FP6-IST-027140

BACS

Bayesian Approach to Cognitive Systems

Instrument: IP
Thematic Priority: IST

D2.2: Report on tracking and saccade selection models

Due date of deliverable: June 30th, 2007

Actual submission date: June 26th, 2007

Start date of project: 01.01.2006 Duration: 50 month

Organisation name of lead contractor fort his deliverable: CNRS - LPPA

Revision 1

Project co-funded by the European Commission within the Sixth Framework Programme (2002–2006)		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Report on tracking and saccade selection models

Francis Colas Pierre Bessière Fabien Flacher Alain Berthoz
Benoît Girard

June 26, 2007

1 Introduction

Ocular movements are generally classified in three categories (Henn, 1993):

- slow movements, which can be stabilization of the overall image either by vestibular or optokinetic stimulation, or pursuit of a slow target in the visual field;
- rapid eye movements, which can be saccades, that is ballistic movements to look at some place in the visual field, or nystagmus, fast phase in stabilization movements to keep eye position is its mechanical limits;
- vergence movements: that adapt the angle between both eye according to the target distance. These movements will not be considered in the following.

The neuronal circuitry to handle these different movements has been thoroughly investigated in the past decades.

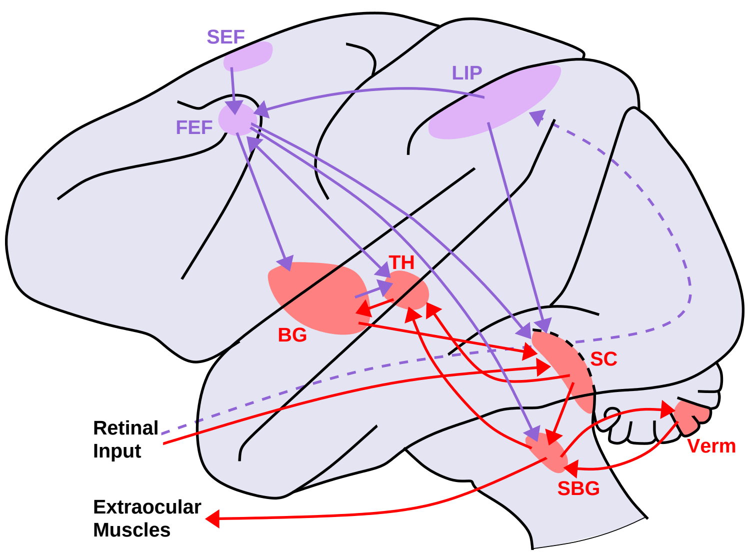

Figure 1: Saccadic circuitry (Macaque monkey). In red, short subcortical loop, in purple, long cortical loop. BG: basal ganglia, FEF: frontal eye fields, LIP: lateral bank of the intraparietal sulcus, SBG: saccade burst generators, SC: superior colliculus, SEF: supplementary eye fields, TH: thalamus, Verm: cerebellar vermis.

Among the numerous saccade related brain regions (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).

