and Wurtz 2000), however complementary studies are necessary to understand how its orientation is encoded. Finally, the structure of the LIP maps is still to be deciphered, even if a continuous topographical organization seems to exist, with an over representation of the central visual field (Ben Hamed et al. 2001). Given the lack of quantitatively defined FEF and

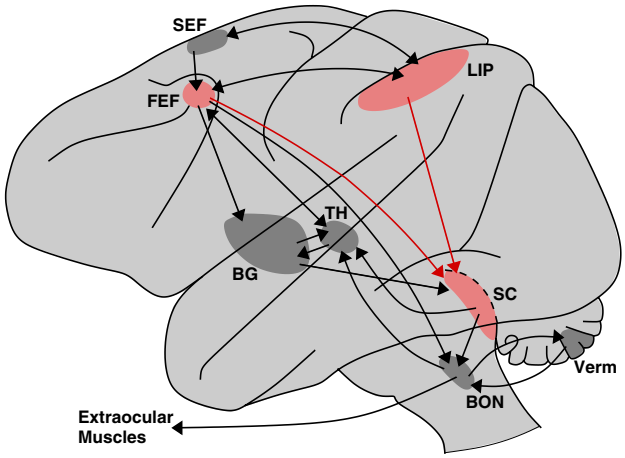

Fig. 1 Premotor and motor circuitry shared by saccade and smooth pursuit movement (Macaque monkey). *BG* basal ganglia, *BON* brainstem oculomotor nuclei, *FEF* frontal eye fields, *LIP* lateral bank of the intraparietal sulcus, *SC* superior colliculus, *SEF* supplementary eye fields, *TH* thalamus, *Verm* cerebellar vermis. In light *red* regions using retinotopic reference frames to encode visual, memory and motor activity, refer to text for more details. Adapted from (Krauzlis 2004) (color in online)

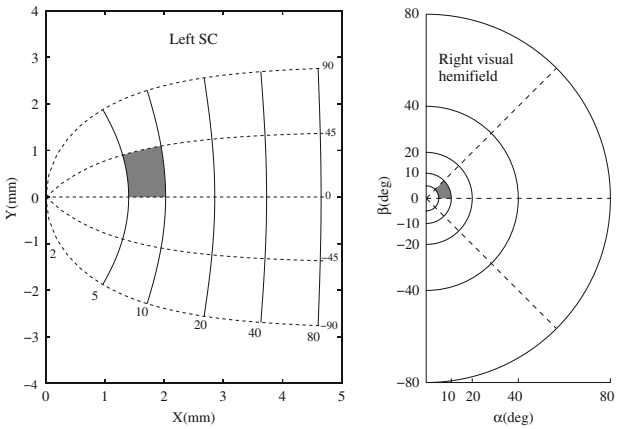

Fig. 2 Macaque collicular mapping. The angular position of targets in the visual field (*right*) are mapped onto the SC surface (*left*) using a logarithmic mapping. The *grey* areas represent the same part of the visual field in both representations

LIP mappings, we assume that they share similar properties with the SC one and thus use the log complex mapping of the SC for all the position encoding variables of our model.

The neurons related to the spatial working memory in SC (Mays and Sparks 1980), FEF (Goldberg and Bruce 1990) and LIP (Gnadt and Andersen 1988; Barash et al. 1991a,b)—also called quasi-visual cells or QV—are capable of dynamic remapping. These cells 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 included in a retinotopic memory map, integrating a remapping mechanism allowing the displacement of the

memorized activity when an eye movement is performed. Neural network models of that type of maps, either in the SC or the FEF, have already been proposed (Droulez and Berthoz 1991; Bozis and Moschovakis 1998; Mitchell and Zipser 2003). Such a mechanism, adapted to Bayesian programming, is used in the representation and memory layers of our model.

To summarize, though not strictly neuromimetic, the layered structure of our Bayesian model is based on log complex retinotopic maps with remapping capabilities, encoding the filtered visual input, the memorized position of targets of interests, and the generation of motor commands.

1.3 Experimental protocol

In order to study selection of eye movement in a controlled task, we use eye movement recordings from a freely moving eye version (Tanner et al. 2007) of the classical MOT task (Pylyshyn and Storm 1988). Eye movements in MOT have only recently attracted interest (Tanner et al. 2007; Fehd and Seiffert 2008; Zelinsky and Neider 2008). The original task was designed to investigate the distribution of covert attention with eye movements constrained by a fixation cross (Cavanagh and Alvarez 2005), while we looked at how free eye movements might optimize the tracking. Figure 3 illustrates this experiment in which participants are presented with a set of targets among a number of distractors. All of these objects are indiscernible 1° large discs and move in a quasi-random pattern. The task is to remember which of these objects are the targets (see Appendix A for a complete description). With this experimental paradigm, the visual scene is composed of simple geometric features therefore allowing for a study of the eye movement selection that occurs in this context.

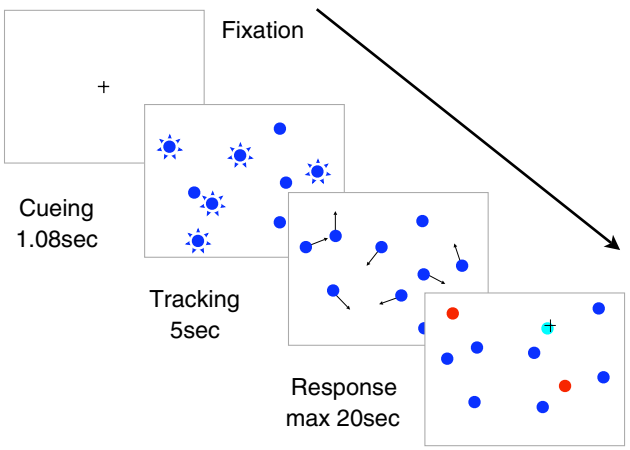

Fig. 3 Typical multiple object tracking experiment. A set of simple objects is presented, the targets are identified as the flashing ones, then the flashing stops and all the objects move around independently. After they stop moving, the subject must identify the targets

First we describe the Bayesian models we propose. Then we present the global results indicating that uncertainty is useful and some specific situations shedding light on the differences between the models.

2 Methods

The model we propose is composed of two parts. The first part deals with the perception and memory of the visual scene (*representation* model). The second part deals with the actual selection of where to look next (*decision* model).

Both models are expressed in a retinal reference frame, with a logcomplex mapping as explained above.

2.1 Representation

The representation part of our model is a dynamic retinotopic map of the visual environment. This representation is structured in two different layers. The first layer is concerned only with the integration of the visual input, i.e. the occupancy of the visual scene without any discrimination between targets and distractors (*occupancy grid*). This model would be homologous to the visual cells.

The second layer is a memory of the position of the targets, reminiscent of the QV cells. It represents the knowledge of the observer about the position of the targets, based on the occupancy representation.

