D2.4: Submitted papers on the Bayesian eye movement selection model

Due date of deliverable: June 30th, 2008
Actual submission date: July 11th, 2008

Start date of project: 01.01.2006
Duration: 50 month

Organisation name of lead contractor for this deliverable: CNRS - LPPA
Revision 1

---

<table>
<thead>
<tr>
<th>Dissemination Level</th>
<th>PU</th>
<th>PP</th>
<th>RE</th>
<th>CO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restricted to other programme participants (including the Commission Services)</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Restricted to a group specified by the consortium (including the Commission Services)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidential, only for members of the consortium (including the Commission Services)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Project co-funded by the European Commission within the Sixth Framework Programme (2002–2006)
Abstract

This deliverable was initially planned for December 2007 and was rescheduled for June 2008. It is composed of two papers, the first one, submitted to the neurocomp08 conference, and the second about to be submitted to a journal. Both detail the Bayesian eye movement selection model, which derives from the previous theoretical work presented in D2.2, and its results confronted with the experimental data presented in D2.5.

Contributors:

- Francis Colas, Benoît Girard, Fabien Flacher (CNRS–LPPA),
- Pierre Bessière (CNRS–Grenoble),
- Thomas Tanner (MPI).
EXPLICIT UNCERTAINTY FOR EYE MOVEMENT SELECTION

F. Colas,1∗ F. Flacher,1 P. Bessière,2 B. Girard1
1 – Laboratoire de Physiologie de la Perception et de l’Action
CNRS - Collège de France
11 place Marcelin Berthelot
75231 Paris Cedex 05
2 – Laboratoire d’Informatique de Grenoble
CNRS - Grenoble Universités
655 avenue de l’Europe
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) [3, 4, 5].

∗Corresponding author, email: francis.colas@college-de-france.fr.
The authors acknowledge the support of the European Project BACS
(Bayesian Approach to Cognitive Systems), FP6-IST-027140.

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 [6, 7], 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 [8], 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 [9]. 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 [10], FEF [11] and LIP [12] 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 [13]. The environment is discretized into a regular grid $G$ (with the logcomplex mapping) and we define a binary variable $Occ_i$ in each cell $i$ 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 $Obs_i$. The observation and occupancy of each cell are linked by a probabilistic relation $P(Obs_i | Occ_i)$ that states it is likely to observe the assumed occupancy of the cell.

The remapping capability of this model relies on the current displacement $Mvt^t$ and the distribution $P(Occ_{i | Obs}^t | Mvt^{t-1})$ 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 $A(c)$ 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:

$$P(Occ_{i}^t | Obs_{i}^{t-1}, Mvt^{t-1}) \propto P(Obs_{i}^t | Occ_{i}^t) \sum_{Occ_{A(c)}^{t-1}} P(Occ_{c}^{t-1} | Mvt^{t-1}) P(Occ_{c'}^{t-1} | Obs_{c'}^{t-1}, Mvt^{t-1})$$

(1)

**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 $P(Tgt_i^t | Tgt_i^{t-1}, Occ^t, Mvt^t)$ 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:

$$P(Tgt_i^t | Obs_{i}^{t-1}, Mvt^{t-1}) \propto \sum_{Tgt_i^{t-1}} \left[ P(Tgt_i^{t-1} | Obs_{i}^{t-1}, Mvt^{t-1}) \times \sum_{Occ_{i}} P(Occ_{i}^t | Obs_{i}^{t-1}, Mvt^{t-1}) \times P(Tgt_i^t | Tgt_i^{t-1}, Occ^t, Mvt^t) \right]$$

(2)

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 $P(Tgt_i^t | Mot^t)$ 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:

$$P(Mot^t | Obs_{i}^{t-1}, Mvt^{t-1}) \propto P(Mot) \prod_{i=1}^{N} \left[ \sum_{Tgt_i^t} P(Tgt_i^t | Obs_{i}^{t-1}, Mvt^{t-1}) P(Tgt_i^t | Mot^t) \right]$$

**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 $I_t^c$ 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:

$$P(Mot^\tau | Obs^{1:T} Mot^{1:T}) \propto P(Mot^\tau | Obs^{1:T} Mot^{1:T}) P(I^\tau_{Mot} | Mot^\tau)$$

with $I^\tau_{Mot} = P(Occ^\tau_c | Obs^{1:T} Mot^{1:T})$ (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: $P(Model | Data) \propto P(Model) \times P(Data | Model)$.