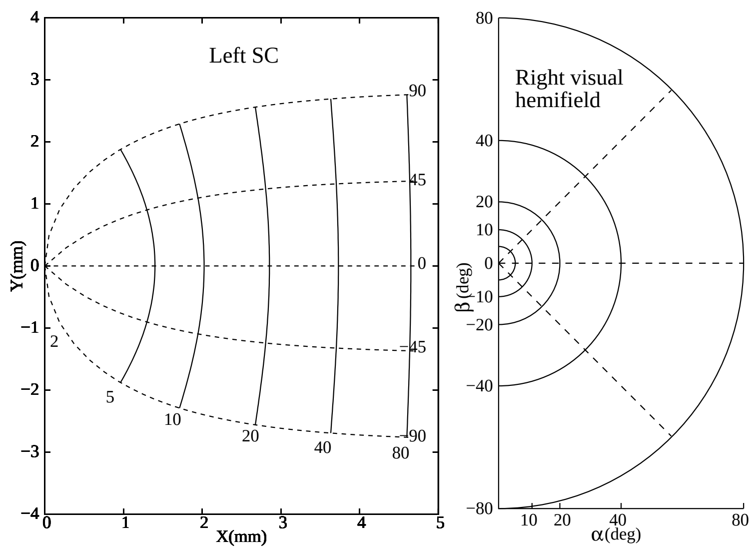

Figure 2: Macaque collicular mapping. The angular position of targets in the visual field (right) are mapped onto the SC surface (left) using a logarithmic mapping.

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 (see fig. 3). 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).

In this document we present models of retinotopic maps for eye movement selection. The aim of these model is to study whether uncertainty is used for eye movement selection. Uncertainty is the consequence of the inverse nature of perception, as well as incompleteness of the models. We choose the Bayesian Programming framework (Lebeltel et al., 2004) to handle and reason with the uncertainty.

More precisely, the models are designed to compute a sequence of probability distributions over the next eye movement to do, based on a sequence of observations of objects in the visual field. At each time step, the input is a set of positions of salient features, provided for instance by lower levels processes. Among these features, also called objects, some will be distinguished at the start as targets. Objects that are not targets will be refered to as distractors. The aim of the model is to remember where all the targets are, regardless of the distractors. This task follows the Multiple Object Tracking paradigm (Pylyshyn and

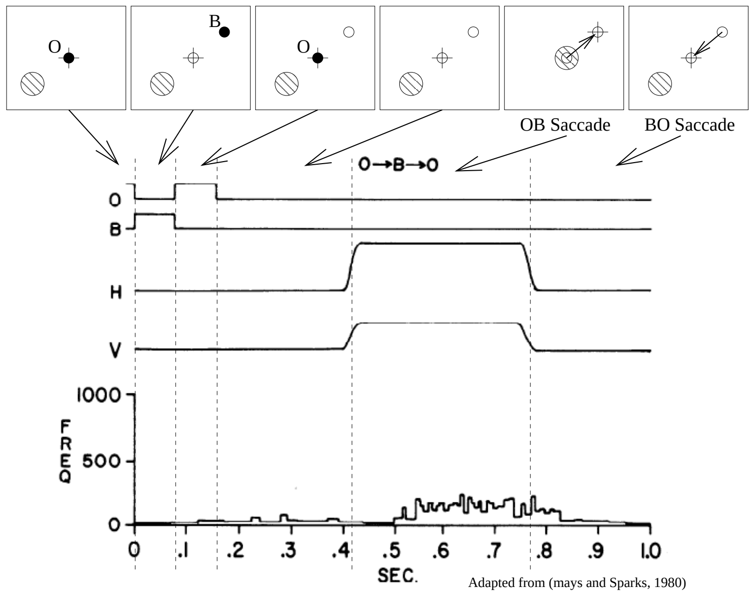

Figure 3: Specific activity of the quasi-visual cells (QV) in the SC. Upper row: the cross represents the current position of the eye center, the hashed surface is the receptive field of the recorded cell. The circles represent the target (B) and the fixation point (O), they are black when the light is on and white when it is off. The monkey is trained to perform an O-B saccade followed by a B-O saccade after the disappearance of both O and B. After the OB saccade, the position of the fixation point is inside the cell's receptive field, and the cell is activated despite the absence of a visual input for O.

Storm, 1988).

We propose a general architecture in two parts. The first part is a map model (see section 2) inspired from the retinotopic memory maps such as QV layers. As an example, figure 4 presents the result of this model.