2.1.1 Occupancy grid

Occupancy grids are a standard way to represent the state of an environment. They were originally introduced for the representation of obstacles in robotics applications (Elfes 1989). The general idea is to discretize the environment into a grid and to assign a variable in each cell of the grid stating whether there is an obstacle or not. The occupancy grid is therefore the collection of probability distributions over each variable in the grid.

We apply this model to the presence of objects in the visual field. More precisely, we introduce a collection O of binary variables $O^t_{(x,y)}$, one for each timestep $t \in [\![0, t_{\max}]\!]$ and location $(x, y) \in \mathcal{G}$ where $\mathcal{G}$ is a regular grid in the retino-centered logcomplex reference frame.[1] We also assume that we have visual inputs in this same reference frame, represented by a collection V of binary variables $V^t_{(x,y)}$ for $t \in [\![1, t_{\max}]\!]$ indicating if an object (either target or distractor) is perceived in the corresponding cell. Finally, we include some past eye movement information M^t in order to model the remapping

[1] Omission of an index or exponent in the variable name indicates the conjunction of all of those variables for the missing index varying in its full range: $O = O^{0\to t_{\max}} = \bigwedge_{t=0}^{t_{\max}} O^t = \bigwedge_{t=0}^{t_{\max}} \bigwedge_{(x,y)\in\mathcal{G}} O^t_{(x,y)}$.

capability exhibited by cortical and subcortical retino-centered memories.

We write the joint probability distribution over all these variables by assuming the occupancy of the cells are independent one from another conditionally to the past eye movement and the former state of the grid. We also assume that the observation corresponding to a cell is independent on all other variables conditionally to the current occupancy in this cell. This is summarized by the following factorization of the joint distribution:

$$
\begin{aligned}
&P(OVM)\\
&\quad = P(O^0)\prod_{t=1}^{t_{\max}} P(O^t\ V^t\ M^t \mid O^{t-1})\\
&\quad = \prod_{(x,y)\in\mathcal{G}} P(O^0_{(x,y)})\\
&\qquad \times \prod_{t=1}^{t_{\max}} \left[\begin{array}{l} P(M^t)\\ \times \prod_{(x,y)\in\mathcal{G}} \left[\begin{array}{l} P(O^t_{(x,y)} \mid M^t\ O^{t-1})\\ \times P(V^t_{(x,y)} \mid O^t_{(x,y)}) \end{array}\right] \end{array}\right]
\end{aligned}
$$

In this expression, $P(O^0_{(x,y)})$ is an arbitrary prior on the occupancy of the visual scene, $P(M^t)$ is a distribution over the eye movement that can be chosen arbitrarily as the results of the inference do not depend on it, provided that it is not zero for the actual eye movements observed. The relation between the occupancy and the observation, $P(V^t_{(x,y)} \mid O^t_{(x,y)})$, is a simple probability matrix chosen to state that there is a high probability of observing an object when there is one and conversely of not observing anything when there is nothing.

The evolution of the grid, with the remapping capability, is specified by the transition model, $P(O^t_{(x,y)} \mid M^t\ O^{t-1})$, which essentially transfers the probability associated to antecedent cells for the given eye movements to the corresponding present cell with an additional uncertainty factor (see Appendix B.1 for details).

With this description, updating the knowledge over the occupancy of the visual field corresponds to the following question for each time t:

$$P(O^t \mid V^{1\to t}\ M^{1\to t}) \tag{1}$$

where $V^{1\to t}$ is the conjunction of all variables V^u for $u \in [\![1, t]\!]$. This expression can be computed in an iterative manner using Bayesian inference:

$$
\begin{aligned}
&P(O^t \mid V^{1\to t}\ M^{1\to t})\\
&\quad \propto \prod_{(x,y)\in\mathcal{G}} P(V^t_{(x,y)} \mid O^t_{(x,y)})\\
&\qquad \times \sum_{O^{t-1}} \left[\begin{array}{l} \prod_{(x,y)\in\mathcal{G}} P(O^t_{(x,y)} \mid M^t\ O^{t-1})\\ \times P(O^{t-1} \mid V^{1\to t-1}\ M^{1\to t-1}) \end{array}\right]
\end{aligned}
$$

However, this expression comprises a summation over all possible grid states, which is computationally intensive. Therefore we approximate the inference over the whole grid by a set of inferences for each cell that depend only on a subset of the grid:

$$
\begin{aligned}
&P(O^t_{(x,y)} \mid V^{1\to t}\ M^{1\to t})\\
&\quad \propto P(V^t_{(x,y)} \mid O^t_{(x,y)})\\
&\qquad \times \sum_{O^{t-1}_{\mathcal{A}(x,y)}} \left[\begin{array}{l} P(O^t_{(x,y)} \mid M^t\ O^{t-1}_{\mathcal{A}(x,y)})\\ \times \prod_{\mathcal{A}(x,y)} P(O^{t-1}_{(x',y')} \mid V^{1\to t-1}\ M^{1\to t-1}) \end{array}\right]
\end{aligned}
$$

where $\mathcal{A}(x, y)$ is the subset of the cells (x', y') of the grid that are the antecedent of the cell (x, y) by the current eye movement M^t.

2.1.2 Positions of the targets

The previous model describes the visual scene without differentiating between targets and distractors. In order to take this two classes into account, we add a set of variables T^t_i to represent the location of each target $i \in [\![1, N]\!]$ at each time $t \in [\![0, t_{\max}]\!]$ in the logcomplex retino-centered reference frame.

This representation is the standard way to represent the location of some objects and serves a different purpose than the occupancy grid, which is only the representation of the visual scene.

The model is extended with this additional variables by adding a new factor in the joint distribution, $P(T^t_i \mid T^{t-1}_i\ O^t\ M^t)$, that represents the dynamic model of targets:

$$
\begin{aligned}
&P(OVMT)\\
&\quad = \prod_{(x,y)\in\mathcal{G}} P(O^0_{(x,y)}) \prod_{i=1}^{N} P(T^0_i)\\
&\qquad \times \prod_{t=1}^{t_{\max}} \left[\begin{array}{l} P(M^t)\\ \times \prod_{(x,y)\in\mathcal{G}} \left[\begin{array}{l} P(O^t_{(x,y)} \mid M^t\ O^{t-1})\\ \times P(V^t_{(x,y)} \mid O^t_{(x,y)}) \end{array}\right]\\ \times \prod_{i=1}^{N} P(T^t_i \mid M^t\ O^t\ T^{t-1}_i) \end{array}\right]
\end{aligned}
$$

The additional factors $P(T^0_i)$ are priors over the positions of the targets that can be set according to the starting position of the targets as shown in the cueing phase.

The dynamic model of targets is very similar to the dynamic model of objects but with the occupancy grid on objects as observation (see Appendix B.2 for details).