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 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:

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

for the target model, and

$$\sqrt[\sum T_{\tau=1}^N]{\prod_{t=1}^T P([Mot = mot_t^{\tau+1} | obs^{\tau,t} mot_t^{\tau+1}])}$$

for the uncertainty model, where $mot_t^\tau$ is the actual eye movement recorded in trial $\tau$ 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 [14] 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 parameters of the model and the results are computed on the remaining 1089 trials.

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 $4.9 \times 10^{63}$. With half the trials, this likelihood ratio is the square root, that is only $7.0 \times 10^{31}$. We preferred presenting the results with a measure independent of the number of trials.

<table>
<thead>
<tr>
<th>Ratio</th>
<th>Constant</th>
<th>Target</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1</td>
<td>280</td>
<td>320</td>
</tr>
<tr>
<td>Target</td>
<td>$3.5 \times 10^{-3}$</td>
<td>1</td>
<td>1.14</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>$3.1 \times 10^{-3}$</td>
<td>0.87</td>
<td>1</td>
</tr>
</tbody>
</table>

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

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

The authors thank Thomas Tanner, Luiz Canto-Pereira, and Heinrich Bülthoff for the experimental results.

References

Bayesian models of eye movement selection with retinotopic maps

Francis Colas¹  Fabien Flacher¹  Thomas Tanner³  Pierre Bessière²
Benoit Girard¹

1) Laboratoire de Physiologie de la Perception et de l’Action
CNRS / Collège de France
11 pl. Marcelin Berthelot
75231 Paris Cedex 05
France
2) Laboratoire d’Informatique de Grenoble
CNRS / Grenoble Universités
655 av. de l’Europe
38334 Montbonnot
France
3) Cognitive and Computational Psychophysics Dept.
MPI for Biological Cybernetics
Spemannstr. 44
72076 Tuebingen
Germany

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 do so, we confront the movement predictions of computational models with human results from a psychophysical task. This task is an freely moving eye version of the Multiple Object Tracking task with the eye movements possibly compensating for lower peripheral resolution. We design several Bayesian models of increasing complexity, whose layered structures are inspired by the neurobiology of the brain areas implied in eye movement selection. 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.

1 Introduction

We usually make several 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. However, as our peripheral definition is poor; we do not know precisely how much information we can gain, depending on what is there to see. The uncertainty is a common issue for both perception – because we can not be sure of what we perceive – and action – because we can not be sure of the consequences of our action. 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).

Bayesian methodology

In order to handle and reason about uncertainty, we use the Bayesian Programming framework (Lebeltel et al., 2004). This framework provides a systematic procedure to build and use a Bayesian model. A Bayesian model represents knowledge with uncertainty using probability distributions and reasons about it with the probabilistic rules. 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.

Eye movement circuitry

Even if we do not have the pretension to build a model of the neurophysiology of the eye movement selection related brain regions, we are 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 in 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.

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).

![Diagram](image)

Figure 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 spatial working memory-related neurons 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).

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 generation of motor commands.

**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, Tanner, in preparation) of the classical MOT task (Pylyshyn and Storm, 1988). In this experiment participants are presented with a set of targets among a number of distractors. All of these objects are indiscernible $1^\circ$ large discs and move in a 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 primitive features therefore allowing for a study of the eye movement selection that occurs in this context.

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).

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

Occupancy grid  Occupancy grids are a standard way to represent the state of an environment. They were originally introduced for obstacles representation 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 $\mathbf{Occ}$ of binary variables $Occ_t^{(x,y)}$, one for each timestep $t \in [0,T]$ and location $(x,y) \in \mathcal{G}$ where $\mathcal{G}$ is a regular grid in the retino-centered logcomplex reference frame.\footnote{We also assume that we have visual inputs in this same reference frame, represented by a collection $\mathbf{Obs}$ of binary variables $Obs_t^{(x,y)}$ for $t \in [1,T]$ indicating if an object (either target or distractor) is perceived in the corresponding cell. Finally, we include some past eye movement information $Mt^t$ in order to model the remapping 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:

$$P(\mathbf{Occ} \cdot \mathbf{Obs} \cdot Mt) = P(\mathbf{Occ}^0) \prod_{t=1}^{T} P(\mathbf{Occ}^t \cdot \mathbf{Obs}^t \cdot Mt^t \mid \mathbf{Occ}^{t-1})$$

$$= \prod_{(x,y) \in \mathcal{G}} P(\mathbf{Occ}^0_{(x,y)})$$

$$\times \prod_{t=1}^{T} P(Mt^t) \prod_{(x,y) \in \mathcal{G}} \left[ P(\mathbf{Occ}^t_{(x,y)} \mid Mt^t \cdot \mathbf{Occ}^{t-1}) \times P(\mathbf{Obs}^t_{(x,y)} \mid \mathbf{Occ}^t_{(x,y)}) \right]$$

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: $Occ = \bigwedge_{t=0}^{T} Occ^t = \bigwedge_{(x,y) \in \mathcal{G}} \mathbf{Occ}^t_{(x,y)}$.
In this expression, \( P(Occ^0_{(x,y)}) \) is an arbitrary prior on the occupancy of the visual scene, \( P(Mvt^t) \) is a distribution over the eye movement that can be chosen arbitrarily as the results of the inference do not depend on it, as far as it is non zero for the actual eye movements observed. The relation between the occupancy and the observation, \( P(Obs^t_{(x,y)} \mid Occ^t_{(x,y)}) \), 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 grid evolution with the remapping capability is specified by the transition model, \( P(Occ^t_{(x,y)} \mid Mvt^t Occ^{t-1}) \), 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 scene corresponds to the following question for each time \( t \):

\[
P(Occ^t \mid Obs^{1-t} Mvt^{1-t})
\]

This expression can be computed in an iterative manner using Bayesian inference:

\[
P(Occ^t \mid Obs^{1-t} Mvt^{1-t})
\approx \prod_{(x,y) \in G} P(Obs^t_{(x,y)} \mid Occ^t_{(x,y)}) \times \sum_{Occ^{t-1}} \left[ \prod_{(x,y) \in G} P(Occ^t_{(x,y)} \mid Mvt^t Occ^{t-1}) \right]
\]

However the summation over all possible grid states 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:

\[
P(Occ^t_{(x,y)} \mid Obs^{1-t} Mvt^{1-t})
\approx P(Obs^t_{(x,y)} \mid Occ^t_{(x,y)}) \times \sum_{Occ^{t-1}_{A(x,y)}} \left[ P(Occ^t_{(x,y)} \mid Mvt^t Occ^{t-1}_{A(x,y)}) \right]
\]

where \( A(x,y) \) is the subset of the grid that are the antecedent of the cell \((x,y)\) by the current eye movement \( Mvt^t \).

**Positions of the targets** The previous model allows for the description of the visual scene without differentiating between targets and distractors. In order to introduce this difference, we add a set of variables \( Tgt^t_i \) to represent the location of each target \( i \in [1, N] \) at each time \( t \in [0, T] \) in the logcomplex retino-centered reference frame.

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

These variables are included in the model with an additional factor \( P(Tgt^t_i \mid Tgt^{t-1}_i Occ^t Mvt^t) \) representing the dynamic model of targets:

\[
P(Occ \mid Obs \mid Mvt \mid Tgt)
= \prod_{(x,y) \in G} P(Occ^0_{(x,y)}) \prod_{i=1}^N P(Tgt^0_i)
\times \prod_{t=1}^T \left[ \prod_{(x,y) \in G} P(Occ^t_{(x,y)} \mid Mvt^t Occ^{t-1}) \right]
\times \prod_{t=1}^T \left[ \prod_{(x,y) \in G} P(Obs^t_{(x,y)} \mid Occ^t_{(x,y)}) \right]
\times \prod_{i=1}^N P(Tgt^t_i \mid Tgt^{t-1}_i Occ^t Mvt^t)
The additional factors $P(Tgt_i^0)$ 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, $P(Tgt_i^t \mid Tgt_i^{t-1} \text{Occ} \text{Mot}^t)$, 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 $t \in [1, T]$:

$$P(Tgt_i^t \mid \text{Obs}^{1-t} \text{Mot}^{1-t}) \tag{2}$$

Bayesian inference leads to the following expression for this question:

$$P(Tgt_i^t \mid \text{Obs}^{1-t} \text{Mot}^{1-t}) \propto \sum_{Tgt_i^{t-1}} \left[ P(Tgt_i^{t-1} \mid \text{Obs}^{1-t-1} \text{Mot}^{1-t-1}) \times P(\text{Occ}^t \mid \text{Obs}^{1-t} \text{Mot}^{1-t}) \times P(Tgt_i^t \mid Tgt_i^{t-1} \text{Occ}^t \text{Mot}^t) \right]$$

where $P(Tgt_i^{t-1} \mid \text{Obs}^{1-t-1} \text{Mot}^{1-t-1})$ is the result of the same inference at the preceding timestep, $P(\text{Occ}^t \mid \text{Obs}^{1-t} \text{Mot}^{1-t})$ 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, we 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 one 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 one 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 ($\pi_A$), one that uses the representation model without explicitly taking into account uncertainty ($\pi_B$), and finally one that uses the representation model and explicitly takes into account uncertainty ($\pi_C$). Each model $\pi_k$ will infer a probability distribution on the next eye movement represented by a new variable $\text{Mot}^t \in G$ at each time $t \in [1, T]$: $P(\text{Mot}^t \mid \text{Obs}^{1-t} \text{Mot}^{1-t} \pi_k)$.

**Constant model** This constant model is a baseline for the other models. We need 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:

$$\forall t \in [1, T], P(\text{Mot}^t \mid \text{Obs}^{1-t} \text{Mot}^{1-t} \pi_A) = P(\text{Mot}^t \mid \pi_A) = P(\text{Mot}^1 \mid \pi_A)$$

In these conditions, it can be shown that the best distribution $P(\text{Mot}^1 \mid \pi_A)$, according to the measure defined section 3.1, assigns the probability of each individual discretized motion to be equal to its frequency in the experimental data. Therefore, we learned this distribution from our experimental data, using only a randomly selected subset in order not to overfit our models.
The second model we propose determines its eye movements based on the knowledge from the representation layer. More precisely, it tends to look at locations where targets are near, 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:

\[
P(Mot\; Obs\; Mvt\; Tgt \mid \pi_B) = \prod_{t=1}^{T} \left[ \prod_{i=1}^{N} P(Obs^t\; Mvt^t \mid \pi_B) \cdot \prod_{i=1}^{N} P(Tgt^t_i \mid Obs^{1-t}\; Mvt^{1-t} \; \pi_B) \cdot P(Mot^t \mid Tgt^t \; \pi_B) \right]
\]

where:

- \(P(Obs^t\; Mvt^t \mid \pi_B)\) is an arbitrary prior that is not used in the inference,
- \(P(Tgt^t_i \mid Obs^{1-t}\; Mvt^{1-t} \; \pi_B)\) is the result of inference 2,
- \(P(Mot^t \mid Tgt^t \; \pi_B)\) is the result of the inference in a fusion submodel over the targets that yields:

\[
P(Mot^t \mid Tgt^t \; \pi_B) \propto P(Mot^t \mid \pi_A) \prod_{i=1}^{N} P(Tgt^t_i \mid Mot^t)
\]

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

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

\[
P(Mot^t \mid Obs^{1-t}\; Mvt^{1-t} \; \pi_B) \propto P(Mot^t \mid \pi_A) \prod_{i=1}^{N} \sum_{Tgt^t_i} P(Tgt^t_i \mid Obs^{1-t}\; Mvt^{1-t} \; \pi_B) P(Tgt^t_i \mid Mot^t)
\]

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.

**Uncertainty model** The behaviour 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 \((x,y) \in G\) at time \(t \in [1,T]\). Any index can fit as long as we can correlate the value of this uncertainty index with the actual uncertainty.

To simplify, as we represent occupancy as binary variables, we choose our uncertainty indices to be equal to this probability of occupancy. The relation between this uncertainty index (probability distribution) and uncertainty is such as a probability
near $\frac{1}{2}$ represents a high uncertainty whereas a probability near 0 or 1 represent a low uncertainty. There are other choices for the space of these variables, such as entropy, but we keep the probability distribution to simplify our computations.