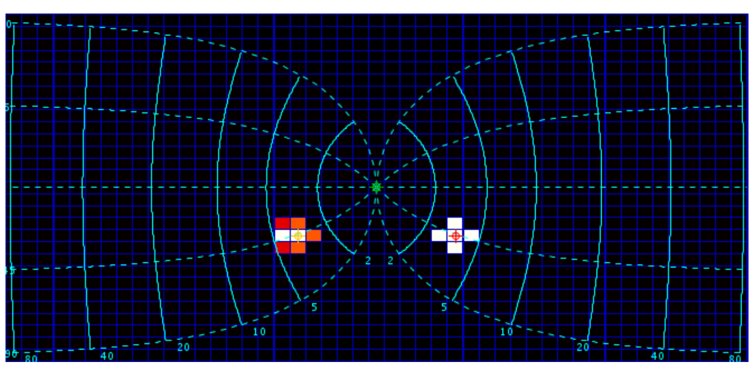

Figure 4: Example of a result of the map model. The two halves of the visual field are presented, each deformed according to the logarithmic mapping. In each cell, the color represents the probability of presence of an object from 0 (in black) to 1 (in white). In this example, two objects where present in the visual field. On the right, the object is stationary, hence a very high probability of presence detected by the map model. On the left, the object is in translation rightwards, hence a moderate probability of presence at the right and left edges of the object where it respectively appears and disappears.

The second part is a so-called motor model, working from the retinotopic map to choose the next eye movement. Figure 5 presents an example of probability distribution on the next eye movement based on the map present in figure 4. In section 3, we propose several of these motor models based on different hypotheses. These models aim at proving or disproving these hypotheses by way of comparison with experimental results.

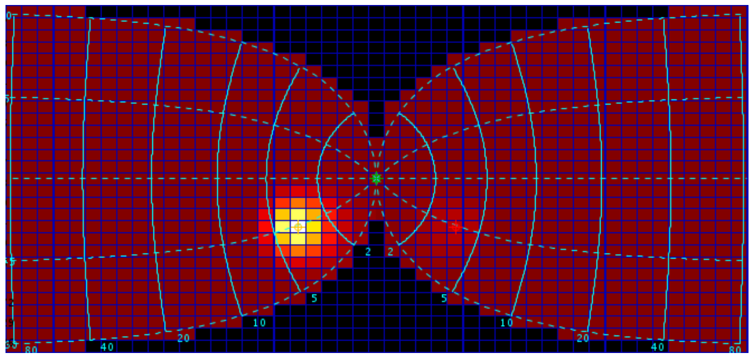

Figure 5: Example of a result of the uncertain model (see section 3.3) in the conditions of figure 4. The model computes a probability distribution over the next eye movement, represented here as the next position of the gaze. This distribution is normalized across all the possible positions. The cells outside of the visual field deformed by the logarithmic mapping are given a zero probability. In this example, there are two bumps of probability corresponding to both objects in the visual field. However, as the object on the left is moving, there is more uncertainty in its neighborhood and therefore a higher probability of looking in this direction.

2 Map Model

This model is based on an occupancy grid. The dynamic model only takes into account eye movement and not object motion. We introduce some notations and the variables of the model, then we present the structure of the map model, and finally the observation model and both dynamic models that are used in this map model.

2.1 Variables

Notations

- (x, y) will be used for the coordinates of a given cell in the retinotopic map;
- T, t note respectively the last temporal index and an arbitrary index before (or including) the last;
- $A^{t_1:t_2}$ is the conjunction of variables A^s for $s \in [\![t_1, t_2]\!]$;
- absence of coordinates subscript for variables defined in each cell of the map means the conjunction of all the associated in the map (note that there are also variables globally defined that do not have subscript at all).

Variables Contrary to usual occupancy grids, we will evaluate two quantities in each cell: the probability of presence of any object (be it target or distractor) and the probability of presence of a target. Therefore we have two binary state variables and one observation variable in each cell. Furthermore the model has access to the last eye movement completed:

- $Obs^t_{(x,y)}$: binary observation of the presence of an object (either target or distractor) at time t in cell (x, y);

- $Obj^t_{(x,y)}$: binary estimation of the presence of an object at time t in cell (x, y);
- $Tgt^t_{(x,y)}$: binary estimation of the presence of a target at time t in cell (x, y);
- Mvt^t: eye movement completed at time t (domain is the set of all (x, y)).