At each time step, the relevant state of the representation can be summarized by the following question for each target $i \in [\![1, N]\!]$ at each timestep $t \in [\![1, t_{\max}]\!]$:

$$P(T^t_i \mid V^{1\to t}\ M^{1\to t}) \tag{2}$$

Bayesian inference leads to the following expression for this question:

$$P(T_i^t \mid V^{1\to t}\ M^{1\to t}) \propto \sum_{T_i^{t-1}} \left[\begin{array}{l} \sum_{O^t} \left[\begin{array}{l} P(T_i^t \mid M^t\ O^t\ T_i^{t-1}) \\ \times P(O^t \mid V^{1\to t}\ M^{1\to t}) \end{array} \right] \\ \times P(T_i^{t-1} \mid V^{1\to t-1}\ M^{1\to t-1}) \end{array} \right]$$

where $P(T_i^{t-1} \mid V^{1\to t-1}\ M^{1\to t-1})$ is the result of the same inference at the preceding timestep, $P(O^t \mid V^{1\to t}\ M^{1\to t})$ the result of question 1 at the same timestep. The summation of the whole grid, which is still computationally intensive, 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 models

Based on this knowledge, the observer has to decide where to look next in order to solve the task. We propose different models in order to test different hypotheses. First, we make the hypothesis that this representation model is useful for producing eye movements. To test this hypothesis, we compare a model that does not use the representation with one that does.

Then, the main hypothesis is that uncertainty, explicitly taken into account, can help in the decision of eye movement. Therefore, we compare a model that does not take into account explicitly the uncertainty with one that does.

In the end, we need to specify three models: one that does not use the representation model (π_A), one that uses the representation model without explicitly taking into account uncertainty (π_B), and finally one that uses the representation model and explicitly takes into account uncertainty (π_C). Each model π_k will infer a probability distribution on the next eye movement represented by a new variable $C^t \in \mathcal{G}$ at each time $t \in [\![1, t_{\max}]\!]$:

$$P(C^t \mid V^{1\to t}\ M^{1\to t}\ \pi_k)$$

This variable is the model's homologue to the motor cells found in LIP, FEF and SC.

2.2.1 Constant model

This model is a baseline for the other models. We look for the best static probabilistic distribution that can account for the experimental eye movement. Formally it is specified as being independent on time and on the observations:

$$\forall t \in [\![1, t_{\max}]\!], \quad P(C^t \mid V^{1\to t}\ M^{1\to t}\ \pi_A) = P(C^t \mid \pi_A) = P(C^1 \mid \pi_A)$$

In these conditions, it can be shown that the best distribution $P(C^1 \mid \pi_A)$, according to the measure defined Sect. 3.1, assigns the probability of each individual discretized motion to be equal to its frequency in the experimental data.[2] Therefore, we learned this distribution from our experimental data, using only a randomly selected subset in order not to overfit our models.

2.2.2 Targets positions

The second model we propose uses the knowledge from the representation layer to determine its eye movements. More precisely, it tends to look at locations where targets are close to another, in a kind of fusion process. Its prior will follow the statistical distribution of eye movements and the likelihood will be based on the distributions on the targets location inferred in the representation layer.

The decomposition is as follows:

$$P(CVMT \mid \pi_B) = \prod_{t=1}^{t_{\max}} \left[\begin{array}{l} P(V^t\ M^t \mid \pi_B) \\ \times \prod_{i=1}^{N} P(T_i^t \mid V^{1\to t}\ M^{1\to t}\ \pi_B) \\ \times P(C^t \mid T^t\ \pi_B) \end{array} \right]$$

where:

- $P(V^t\ M^t \mid \pi_B)$ is an arbitrary prior that is not used in the inference,
- $P(T_i^t \mid V^{1\to t}\ M^{1\to t}\ \pi_B)$ is the result of inference 2,
- $P(C^t \mid T^t\ \pi_B)$ is the result of the inference in a fusion submodel over the targets that yields:

$$P(C^t \mid T^t\ \pi_B) \propto P(C^t \mid \pi_A) \prod_{i=1}^{N} P(T_i^t \mid C^t)$$

where $P(C^t \mid \pi_A)$ is the prior taken from the constant model and $P(T_i^t \mid C^t)$ a distribution centered on C^t that expresses a proximity between C^t and T_i^t (concretely a Gaussian distribution centered on C^t).

With this model, the distribution on eye movement can be computed with the following expression:

$$P(C^t \mid V^{1\to t}\ M^{1\to t}\ \pi_B) \propto P(C^t \mid \pi_A) \times \prod_{i=1}^{N} \sum_{T_i^t} \left[\begin{array}{l} P(T_i^t \mid V^{1\to t}\ M^{1\to t}\ \pi_B) \\ \times P(T_i^t \mid C^t) \end{array} \right]$$

[2] When restricted to time independence and assuming a uniform prior over such models, our measure is a multinomial likelihood which leads to a Dirichlet distribution according to the experimental frequencies. The maximum of this Dirichlet distribution is the histogram of the experimental frequencies.

In short, this model is the product between the prior on eye movement and each distribution on the targets convolved by a Gaussian distribution. This expression shows that this model is attracted towards the targets but without necessarily looking at one in particular as balance between the distributions on the targets can lead to a peak in some weighted sum of their locations.

2.2.3 Uncertainty model

The behavior of the preceding 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 than when and where there is less uncertainty.

Therefore, we introduce a new set of variables $I^t_{(x,y)} \in [0, 1]$, representing an index of the uncertainty at cell $(x, y) \in \mathcal{G}$ at time $t \in [\![1, t_{\max}]\!]$. Any index can fit as long as we can correlate the value of this uncertainty index with the actual uncertainty.

For simplification, we choose our uncertainty indices to be equal to this probability of occupancy, as we represent occupancy as binary variables. The relation between this uncertainty index (probability distribution) and uncertainty is such that a probability near $\frac{1}{2}$ represents a high uncertainty whereas a probability near 0 or 1 represent a low uncertainty. Other spaces can be chosen for these variables, such as entropy, but we keep the probability distribution to simplify our computations.

As mentioned above, this model is structured around a prior probability of motion which is filtered by these uncertainty variables in order to enhance the probability of eye movements towards uncertain regions. The prior probability is the result of the preceding model π_B.[3]

The decomposition of this model is as follows:

$$P(CVMI \mid \pi_C) = \prod_{t=1}^{t_{\max}} \left[\begin{array}{l} P(V^t\ M^t \mid \pi_C) \\ \times P(C^t \mid V^{1\to t}\ M^{1\to t}\ \pi_B) \\ \times \prod_{(x,y)\in\mathcal{G}} P(I^t_{(x,y)} \mid C^t\ \pi_C) \end{array} \right]$$

where:

- $P(V^t\ M^t \mid \pi_C)$ is an arbitrary prior that is not used in the inference,
- $P(C^t \mid V^{1\to t}\ M^{1\to t}\ \pi_B)$ is the result of the previous model,
- $P(I^t_{(x,y)} \mid C^t\ \pi_C)$ is a beta distribution that expresses that for a given eye movement proposal C^t, $I^t_{C^t}$ is more likely near $\frac{1}{2}$ and distribution on $I^t_{(x,y)}$ for $(x, y) \neq C^t$ is uniform.