As above, this model is structured around a prior probability of motion which is filtered by these uncertainty variables such as to enhance probability of eye movement towards uncertain regions. The prior probability is the result of the preceding model $\pi_B$.

The decomposition of this model is as follows:

$$P(Mot \mid Obs \cdot Mvt \cdot I \mid \pi_C)$$

$$= \prod_{t=1}^{T} \left[ \frac{P(Obs^t \cdot Mvt^t \mid \pi_C)}{P(Mot^t \mid Obs^{1-t} \cdot Mvt^{1-t} \cdot \pi_B)} \prod_{(x,y) \in G} P(I^t_{(x,y)} \mid Mot^t \cdot \pi_C) \right]$$

Where:

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

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

$$P(Mot^t \mid Obs^{1-t} \cdot Mvt^{1-t} \cdot I^{1-t} \cdot \pi_C)$$

$$\propto P(Mot^t \mid Obs^{1-t} \cdot Mvt^{1-t} \cdot \pi_B)P(I^t_{Mot^t} \mid Mot^t \cdot \pi_C)$$

where $\forall (x,y), t \in G \times [1,T], I^t_{(x,y)} = P(Occ^t_{(x,y)} \mid Obs^{1-t} \cdot Mvt^{1-t})$ as computed by 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

In order to compare these models, we first introduce a comparison method to assess the relative prediction power of the models based on experimental data. We present their results and comment them with respect to the specific behaviour 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 (figure 3). 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 in themselves are not really significant as, when the possibilities are numerous, they tend to be very small. However, their comparison across models

---

\[3\] This is a matter of presentation of the model. The complete expression of $\pi_C$ can be written without reference to model $\pi_B$ but the addition of uncertainty would be less clear.
Figure 3: Example of probability distributions computed by each model in the same configuration. The two halves of the representations are drawn side by side. The plain cyan lines are the iso-eccentricities and the cyan dotted lines are the iso-directions. The color 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. Panel (a) is the probability 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. Panel (d) shows the position of the targets (dark red) and distractors (light green) in the visual field.

(which share the same number of possibilities) indicates which model is a better predictor of the recorded eye movements. This process is known as Maximum Likelihood method.

However, except in very special cases, the likelihood of a model would decrease exponentially toward zero, and the likelihood ratio between two models will diverge or converge exponentially toward zero with the number of trials. 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 $N$th root where $N$ is the total number of trial, while providing a measure converging to a non-zero value as the number of trial grows.

More precisely, let $mot_n^t$ be the $t$th eye movement recorded during trial $n$. The likelihood of a model $\pi$ for trial $n$ is:

$$\prod_{t=1}^{T} P([Mot^t = mot_{n}^{t+1}] \mid obs_{n}^{1-t} mot_{n}^{1-t} \pi)$$

The global likelihood of model $\pi$ is:

$$\prod_{n=1}^{N} \prod_{t=1}^{T} P([Mot^t = mot_{n}^{t+1}] \mid obs_{n}^{1-t} mot_{n}^{1-t} \pi)$$

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

$$\mu(\pi) = \sqrt[\prod_{n=1}^{N} \prod_{t=1}^{T} P([Mot^t = mot_{n}^{t+1}] \mid obs_{n}^{1-t} mot_{n}^{1-t} \pi)} (3)$$
3.2 Results and analysis

The data set is gathered from 11 subjects with 110 trials each for a total of 1210 trials (see Tanner et al., 2007, Tanner, in preparation, for details). Each trial was regularly discretized in time in $T = 24$ observations for a grand total of 29040 data points. 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 $N = 1089$ trials.

<table>
<thead>
<tr>
<th>Constant</th>
<th>Target</th>
<th>Uncertainty</th>
<th>Model</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1$</td>
<td>$3.5 \times 10^{-3}$</td>
<td>$3.1 \times 10^{-3}$</td>
<td>$280$</td>
<td>$1$</td>
</tr>
<tr>
<td>$1.14$</td>
<td>$1.40$</td>
<td>$0.87$</td>
<td>Target</td>
<td>Uncertainty</td>
</tr>
</tbody>
</table>

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

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 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.