2.2 Main Model

Decomposition We specify the joint probability distribution over all these variables. To simplify this expression, we use conditional independencies that we summarize in a decomposition:

$$\begin{array}{rl} & P(Obs^{0:T}\ Obj^{0:T}\ Tgt^{0:T}\ Mvt^{0:T}) \\ = & P(Mvt^0) \prod\limits_{(x,y)} P(Obj^0_{(x,y)}) P(Obs^0_{(x,y)} \mid Obj^0_{(x,y)}) P(Tgt^0_{(x,y)}) \\ \times & \prod\limits_{t=1}^{T} P(Mvt^t) \prod\limits_{(x,y)} \left[\begin{array}{l} P(Obj^t_{(x,y)} \mid Obj^{t-1}\ Mvt^t) \\ P(Obs^t_{(x,y)} \mid Obj^t_{(x,y)}) \\ P(Tgt^t_{(x,y)} \mid Tgt^{t-1}\ Mvt^t\ Obj^t_{(x,y)}) \end{array} \right. \end{array}$$

This decomposition states that the cells are conditionnally independent of the others, given eye movement and the previous state of the map. Furthermore, in each cell we can recognize a structure similar to a hidden Markov model on $Obj^t_{(x,y)}$ with observation $Obs^t_{(x,y)}$.

Parametrical Forms To be able to do the computation, we have to specify each distribution that appear in the decomposition:

- $\forall t \in [\![0, T]\!], P(Mvt^t)$ unknown (will disappear in the inferences due to our choice of questions);
- $\forall (x, y), P(Obj^0_{(x,y)})$ initial state of the map, whose influence decreases rapidly with time;
- $\forall (x, y), P(Tgt^0_{(x,y)})$ prior according to the initial distribution of target/distractor.
- $\forall (x, y), t \in [\![0, T]\!], P(Obs^t_{(x,y)} \mid Obj^t_{(x,y)})$ observation model (see section 2.3);
- $\forall (x, y), t \in [\![0, T]\!], P(Obj^t_{(x,y)} \mid Obj^{t-1}\ Mvt^t)$ dynamic model of displacement of objects in the map when the eyes are moving (see section 2.4);
- $\forall (x, y), t \in [\![0, T]\!], P(Tgt^t_{(x,y)} \mid Tgt^{t-1}\ Mvt^t\ Obj^t_{(x,y)})$ dynamic model of the displacement of targets in the map when the eyes are moving (see section 2.5).

Questions and Inferences The first question is the update of the "object map":

$$\forall (x, y), P(Obj^T_{(x,y)} \mid Obs^{0:T}\ Mvt^{0:T}) \tag{1}$$

The Bayesian inference leads to:

$$\begin{array}{rl} & P(Obj^T_{(x,y)} \mid Obs^{0:T}\ Mvt^{0:T}) \\ \propto & P(Obs^T_{(x,y)} \mid Obj^T_{(x,y)}) \\ \times & \sum\limits_{Obj^{T-1}} \left[\begin{array}{l} \prod_{(x',y')} P(Obj_{(x',y')} \mid Obs^{0:T-1}\ Mvt^{0:T-1}) \\ P(Obj^T_{(x,y)} \mid Obj^{T-1}\ Mvt^T) \end{array} \right. \end{array}$$

This expression allows for a recursive estimation of this question. Due to the dynamic model, the summation will only be needed on a subset of antecedent cells (defined by the eye movement) which will further reduce the complexity of the calculation (see section 2.4).