This model computes the posterior probability distribution on next eye movement using the following expression:

$$P(C^t \mid V^{1\to t}\ M^{1\to t}\ I^{1\to t}\ \pi_C) \propto P(C^t \mid V^{1\to t}\ M^{1\to t}\ \pi_B) \times P(I^t_{C^t} \mid C^t\ \pi_C)$$

where:

$$\forall (x, y), t \in \mathcal{G} \times [\![1, t_{\max}]\!],\quad I^t_{(x,y)} = P(O^t_{(x,y)} \mid V^{1\to t}\ M^{1\to t})$$

as computed by Eq. 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

The output of our models is a probability distribution over the eye position at each timestep. For such complex objects, there are neither significance test nor an appropriate sensitivity analysis and the comparison is done using their respective likelihood. However the likelihood is highly dependent on the size of the data set. Therefore we first introduce a comparison method that does not depend on the size of the data set. Then we present their results and comment them with respect to the specific behavior of each model. Finally, we illustrate the main differences between the various models by giving examples of specific situations.

3.1 Comparison method

The decision models compute a probability distribution over the possible eye movements at one moment, based on past observations and their respective hypotheses (Fig. 4). We can therefore compute, for each model, the probability of the actual eye movements recorded from subjects in a given situation, as well as the probability of the whole set of recordings with an additional independency assumption.

Probability values are only relative measures as, when the possibilities are numerous, they tend to be very small. However, their comparison across models (which share the same number of possibilities) indicates which model is a better predictor of the recorded eye movements. This process is known as the *Maximum Likelihood* method.

[3] This is a matter of presentation of the model. The complete expression of π_C can be written without reference to model π_B but the addition of uncertainty would be less clear.

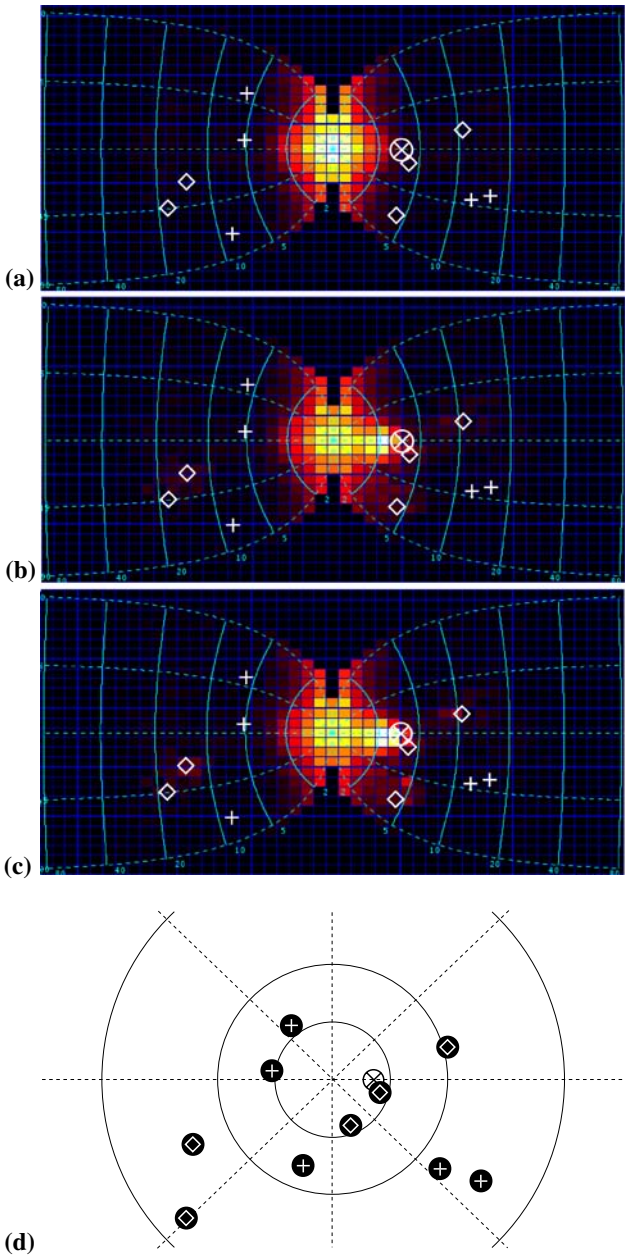

Fig. 4 Example of probability distributions computed by each decision model in the same configuration. The two halves of the representations are drawn side-by-side. The *plain lines* are the iso-eccentricities and the *dotted lines* are the iso-directions. The *brightness* of the cell indicates the probability of the associated eye movement: a dark cell for a low probability and a white cell for a high probability for the eye movement toward this cell. *Diamond* position of a target, *plus sign* position of a distractor, *crossed circle* next eye displacement. **a** is the probability distribution of constant model. **b** shows the probability distribution for the target model that shows a preference for the targets. **c** shows the probability distribution for the uncertainty model that highlights some of the targets. **d** shows the position of the targets and distractors in the visual field. Note that the probability distributions for model **c** favors the next eye movement

However, except in very special cases, the likelihood of a model would decrease exponentially toward zero with the increase of the number of trials, while the likelihood ratio between two models will diverge or converge exponentially toward zero. Therefore, we compare our decision models using the geometric mean of the likelihood of the observed eye movements over each trial. The geometric mean allows to be a substitute for the complete likelihood, as it is its Nth root where N is the total number of trials, while providing a measure converging to a non-zero value as the number of trials grows.

More precisely, let c_n^t be the tth eye movement recorded during trial n. The likelihood of a model π for trial n is:

$$\prod_{t=1}^{t_{\max}} P([C^t = c_n^{t+1}] \mid v_n^{1\to t}\, c_n^{1\to t}\, \pi)$$

The global likelihood of model π is:

$$\prod_{n=1}^{N} \prod_{t=1}^{t_{\max}} P([C^t = c_n^{t+1}] \mid v_n^{1\to t}\, c_n^{1\to t}\, \pi)$$

Finally we define our measure μ to be the geometric mean of the likelihood over all the trials:

$$\mu(\pi) = \sqrt[N]{\prod_{n=1}^{N} \prod_{t=1}^{t_{\max}} P([C^t = c_n^{t+1}] \mid v_n^{1\to t}\, c_n^{1\to t}\, \pi)} \tag{3}$$

3.2 Results and analysis

The data set is gathered from 11 subjects with 110 trials each for a total of 1,210 trials (Tanner et al. 2007). Each trial was regularly discretized in time in $t_{\max} = 24$ observations (with a timestep of 200 ms) for a grand total of 29,040 data points. The eye movement variable M^t is build from the difference in gaze position between two successive timesteps. Part of the data set (124 random trials) was used to determine the parameters of the various models and the results are computed on the remaining $N = 1{,}089$ trials.