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 $4.9 \times 10^{63}$. With half the trials, this likelihood ratio is the square root, that is only $7.0 \times 10^{31}$. 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 a light on the behaviour of the various decision models we proposed.

Examples where $\pi_C$ is better than $\pi_B$. 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 $\pi_C$ was especially better than model $\pi_B$, exemplified in figure 4. 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. 4a). In these case, the target model is simply attracted by the target whereas the uncertainty model is, in addition, attracted by both objects due to their uncertainty.

The second category consists in occurrences of an eye movement towards a distractor (Fig. 4b). In this case, the target model has no incentive for looking at this location whereas there is always some uncertainty to investigate for model $\pi_C$. 

10
Examples where $\pi_B$ is better than $\pi_C$. 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 happens mainly when the eye movement happens in the middle of several targets but not on a particular one (example Fig. 5a). In this case, the fusion on the targets operated by model $\pi_B$ can present a maximum 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 $\pi_C$.

A second interesting case is depicted by Fig. 5b. The eye movement occurs in between a target and a distractor. However, the occupancy grid at that time (Fig. 5c) 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 $\pi_C$ predicts a high probability of eye movement.

Examples where $\pi_A$ is better than $\pi_B$ or $\pi_C$. Finally, the best model can be the constant one for some particular configurations and movements. This occurs mostly for fixations that are not directed on objects (for example Fig. 6a). Indeed model $\pi_A$ is simply the global distribution of eye movements that are mostly of low amplitude (see Fig. 3a) and the other models are mostly attracted to targets or the uncertainty attached to objects.

Fig. 6b 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. 6c, for which the situation is similar, the participant looked towards this group of targets, as predicted by both models $\pi_B$ and $\pi_C$. 
Figure 5: Examples of eye movements better predicted by model $\pi_B$ than model $\pi_C$. The scene is presented in an eye centered reference frame. The targets are in dark red and the distractors in light green. The actual eye movement recorded in this situation is indicated by the black cross. (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.

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 $\pi_C$.
Figure 6: Examples of eye movements better predicted by model $\pi_A$ than models $\pi_B$ or $\pi_C$. The scene is presented in an eye centered reference frame. The targets are in dark red and the distractors in light green. The actual eye movement recorded in this situation is indicated by the black cross. (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.

which assigns 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.

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, which means our model can only track as much targets as its number of target position variables. We used a fixed number of targets for the model, as it was suggested by the human performance in this particular experimental design (Tanner, in preparation). Note that this number 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. Unexpectedly, we found in the psychophysical data that the distribution of the objects’ positions is quite uniform in the logcomplex mapping. This implies 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.

Acknowledgements

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

A Experimental protocol

Figure 7: 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. Reprinted from Pylyshyn and Annan (2006)

This experiment is an adaptation of the classical Multiple Object Tracking paradigm from Pylyshyn and Storm (1988) (see Fig. 7) 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 Material and Methods

Participants Eleven subjects participated in the experiment with normal or corrected vision. Each session consists of 110 trials.
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 1024 × 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 darkened, sound attenuated room. 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°.

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/m² against a black background, in a room illuminated with diffuse D65 light (70 cd/m²).

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 1080 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 seconds 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 A in (Tanner, in preparation).

B Dynamic models

B.1 Dynamic object model

This dynamic model provides the transition probability distribution \( P(Occ^t_{(x,y)} \mid Mvt^t \ Occ^{t-1} ) \) 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 \((\rho, \theta)\) 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:

\[
P((x, y)_{\text{ant}} \ (\rho, \theta)_{\text{ant}} \ Occ^t_{(x,y)} \ Occ^{t-1} Mvt^t) = \ P((x, y))P(Mvt^t)P(Occ^t_{(x,y)})P((\rho, \theta) \ | \ (x, y)) \times P((\rho, \theta)_{\text{ant}} \ | \ (\rho, \theta) \ Mvt^t)P((x, y)_{\text{ant}} \ | \ (\rho, \theta)_{\text{ant}}) \times \prod_{(x', y')} P(Occ^t_{(x', y')} \ | \ Occ^t_{(x,y)} \ (x, y) + \text{ant})
\]

where:

- \( P((x, y)) \) is an arbitrary unused distribution;
- \( P(Mvt^t) \) is an arbitrary unused distribution;
- \( P(Occ^t_{(x,y)}) \) is a uniform distribution;
- \( P((\rho, \theta) \ | \ (x, y)) \) is a uniform distribution on the inverse image of the position \((x, y)\) by the logcomplex mapping;
- \( P((\rho, \theta)_{\text{ant}} \ | \ (\rho, \theta) \ Mvt^t) \) is a Dirac distribution on the image of \((\rho, \theta)\) by eye movement \(Mvt^t\);
- \( P((x,y)_{\text{ant}} \mid (\rho,\theta)_{\text{ant}}) \) is a Dirac distribution on the cell corresponding to position \((\rho,\theta)_{\text{ant}}\);
- \( P(Occt^{-1}(x',y') \mid Occ^t(x,y))(x^{-1},y^{-1}) \) 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 \( P(Occ^t_t(x,y) \mid Mvt^t \text{Occ}^{t-1}) \) using the following expression:

\[
P(Occ^t_t(x,y) \mid \text{Occ}^{t-1} \text{Mvt}^t)
\propto \sum_{(\rho,\theta)} P((\rho,\theta) \mid (x,y))P(Occ^t_t(x',y') \mid \text{Occ}^{t-1}(\tilde{x},\tilde{y}))
\]

where \((\tilde{x},\tilde{y})\) are the coordinates of the cell corresponding to the image of \((\rho,\theta)\) by eye motion \(\text{Mvt}^t\).

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 \( P(Tgt^t_i \mid Tgt^t_i-1 \text{Occ}^t \text{Mvt}^t) \) used in the representation model.

The decomposition is as follows:

\[
P(Tgt^t_i \mid Tgt^t_i-1 \text{Occ}^t \text{Mvt}^t)
= P(Tgt^t_i)P(Occt^t \mid \text{Mvt}^t) \times P((\rho,\theta)_{\text{ant}} \mid (\rho,\theta)\text{Mvt}^t)P(Tgt^t_i \mid (\rho,\theta)_{\text{ant}})
\times \prod_{(x,y)} P(Occ^t_t(x,y) \mid Tgt^t_i)
\]

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

This model is used to compute \( P(Tgt^t_i \mid Tgt^t_i-1 \text{Occ}^t \text{Mvt}^t) \) with the following expression:

\[
P(Tgt^t_i \mid Tgt^t_i-1 \text{Mvt}^t \text{Occ}^t)
\propto |\mathcal{E}(Tgt^t_i-1,\text{Mvt}^t)| \prod_{(x,y)} P(Occ^t_t(x,y) \mid Tgt^t_i)
\]

where \( |\mathcal{E}(Tgt^t_i-1,\text{Mvt}^t)| \) is the size of the set of the polar positions \((\rho,\theta)\) that are in relation with \( Tgt^t_i-1 \) by the eye movement \(\text{Mvt}^t\). This set can be obtained by sampling like in the dynamic model.
C Implementation details

The models presented are implemented in the Java language. In all the examples, the grid $G$ is composed of $24 \times 29$ 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, the observation model $P(Obs^t_{(x,y)} \mid Occ^t_{(x,y)})$ of the representation model is a $2 \times 2$ 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 $P(Occ^t_{(x,y)} \mid Tgt^t)$ is of the form $0.5 + \frac{0.25}{1 + \frac{d((x,y),Tgt^t)}{0.02}}$ for an occupied cell and $0.5 - \frac{0.25}{1 + \frac{d((x,y),Tgt^t)}{0.02}}$ otherwise with $d((x,y),Tgt^t)$ the distance between cell $(x,y)$ and position $Tgt^t$ in mm. The target fusion model $P(Tgt^t \mid Mot^t)$ is a mixture between a Gaussian and a uniform distribution: $\propto 0.2275 + \exp \left(-\frac{d(Tgt^t, Mot^t)}{0.25}\right)^2$. The uncertainty fusion distribution $P(I^t_{(x,y)} \mid Mot^t \pi_C)$ is a symmetrical beta distribution with parameter $0.075$.

References