The second question is the update of the "target map":

$$\forall(x,y), P(Tgt^T_{(x,y)} \mid Obs^{0:T}\ Mvt^{0:T}) \tag{2}$$

Again, the Bayesian inference leads to:

$$\begin{array}{ll} & P(Tgt^T_{(x,y)} \mid Obs^{0:T}\ Mvt^{0:T}_{(x,y)}) \\ \propto & \sum_{Obj^T_{(x,y)}} \left[\begin{array}{l} P(Obj^T_{(x,y)} \mid Obs^{0:T}\ Mvt^{0:T}) \\ \sum_{Tgt^{T-1}} \left[\begin{array}{l} \prod_{(x',y')} P(Tgt_{(x',y')} \mid Obs^{0:T-1}\ Mvt^{0:T-1}) \\ P(Tgt^T_{(x,y)} \mid Tgt^{T-1}\ Mvt^T\ Obj^T_{(x,y)}) \end{array} \right. \end{array} \right. \end{array}$$

This expression is the same as for the first question, with the small addition of both alternative with respect to the presence or absence of object. See dynamic model of targets for details.

The expressions of these inferences do only depend on the structure of the model. However, instanciated parametric forms are needed to do actual computations.

2.3 Observation Model

The observation model is simply a 2-by-2 matrix describing the probability to observe an object when there is one and when there isn't any.

Parameters The are two free parameters. We decide arbitrarily to have them represent the right observation probability o_0 and o_1. The probability table is therefore:

$Obs^T_{(x,y)} \backslash Obj^T_{(x,y)}$	False	True
False	o_0	$1-o_1$
True	$1-o_0$	o_1

Table 1: Observation model.

2.4 Dynamic Model

Contrary to the observation model, that could be directly defined as a conditional distribution, we will build a whole Bayesian program to which we will ask the questions needed in the map model.

This subprogram will explicitely refer to visual coordinates, as eye movements are expessed in this reference frame.

Variables Coordinates will appear as variables. Therefore, there will be only one model instanciated with different coordinates rather than a collection of models each related to only one given cell.

- (x, y): coordinates of the cell in the logcomplex map;
- (x^{-1}, y^{-1}): coordinates of an antecedent cell in the logcomplex map;
- (ρ, θ): coordinates in the polar reference frame;
- (ρ^{-1}, θ^{-1}): coordinates of antecedent in polar frame;
- $Obj^t_{(x,y)}$: presence of an object in the cell corresponding to (x, y);
- Obj^{t-1}: presence of an object in each position at time $t-1$;
- Mvt^t: eye motion completed.

Decomposition The decomposition is as follows:

$$
\begin{aligned}
& P((x,y)\ (x^{-1},y^{-1})\ (\rho,\theta)\ (\rho^{-1},\theta^{-1})\ Obj^t_{(x,y)}\ Obj^{t-1}\ Mvt^t) \\
= \ & P((x,y))P(Mvt^t)P(Obj^t_{(x,y)}) \\
\times \ & P((\rho,\theta) \mid (x,y))P((\rho^{-1},\theta^{-1}) \mid (\rho,\theta)\ Mvt^t)P((x^{-1},y^{-1}) \mid (\rho^{-1},\theta^{-1})) \\
\times \ & \prod_{(x',y')} P(Obj^{t-1}_{(x',y')} \mid Obj^t_{(x,y)}\ (x^{-1},y^{-1}))
\end{aligned}
$$

Parametrical Forms We give a parametrical form for each of the distributions involved in the decomposition:

- $P((x,y))$: unknown;
- $P(Mvt^t)$: unknown;
- $P(Obj^t_{(x,y)})$: uniform;
- $P((\rho,\theta) \mid (x,y))$: uniform distribution on the inverse image of the cell (x,y) by the logcomplex mapping;
- $P((\rho^{-1},\theta^{-1}) \mid (\rho,\theta)\ Mvt^t)$: dirac on the image of (ρ,θ) by eye movement Mvt^t;
- $P((x^{-1},y^{-1}) \mid (\rho^{-1},\theta^{-1}))$: dirac on the cell corresponding to position (ρ^{-1},θ^{-1});
- $P(Obj^{t-1}_{(x',y')} \mid Obj^t_{(x,y)}\ (x^{-1},y^{-1}))$: sharp 2-by-2 matrix on the value of $Obj^t_{(x,y)}$ if $(x',y') = (x^{-1},y^{-1})$, uniform otherwise.