Table 1 presents the ratio of the measure for each pair of our three decision models computed for this data set. It shows that the model which 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 Ratio of the measures for each pair of models

Model	Model		
	Constant (π_A)	Target (π_B)	Uncertainty (π_C)
Constant (π_A)	1	280	320
Target (π_B)	3.5×10^{-3}	1	1.14
Uncertainty (π_C)	3.1×10^{-3}	0.87	1

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

As explained above, the choice of the geometric mean prevents the measure to converge toward zero and prevents their ratios 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}. This shows that the likelihood ratio is indeed not a stable measure with respect to the number of trials. We preferred a stable measure in order to have a more meaningful value.

3.3 Typical situations

These results show a global agreement of the model with the actual eye movements of the human participants. However, there are some configurations where the models can have different relative performances. The analysis of such examples can shed some light on the behavior of the various decision models we proposed.

3.3.1 Examples where π_C is more likely than π_B

The global result shows that it is better to take into account uncertainty explicitly for the choice of the eye movement. We can further investigate by looking at the frames where the difference in the likelihood is greatest.

We isolated two different categories of configurations where model π_C was especially better than model π_B, exemplified in Fig. 5. The first category consists in scenes where a target and a distractor are in a close vicinity and the eye movement of the participant is around those objects (Fig. 5a). In these case, the target model is simply attracted by the target whereas the uncertainty model is additionally attracted by both objects due to their uncertainty.

The second category consists in occurrences of an eye movement towards a distractor (see Fig. 5b). In this case, the target model has no incentive for looking at this location whereas there is always some uncertainty to investigate for model π_C.

3.3.2 Examples where π_B is more likely than π_C

Even if the global results are in favor of the model with explicit uncertainty, there are cases where the target model better predicts the eye movements. This happen mainly when the eye movements occur in the middle of several targets but not on a particular one (Fig. 6a). In this case, the fusion on the targets employed by model π_B can present a maximum

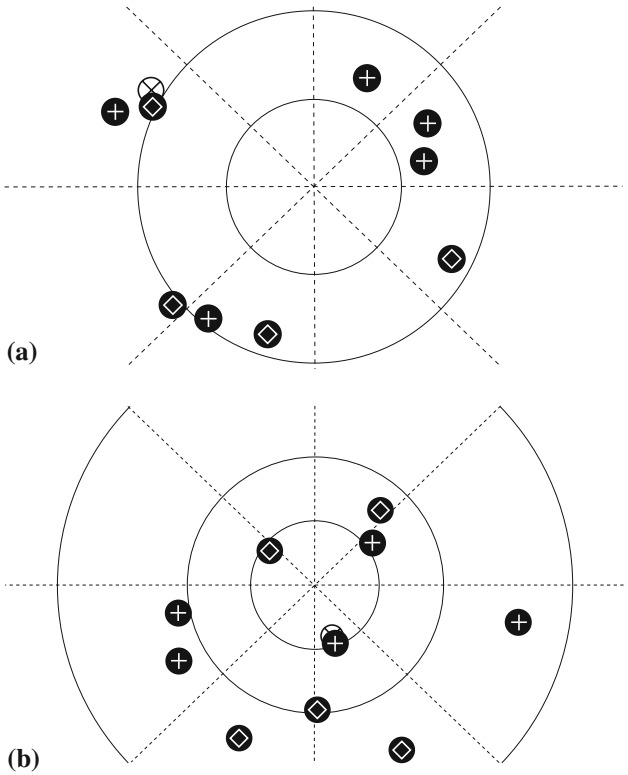

Fig. 5 Examples of eye movements better predicted by model π_C than model π_B. The scene is presented in an eye centered reference frame. *Diamond* position of a target, *plus sign* position of a distractor, *crossed circle* next eye displacement. **a** The actual eye movement occurs towards both a target and a distractor. **b** The actual eye movement occurs towards an isolated distractor

in a center of mass of the targets, whereas the absence of objects—and therefore the low uncertainty—will lower the probability of this particular eye movement by model π_C.

Figure 6b illustrates a second interesting case. The eye movement occurs in between a target and a distractor. However, the occupancy grid at that time (Fig. 6c) shows that the target is moving and the eye movement is near the previous position of the target shown by a peak of occupancy in the corresponding cell. Therefore the eye movement is near the representation of the target. On the other hand, there is also a great patch near the center of the visual field with a moderate level of uncertainty where, consequently, model π_C predicts a high probability of eye movement.

3.3.3 Examples where π_A is more likely than π_B or π_C

Finally, the constant model can also be the most likely one for some particular configurations and movements. This occurs mostly for fixations that are not directed to objects (for example Fig. 7a). Indeed model π_A is simply the global distribution of eye movements that are mostly of low amplitude (see Fig. 4a) and the other models are mostly attracted to targets or the uncertainty attached to objects.

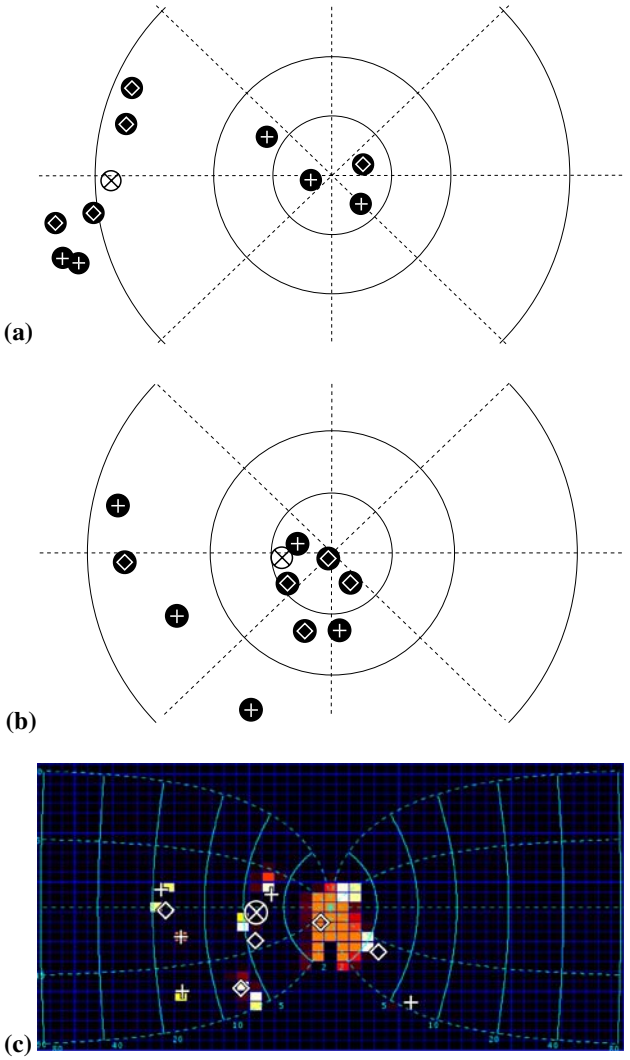

Fig. 6 Examples of eye movements better predicted by model π_B than model π_C. The scene is presented in an eye centered reference frame. *Diamond* position of a target, *plus sign* position of a distractor, *crossed circle* next eye displacement. **a** The actual eye movement occurs in between several targets. **b** The actual eye movement occurs towards an isolated distractor. **c** Occupancy grid for the same configuration depicted in **b** showing the eye movement is near the past location of the target

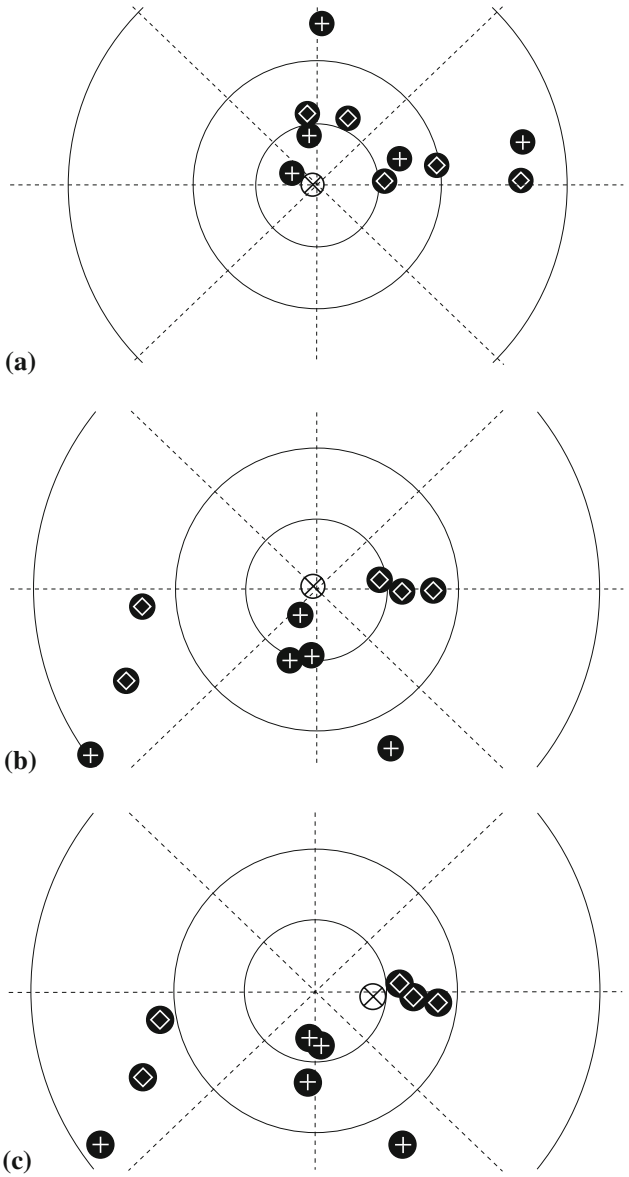

Fig. 7 Examples of eye movements better predicted by model π_A than models π_B or π_C. The scene is presented in an eye centered reference frame. *Diamond* position of a target, *plus sign* position of a distractor, *crossed circle* next eye displacement. **a** The actual eye movement is a fixation without object. **b** The actual eye movement is also a fixation although there is a group of targets on the right. **c** Situation following **b** where the eye movement is towards the group of targets

Figure 7b shows another occurrence of this situation with a group of target on the right towards which the other models predict a high probability of movement. It happens that, on the next frame, shown Fig. 7c, for which the situation is similar, the participant looked towards this group of targets, as predicted by both models π_B and π_C.

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.

In addition, 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 higher likelihood than the constant one which assigns a lower probability as the eccentricity grows. On the other hand, the difference between the target model and the uncertainty model is due to the filtering of the eye movements distribution from the target model by the uncertainty. The difference is less substantial 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 stimuli in order to manipulate uncertainty more precisely.

The stimulus is adapted from the classical MOT task used primarily to study attention. Our model uses a set of variables to track the position of the targets. This set of variable is fixed and finite (five in our model), which means our model can only track as much targets as its number of target position variables. The human subjects, however, are also informed about the number of targets in the instructions. Experimental evidence suggests that human performance drops if the number of target gets too high. For the particular experimental design we used, the maximum number of targets consistently tracked was 5, which justifies our choice of the number of target variables. Other experimental studies suggest that this maximum number of target is not fixed and seems to depend on factors such as speed and spacing of the objects (Alvarez and Franconeri 2007). In addition, each of our target variables cover the whole visual field (encoded in the logcomplex mapping) although there are works indicating that some representation capacities are separated across the hemifields (Alvarez and Cavanagh 2005). It could be interesting to test this in our model with a set of target variables for the left part and another for the right part. However, due both to eye movements and targets movements, the targets sometimes change side, implying some additional mechanism of communication between these variables.

Finally, one of the main features of our model is to place all computations and representation in the logcomplex mapping found in the neurophysiology of some retinotopic maps. To our surprise, we found in the psychophysical data that the distribution of the objects positions is quite uniform in the logcomplex mapping. This suggests a particular strategy for the eye movements. One interpretation could be that the eye movements are chosen in order to maximize the use of the representation: that is, so that the objects are uniformly distributed in this representation. This seems to be an indirect confirmation that eye movements are governed by structures using this particular mapping.

Acknowledgments The authors acknowledge the support of the European Project BACS (Bayesian Approach to Cognitive Systems), FP6-IST-027140. The authors thank Luiz Canto-Pereira, Heinrich Bülthoff, and Cristóbal Curio for their involvement in the experimental aspects of this work. The authors also thank warmly Julien Diard for the insightful discussions about the preliminary model design.

Appendix A: Experimental protocol

This experiment is an adaptation of the classical MOT paradigm from Pylyshyn and Storm (1988) (see Fig. 3) but with eye movements. In the original task, participants were asked to keep track of a given number of targets among identical distractors as they all move independently on the screen. Participants had to keep their gaze at a fixating point located on the center of the screen. Therefore the targets will occasionally be located in the periphery of the visual field, in the low resolution areas of the visual field. Therefore we expect eye movements to occur in order to keep track of targets.

A.1 Materials and methods

A.1.1 Participants

Eleven subjects participated in the experiment with normal or corrected vision. Each session consists of 110 trials.