Question This model is designed to answer the questions $P(Obj^t_{(x,y)} \mid Obj^{t-1}\ Mvt^t)$ for the main model:

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

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

We implement this summation by sampling the distribution $P((\rho,\theta) \mid (x,y))$.

2.5 Dynamic Model of Target

This model is essentially the same as for plain objects with the addition of object presence as a pseudo observation.

Variables

- (x,y): coordinates of the cell in the logcomplex map;
- (x^{-1},y^{-1}): coordinates of an antecedent cell in the logcomplex map;
- (ρ,θ): coordinates in the polar reference frame;
- (ρ^{-1},θ^{-1}): coordinates of antecedent in polar frame;
- $Tgt^t_{(x,y)}$: presence of a target in the cell corresponding to (x,y);
- $Obj^t_{(x,y)}$: presence of an object in the cell corresponding to (x,y);
- Tgt^{t-1}: presence of a target in each position at time $t-1$;
- Mvt^t: eye motion completed.

Decomposition

$$
\begin{aligned}
& P((x,y)\ (x^{-1},y^{-1})\ (\rho,\theta)\ (\rho^{-1},\theta^{-1})\ Tgt^t_{(x,y)}\ Tgt^{t-1}\ Mvt^t\ Obj^t_{(x,y)}) \\
= \quad & P((x,y))P(Mvt^t)P(Tgt^t_{(x,y)})P(Obj^t_{(x,y)} \mid Tgt^t_{(x,y)}) \\
\times \quad & P((\rho,\theta) \mid (x,y))P((\rho^{-1},\theta^{-1}) \mid (\rho,\theta)\ Mvt^t)P((x^{-1},y^{-1}) \mid (\rho^{-1},\theta^{-1})) \\
\times \quad & \prod_{(x',y')} P(Tgt^{t-1}_{(x',y')} \mid Tgt^t_{(x,y)}\ (x^{-1},y^{-1}))
\end{aligned}
$$

Parametrical Forms

- $P((x,y))$: unknown;
- $P(Mvt^t)$: unknown;
- $P(Tgt^t_{(x,y)})$: uniform;
- $P(Obj^t_{(x,y)} \mid Tgt^t_{(x,y)})$: $(1-o'_1, o'_1)$ if target, uniform otherwise;
- $P((\rho,\theta) \mid (x,y))$: uniform distribution on the inverse image of the cell (x,y) by the logcomplex mapping;
- $P((\rho^{-1},\theta^{-1}) \mid (\rho,\theta)\ Mvt^t)$: dirac on the image of (ρ,θ) by eye movement Mvt^t;
- $P((x^{-1},y^{-1}) \mid (\rho^{-1},\theta^{-1}))$: dirac on the cell corresponding to position (ρ^{-1},θ^{-1});
- $P(Tgt^{t-1}_{(x',y')} \mid Tgt^t_{(x,y)}\ (x^{-1},y^{-1}))$: sharp 2-by-2 matrix on the value of $Tgt^t_{(x,y)}$ if $(x',y') = (x^{-1},y^{-1})$, uniform otherwise.

Question This model is designed to answer the questions $P(Tgt^t_{(x,y)} \mid Tgt^{t-1}\ Mvt^t\ Obj^t_{(x,y)})$ for the main model:

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

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

We implement this summation in the same way as for the object dynamic model by sampling the distribution $P((\rho,\theta) \mid (x,y))$.

3 Motor Models

Based on this unique map representation, we wish to compare different models to choose eye movement. The first one will not rely on the map. Then we will test a model relying only on the presence of targets and objects, and finally a last model relying on the uncertainty of targets and objects.