A.1.2 Apparatus

The stimulus is presented on a calibrated 21″ Sony CPD-500 CRT monitor with a refresh rate of 100 Hz and a resolution of $1{,}024 \times 768$. Participants are positioned in front of the monitor at a distance of 65 cm; at this distance the display subtended a visual angle of 33° by 25°. A chin rest ensures that no head movement occurs during the experimental session. All experimental sessions are performed in a sound attenuated room with controlled artificial lighting. Eye movements are recorded by an eye tracker system (EyeLink II, SR Research Ltd.) with a sampling rate of 250 Hz and an accuracy of ca. 0.3°. The model was simulated offline with a timestep of 200 ms using the difference in eye position between two timesteps. No analysis of saccades, micro-saccades, pursuit or fixation was needed in this respect.

A.2. Procedure

The display consists of ten identical objects, each one a white circle subtending 1° of visual angle, with a luminance of 90 cd/m^2 against a black background, in a room illuminated with diffuse D65 light (70 cd/m^2).

Targets and distractors are identical with the exception of the initial phase in the beginning of each trial. In this phase, five targets are cued by a series of three flashes, with a total duration of 1,080 ms. After this initial phase, all objects begin

to move in different directions, chosen from among 8 directions of the compass with a mean velocity of 5.1° per second.

The objects have random initial locations, directions and speeds during trials but are constrained to keep a minimum distance of 1.5° (Pylyshyn and Storm 1988).

Trials last 5 s and on the end of each trial participants are asked to select targets with a mouse.

More details can be found in the description of experiment B in (Tanner et al. 2007, paper in preparation).

Appendix B: Dynamic models

B.1 Dynamic object model

This dynamic model provides the transition probability distribution $P(O^t_{(x,y)} \mid M^t\ O^{t-1})$ that governs the evolution of the grid with the remapping capability. In order to stress the issue of the logcomplex mapping, we explicitly refer to the visual coordinates (ρ, θ) as well as the logcomplex coordinates (x, y). We also consider coordinates $(\rho, \theta)_{\text{ant}}$ and $(x, y)_{\text{ant}}$ to denote coordinates at the previous time step. In the end, the decomposition is as follows:

$$\begin{aligned} &P((x,y)\ (x,y)_{\text{ant}}\ (\rho,\theta)\ (\rho,\theta)_{\text{ant}}\ O^t_{(x,y)}\ O^{t-1}\ M^t) \\ &\quad = P((x,y))P(M^t)P(O^t_{(x,y)})P((\rho,\theta) \mid (x,y)) \\ &\qquad \times P((\rho,\theta)_{\text{ant}} \mid (\rho,\theta)\ M^t)P((x,y)_{\text{ant}} \mid (\rho,\theta)_{\text{ant}}) \\ &\qquad \times \prod_{(x',y')} P(O^{t-1}_{(x',y')} \mid O^t_{(x,y)}\ (x,y)_{\text{ant}}) \end{aligned}$$

where:

- $P((x, y))$ is an arbitrary unused distribution;
- $P(M^t)$ is an arbitrary unused distribution;
- $P(O^t_{(x,y)})$ is a uniform distribution;
- $P((\rho, \theta) \mid (x, y))$ is a uniform distribution on the inverse image of the position (x, y) by the logcomplex mapping;
- $P((\rho, \theta)_{\text{ant}} \mid (\rho, \theta)\ M^t)$ is a Dirac distribution on the image of (ρ, θ) by eye movement M^t;
- $P((x, y)_{\text{ant}} \mid (\rho, \theta)_{\text{ant}})$ is a Dirac distribution on the cell corresponding to position $(\rho, \theta)_{\text{ant}}$;
- $P(O^{t-1}_{(x',y')} \mid O^t_{(x,y)}\ (x^{-1}, y^{-1}))$ is a transition matrix that states there is a great probability to keep the same occupancy if $(x', y') = (x, y)_{\text{ant}}$, and is a uniform distribution otherwise.

This model is used to compute the question $P(O^t_{(x,y)} \mid M^t\ O^{t-1})$ using the following expression:

$$\begin{aligned} &P(O^t_{(x,y)} \mid O^{t-1}\ M^t) \\ &\quad \propto \sum_{(\rho,\theta)} P((\rho,\theta) \mid (x,y))P(O^{t-1}_{(\hat{x},\hat{y})} \mid O^t_{(x,y)}\ (\hat{x},\hat{y})) \end{aligned}$$

where $(\hat{x}, \hat{y})$ are the coordinates of the cell corresponding to the image of (ρ, θ) by eye motion M^t.

This summation can be implemented by sampling the distribution $P((\rho, \theta) \mid (x, y))$.

B.2 Dynamic target model

This dynamic target model is common to every target and combines both the prediction of the position of the target based only on eye movement (remapping) and the update of this position according to the occupancy grid. It provides the distribution $P(T^t_i \mid T^{t-1}_i\ O^t\ M^t)$ used in the representation model.

The decomposition is as follows:

$$\begin{aligned} &P(T^t_i\ T^{t-1}_i\ (\rho,\theta)\ (\rho,\theta)_{\text{ant}}\ M^t\ O^t) \\ &\quad = P(T^t_i)P(M^t)P((\rho,\theta) \mid T^t_i) \\ &\qquad \times P((\rho,\theta)_{\text{ant}} \mid (\rho,\theta)\ M^t)P(T^{t-1}_i \mid (\rho,\theta)_{\text{ant}}) \\ &\qquad \times \prod_{(x,y)} P(O^t_{(x,y)} \mid T^t_i) \end{aligned}$$

where:

- $P(T^t_i)$ is a uniform distribution;
- $P(M^t)$: is an arbitrary unused distribution;
- $P((\rho, \theta) \mid T^t_i)$ is a uniform distribution on the inverse image of the position T^t_i by the logcomplex mapping;
- $P((\rho^{-1}, \theta^{-1}) \mid (\rho, \theta)\ M^t)$ is Dirac distribution on the image of (ρ, θ) by eye movement M^t;
- $P(T^{t-1}_i \mid (\rho, \theta)_{\text{ant}})$ is a Dirac on the cell corresponding to position $(\rho, \theta)_{\text{ant}}$;
- $P(O^t_{(x,y)} \mid T^t_i)$ states that it is more probable to have an occupied cell in a neighborhood of T^t_i, and that it is uniform elsewhere.

This model is used to compute the question $P(T^t_i \mid T^{t-1}_i\ O^t\ M^t)$ with the following expression:

$$\begin{aligned} &P(T^t_i \mid T^{t-1}_i\ M^t\ O^t) \\ &\quad \propto \left|\mathcal{E}(T^{t-1}_i, M^t)\right| \prod_{(x,y)} P(O^t_{(x,y)} \mid T^t_i) \end{aligned}$$

where $\left|\mathcal{E}(T^{t-1}_i, M^t)\right|$ is the size of the set of the polar positions (ρ, θ) that are in relation with T^{t-1}_i by the eye movement M^t. This set can be obtained by sampling like in the dynamic model.