3.1 Constant Model

This model will be the baseline to which compare the next models. The underlying assumption is that eye movement depends neither on observations, nor on history during the experiment.

Variable, Decomposition, Inference This model has only one variable Mot^T, the motor command planned. Therefore the "joint" distribution is reduced to $P(Mot^T)$, which is also the question.

Parametrical Form This model is specified with a histogram learned from data from human observers.

3.2 Classical Model

This models relies on the assumption that motor decisions are taken based on a given distribution of targets and distractors.

Variables This models includes the past observations $Obs^{0:T}$, the past actions $Mvt^{0:T}$, the presence of objects and targets Obj^T, Tgt^T, and the planned motor command Mot^T.

Decomposition We make the hypothesis that the planned motor command does not depend on past commands and observations, conditionnally to the knowledge of the positions of objects and targets. Thus we can write the following decomposition:

$$\begin{aligned} & P(Obs^{0:T}\ Mvt^{0:T}\ Obj^T\ Tgt^T\ Mot^T) \\ = \ & P(Obs^{0:T}\ Mvt^{0:T})P(Obj^T \mid Obs^{0:T}\ Mvt^{0:T})P(Tgt^T \mid Obs^{0:T}\ Mvt^{0:T})P(Mot^T \mid Obj^T\ Tgt^T) \end{aligned}$$

Parametrical Forms We specify each distribution in the above decomposition. $P(Mot^T \mid Obj^T\ Tgt^T)$ is defined by a fusion submodel.

- $P(Obs^{0:T}\ Mvt^{0:T})$: unknown;
- $P(Obj^T \mid Obs^{0:T}\ Mvt^{0:T})$: question to the map model (equation 1);
- $P(Tgt^T \mid Obs^{0:T}\ Mvt^{0:T})$: question to the map model (equation 2);
- $P(Mot^T \mid Obj^T\ Tgt^T)$: question to following submodel:

$$P(Mot^T\ Obj^T\ Tgt^T) = P(Mot^T) \prod_{(x,y)} P(Obj^T_{(x,y)} \mid Mot^T)P(Tgt^T_{(x,y)} \mid Mot^T)$$

  where $P(Obj^T_{(x,y)} \mid Mot^T)$ and $P(Tgt^T_{(x,y)} \mid Mot^T)$ are learnt or specified by hand.

Question The question is $P(Mot^T \mid Obs^{0:T}\ Mvt^{0:T})$. The Bayesian inference yields ultimately:

$$\begin{aligned} & P(Mot^T \mid Obs^{0:T}\ Mvt^{0:T}) \\ \propto \ & \sum_{Obj^T, Tgt^T} \prod_{(x,y)} P(Obj^T_{(x,y)} \mid Obs^{0:T}\ Mvt^{0:T})P(Tgt^T_{(x,y)} \mid Obs^{0:T}\ Mvt^{0:T})P(Mot^T \mid Obj^T\ Tgt^T) \end{aligned}$$

This inference can be approximated by sampling the map space or even choosing the most probable map as a representing instance.

3.3 Uncertain Model

This model relies on the hypothesis that eye movement is based on uncertainty on the visual scene, rather than on the positions of the objects and targets.

Variables Like the classical model, this one includes observations, actions, presence of objects and targets, and planned motor commands. This model also includes a last set of variables, $(I^T_{Obj_{(x,y)}}, I^T_{Tgt_{(x,y)}})$, to represent the uncertainty attached to the presence of targets and objects at a given location.