Appendix C: Implementation details

The models presented are implemented in the Java language. In all the examples, the grid $\mathcal{G}$ is composed of 24×29 cells

for each hemifield and we used a timestep of 200 ms for the representation and decision models.

Additionally, some of the probability distributions described as factors in the decompositions are parametric forms that need precise values to be involved in actual computations. We explored the parametrical space and evaluated each parameter set with our measure computed on a subset of the experimental data.

Finally, in the representation model, the observation model $P(V^t_{(x,y)} \mid O^t_{(x,y)})$ is a 2×2 matrix with value 0.9 on the diagonal and 0.1 elsewhere

$$\begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}.$$

The transition matrix of the dynamic model is

$$\begin{pmatrix} 0.95 & 0.1 \\ 0.05 & 0.9 \end{pmatrix}.$$

The target observation model $P(O^t_{(x,y)} \mid T^t_i)$ is of the form $0.5 + \frac{0.25}{1+\left(\frac{d((x,y),T^t_i)}{0.02}\right)^2}$ for an occupied cell and $0.5 - \frac{0.25}{1+\left(\frac{d((x,y),T^t_i)}{0.02}\right)^2}$ otherwise with $d((x,y),T^t_i)$ the distance between cell (x,y) and position T^t_i in mm. The target fusion model $P(T^t_i \mid C^t)$ is a mixture between a Gaussian and a uniform distribution: $\propto 0.25 + \exp -\frac{d(T^t_i,C^t)^2}{0.25}$. In the uncertainty decision model, the uncertainty fusion distribution $P(I^t_{(x,y)} \mid C^t\, \pi_C)$ is a symmetrical beta distribution with parameter 0.075.

References

Alvarez GA, Cavanagh P (2005) Independent resources for attentional tracking in the left and right visual hemifields. Psychol Sci 16(8):637–643

Alvarez GA, Franconeri SL (2007) How many objects can you attentively track? Evidence for a resource-limited tracking mechanism. J Vis 7(13):1–10. http://journalofvision.org/7/13/14/

Barash S, Bracewell R, Fogassi L, Gnadt J, Andersen R (1991a) Saccade-related activity in the lateral intraparietal area. I. Temporal properties; comparison with area 7a. J Neurophysiol 66(3): 1095–1108

Barash S, Bracewell R, Fogassi L, Gnadt J, Andersen R (1991b) Saccade-related activity in the lateral intraparietal area. II. Spatial properties. J Neurophysiol 66(3):1109–1124

Ben Hamed S, Duhamel JR, Bremmer F, Graf W (2001) Representation of the visual field in the lateral intraparietal area of macaque monkeys: a quantitative receptive field analysis. Exp Brain Res 140:127–144

Bessière P, Laugier C, Siegwart R (2008) Probabilistic reasoning and decision making in sensory-motor systems. Springer, Berlin

Bozis A, Moschovakis A (1998) Neural network simulations of the primate oculomotor system III. An one-dimensional, one-directional model of the superior colliculus. Biol Cybern 79:215–230

Cavanagh P, Alvarez GA (2005) Tracking multiple targets with multifocal attention. Trends Cogn Sci 9(7): 349–354

Droulez J, Berthoz A (1991) A neural network model of sensoritopic maps with predictive short-term memory properties. Proc Natl Acad Sci 88:9653–9657

Elfes A (1989) Occupancy grids: a probabilistic framework for robot perception and navigation. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, USA

Fehd HM, Seiffert AE (2008) Eye movements during multiple object tracking: where do participants look. Cognition 108(1):201–209

Gnadt J, Andersen R (1988) Memory related motor planning activity in the posterior arietal cortex of the macaque. Exp Brain Res 70(1):216–220

Goldberg M, Bruce C (1990) Primate frontal eye fields. III. Maintenance of a spatially accurate saccade signal. J Neurophysiol 64(2):489–508

Herrero L, Rodríguez F, Salas C, Torres B (1998) Tail and eye movememnts evoked by electrical microstimulation of the optic tectum in goldfish. Exp Brain Res 120:291–305

Krauzlis R (2004) Recasting the smooth pursuit eye movement system. J Neurophysiol 91(2):591–603

Lebeltel O, Bessière P, Diard J, Mazer E (2004) Bayesian robots programming. Auton Robots 16(1):49–79

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

McIlwain J (1976) Large receptive fields and spatial transformations in the visual system. In: Porter R (ed) Neurophysiology II, Int Rev Physiol, vol 10. University Park Press, Baltimore, pp 223–248

McIlwain J (1983) Representation of the visual streak in visuotopic maps of the cat's superior colliculus: influence of the mapping variable. Vis Res 23(5):507–516

Mitchell J, Zipser D (2003) Sequential memory-guided saccades and target selection: a neural model of the frontal eye fields. Vis Res 43:2669–2695

Moschovakis A, Scudder C, Highstein S (1996) The microscopic anatomy and physiology of the mammalian saccadic system. Prog Neurobiol 50:133–254

Ottes F, van Gisbergen JA, Eggermont J (1986) Visuomotor fields of the superior colliculus: a quantitative model. Vis Res 26(6): 857–873

Pylyshyn Z, Storm R (1988) Tracking multiple independent targets: evidence for a parallel tracking mechanism. Spatial Vis 3(3):1–19

Robinson D (1972) Eye movements evoked by collicular stimulation in the alert monkey. Vis Res 12:1795–1808

Schwarz E (1980) Computational anatomy and functional architecture of striate cortex: A spatial mapping approach to perceptual coding. Vis Res 20:645–669

Scudder C, Kaneko C, Fuchs A (2002) The brainstem burst generator for saccadic eye movements. A modern synthesis. Exp Brain Res 142:439–462

Siminoff R, Schwassmann H, Kruger L (1966) An electrophysiological study of the visual projection to the superior colliculus of the rat. J Comp Neurol 127:435–444

Sommer M, Wurtz R (2000) Composition and topographic organization of signals sent from the frontal eye fields to the superior colliculus. J Neurophysiol 83:1979–2001

Tanner T, Canto-Pereira L, Bülthoff H (2007) Free vs. constrained gaze in a multiple-object-tracking-paradigm. In: 30th European Conference on Visual Perception, Arezzo, Italy

Wurtz R, Sommer M, Paré M, Ferraina S (2001) Signal transformation from cerebral cortex to superior colliculus for the generation of saccades. Vis Res 41:3399–3412

Zelinsky GJ, Neider MB (2008) An eye movement analysis of multiple object tracking in a realistic environment. Vis Cogn 16(5):553–566