Decomposition The deomposition is similar to the classical model with the additional uncertainty variables I^t:

$$
\begin{aligned}
& P(Obs^{0:T}\ Mvt^{0:T}\ Obj^T\ Tgt^T\ Mot^T\ I^T_{Obj}\ I^T_{Tgt}) \\
= \ & P(Obs^{0:T}\ Mvt^{0:T}) P(Obj^T \mid Obs^{0:T}\ Mvt^{0:T}) P(Tgt^T \mid Obs^{0:T}\ Mvt^{0:T}) \\
\times \ & \prod_{(x,y)} P(I^T_{Obj_{(x,y)}} \mid Obs^{0:T}\ Mvt^{0:T}) P(I^T_{Tgt_{(x,y)}} \mid Obs^{0:T}\ Mvt^{0:T}) P(Mot^T \mid I^T)
\end{aligned} \tag{3}
$$

Parametrical Forms Again, distribution on Mot^T is defined using a fusion submodel:

- $P(Obs^{0:T}\ Mvt^{0:T})$: unknown;
- $P(Obj^T \mid Obs^{0:T}\ Mvt^{0:T})$: question to map model (equation 1);
- $P(Tgt^T \mid Obs^{0:T}\ Mvt^{0:T})$: question to map model (equation 2);
- $P(I^T_{Obj_{(x,y)}} \mid Obs^{0:T}\ Mvt^{0:T})$: dirac on $P(obj^T_{(x,y)} \mid Obs^{0:T}\ Mvt^{0:T})$;
- $P(I^T_{Tgt_{(x,y)}} \mid Obs^{0:T}\ Mvt^{0:T})$: dirac on $P(tgt^T_{(x,y)} \mid Obs^{0:T}\ Mvt^{0:T})$;
- $P(Mot^T \mid I^T)$: question to submodel:

$$
P(Mot^T\ I^T) = P(Mot^T) \prod_{(x,y)} P(I^T_{Obj_{(x,y)}} \mid Mot^T) P(I^T_{Tgt_{(x,y)}} \mid Mot^T)
$$

  where $P(I^T_{Obj_{(x,y)}} \mid Mot^T)$ and $P(I^T_{Tgt_{(x,y)}} \mid Mot^T)$ are learned or specified by hand.

Question The question is $P(Mot^T \mid Obs^{0:T}\ Mvt^{0:T})$. Bayesian inference yields ultimately:

$$
\begin{aligned}
& P(Mot^T \mid Obs^{0:T}\ Mvt^{0:T}) \\
\propto \ & P(Mot^T) \prod_{(x,y)} P(i^T_{Obj_{(x,y)}} \mid Mot^T) P(i^T_{Tgt_{(x,y)}} \mid Mot^T)
\end{aligned}
$$

where $i^T_{Obj_{(x,y)}} = P(obj^T_{(x,y)} \mid Obs^{0:T}\ Mvt^{0:T})$ and $i^T_{Tgt_{(x,y)}} = P(tgt^T_{(x,y)} \mid Obs^{0:T}\ Mvt^{0:T})$ as answered by the map model above (equations 1 and 2).

4 Conclusion and future work

We presented in this report a first Bayesian model of retinotopic map of objects and targets. Then we detailed three different models of eye movement (being tracking or saccade) based on different hypotheses.

We plan to develop further the parametrical forms in these models as well as learning the parameters from experimental data. We also plan to use experimental data to compare the various eye movement models we proposed.

References

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

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

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

Bozis, A. and 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.

Droulez, J. and Berthoz, A. (1991). A neural network model of sensoritopic maps with predictive short-term memory properties. *Proceedings of the National Academy of Science*, 88:9653–9657.

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

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

Henn, V. (1993). Neuronal control of eye movements. In Berthoz, A., editor, *Multisensory control of movement*, pages 7–26. Oxford University Press, Oxford, UK.

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

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

Mays, L. and 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., editor, *Neurophysiology II*, volume 10 of *Int Rev Physiol*, pages 223–248. University Park Press, Baltimore.

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

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

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

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

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

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

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

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

Siminoff, R., Schwassmann, H., and 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. and Wurtz, R. (2000). Composition and topographic organization of signals sent from the frontal eye fields to the superior colliculus. *Journal of Neurophysiology*, 83:1979–2001.

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