Bayesian Models for Active Multimodal Perception

João Filipe Ferreira, Cátia Pinho, Jorge Lobo, Pierre Bessière and Jorge Dias

Workshop on Bayesian Models and Contraction Theory, Chamonix, France, January 2008
Outline

1. Introduction
   - Motivations
   - Expected Scientific Contributions
   - Expected Practical Contributions

2. Bayesian Volumetric Maps for Multimodal Integration
   - Occupancy Grids and Spatial Configuration
   - Bayesian Sensor Models
   - Bayesian Multimodal Sensor Fusion

3. Ongoing and Future Work
   - What Next?
   - Entropy-Based Active Exploration

4. Bibliography
Outline

1 Introduction
- Motivations
- Expected Scientific Contributions
- Expected Practical Contributions

2 Bayesian Volumetric Maps for Multimodal Integration
- Occupancy Grids and Spatial Configuration
- Bayesian Sensor Models
- Bayesian Multimodal Sensor Fusion

3 Ongoing and Future Work
- What Next?
- Entropy-Based Active Exploration

4 Bibliography
Outline

1 Introduction
   • Motivations
   • Expected Scientific Contributions
   • Expected Practical Contributions

2 Bayesian Volumetric Maps for Multimodal Integration
   • Occupancy Grids and Spatial Configuration
   • Bayesian Sensor Models
   • Bayesian Multimodal Sensor Fusion

3 Ongoing and Future Work
   • What Next?
   • Entropy-Based Active Exploration

4 Bibliography
Outline

1 Introduction
   • Motivations
   • Expected Scientific Contributions
   • Expected Practical Contributions

2 Bayesian Volumetric Maps for Multimodal Integration
   • Occupancy Grids and Spatial Configuration
   • Bayesian Sensor Models
   • Bayesian Multimodal Sensor Fusion

3 Ongoing and Future Work
   • What Next?
   • Entropy-Based Active Exploration

4 Bibliography
Outline

1 Introduction
   - Motivations
   - Expected Scientific Contributions
   - Expected Practical Contributions

2 Bayesian Volumetric Maps for Multimodal Integration
   - Occupancy Grids and Spatial Configuration
   - Bayesian Sensor Models
   - Bayesian Multimodal Sensor Fusion

3 Ongoing and Future Work
   - What Next?
   - Entropy-Based Active Exploration

4 Bibliography
Outline

1 Introduction
   - Motivations
   - Expected Scientific Contributions
   - Expected Practical Contributions

2 Bayesian Volumetric Maps for Multimodal Integration
   - Occupancy Grids and Spatial Configuration
   - Bayesian Sensor Models
   - Bayesian Multimodal Sensor Fusion

3 Ongoing and Future Work
   - What Next?
   - Entropy-Based Active Exploration

4 Bibliography
Main Scientific Motivation
From Perceptual Epistemology to a Perceptual Problem

Problem — An observer watches a non-static 3D scene containing several moving objects:

- How does the observer perceive:
  - his own motion (egomotion);
  - the 3D structure of objects in the scene;
  - the 3D trajectory and velocity of moving objects (independent motion)?
Main Scientific Motivation
From Perceptual Epistemology to a Perceptual Problem

**Problem** — An observer watches a non-static 3D scene containing several moving objects:

- How does the observer perceive:
  - his own motion (*egomotion*);
  - the *3D structure* of objects in the scene;
  - the 3D trajectory and velocity of moving objects (*independent motion*)?
The Bayesian Framework
Perceptual Uncertainty and Ambiguity

- Biological uncertainties:
  - physical constraints on sensors
  - discretisation
    (analogue-to-spike train)
  - neural noise (firing apparently not due to stimuli)
  - structural constraints on
    neural representations and computations

- Artificial uncertainties:
  - sensor accuracy and precision
  - discretisation
    (analogue-to-digital)
  - noise not accounted by artificial perception models
  - round-off effects and data representation limitations

- Ambiguities:
Bayesian Models for Active Multimodal Perception

Introduction
Motivations

The Bayesian Framework
Perceptual Uncertainty and Ambiguity

- Biological uncertainties:
  - physical constraints on sensors
  - discretisation (analogue-to-spike train)
  - neural noise (firing apparently not due to stimuli)
  - structural constraints on neural representations and computations

- Artificial uncertainties:
  - sensor accuracy and precision
  - discretisation (analogue-to-digital)
  - noise not accounted by artificial perception models
  - round-off effects and data representation limitations

Ambiguities:
The Bayesian Framework
Perceptual Uncertainty and Ambiguity

- Biological uncertainties:
  - physical constraints on sensors
  - discretisation (analogue-to-spike train)
  - neural noise (firing apparently not due to stimuli)
  - structural constraints on neural representations and computations

- Artificial uncertainties:
  - sensor accuracy and precision
  - discretisation (analogue-to-digital)
  - noise not accounted by artificial perception models
  - round-off effects and data representation limitations

- Ambiguities:
Recently, several hypothesis regarding biological perception have surfaced [Ferreira and Castelo-Branco(2007)]:

1. Perception as a process of unconscious, probabilistic inference ⇒ deals with perceptual uncertainty, ambiguity and conflicts.
2. The brain coding uncertainty in its internal representations and computations (e.g., population codes).
3. Perceptual brain is not feedforward ⇒ complex network of connections, although some modularity is preserved.
Recently, several hypothesis regarding biological perception have surfaced [Ferreira and Castelo-Branco(2007)]:

1. Perception as a process of unconscious, probabilistic inference ⇒ deals with perceptual uncertainty, ambiguity and conflicts.
2. The brain coding uncertainty in its internal representations and computations (e.g., population codes).
3. Perceptual brain is not feedforward ⇒ complex network of connections, although some modularity is preserved.
Let $\Phi$ be a variable to represent some characteristics of the observed phenomenon and $S$ be a variable to represent a sensation.

\[ P(\Phi S) = P(\Phi)P(S|\Phi) \]

- $P(\Phi)$ is the prior over the phenomenon;
- $P(S|\Phi)$ is the likelihood of the sensation, given the phenomenon: it is the direct sensor model.
- $P(\Phi S)$ is the corresponding joint probability distribution.
The probabilistic question of perception is \( P(\Phi|S) \), the probability distribution on the phenomenon, in view of some given sensation.

This question, which is the posterior distribution on the phenomenon after some observation, is solved by standard Bayesian inference:

\[
P(\Phi|S) = \frac{P(\Phi S)}{P(S)} = \frac{P(\Phi S)}{\sum_{\Phi} P(\Phi S)} = \frac{P(\Phi)P(S|\Phi)}{\sum_{\Phi} P(\Phi)P(S|\Phi)}
\]

This expression can be computed since it only involves the prior and the likelihood that are specified in the decomposition.
This research is also integrated within the PhD work titled Bayesian Cognitive Models for 3D Structure and Motion Multimodal Perception — http://paloma.isr.uc.pt/~jfilipe/BayesianMultimodalPerception.
Outline

1 Introduction
   - Motivations
   - Expected Scientific Contributions
   - Expected Practical Contributions

2 Bayesian Volumetric Maps for Multimodal Integration
   - Occupancy Grids and Spatial Configuration
   - Bayesian Sensor Models
   - Bayesian Multimodal Sensor Fusion

3 Ongoing and Future Work
   - What Next?
   - Entropy-Based Active Exploration

4 Bibliography
We mainly expect to contribute in developing computational models which:

1. are based on the perceptual modalities of vision, audition and vestibular sensing;
2. perform perceptual fusion within a Bayesian framework;
3. do not involve processes such as scene interpretation, classification (except for "background" / “independently moving” object segmentation and labelling), as in the perceptual brain’s dorsal pathway(s).
4. will serve as a framework for implementing short-term egocentric spatial memory for active perception.
We mainly expect to contribute in developing computational models which:

1. are based on the perceptual modalities of vision, audition and vestibular sensing;
2. perform perceptual fusion within a Bayesian framework;
3. do not involve processes such as scene interpretation, classification (except for “background”/“independently moving” object segmentation and labelling), as in the perceptual brain's dorsal pathway(s).
4. will serve as a framework for implementing short-term egocentric spatial memory for active perception.
Bayesian Models for Active Multimodal Perception

Introduction

Expected Scientific Contributions

Main Goals

We mainly expect to contribute in developing computational models which:

1. are based on the perceptual modalities of vision, audition and vestibular sensing;
2. perform perceptual fusion within a Bayesian framework;
3. do not involve processes such as scene interpretation, classification (except for “background”/“independently moving” object segmentation and labelling), as in the perceptual brain’s dorsal pathway(s);
4. will serve as a framework for implementing short-term egocentric spatial memory for active perception.
We mainly expect to contribute in developing computational models which:

1. are based on the perceptual modalities of vision, audition and vestibular sensing;
2. perform perceptual fusion within a Bayesian framework;
3. do not involve processes such as scene interpretation, classification (except for “background” / “independently moving” object segmentation and labelling), as in the perceptual brain’s dorsal pathway(s).
4. will serve as a framework for implementing short-term egocentric spatial memory for active perception.
We mainly expect to contribute in developing computational models which:

1. are based on the perceptual modalities of vision, audition and vestibular sensing;
2. perform perceptual fusion within a Bayesian framework;
3. do not involve processes such as scene interpretation, classification (except for “background”/“independently moving” object segmentation and labelling), as in the perceptual brain’s dorsal pathway(s).
4. will serve as a framework for implementing short-term egocentric spatial memory for active perception.
Outline

1 Introduction
   - Motivations
   - Expected Scientific Contributions
   - Expected Practical Contributions

2 Bayesian Volumetric Maps for Multimodal Integration
   - Occupancy Grids and Spatial Configuration
   - Bayesian Sensor Models
   - Bayesian Multimodal Sensor Fusion

3 Ongoing and Future Work
   - What Next?
   - Entropy-Based Active Exploration

4 Bibliography
An artificial multimodal perception system with gaze control capabilities for image stabilisation and perceptual attention purposes has been constructed, consisting of:

1. a stereovision setup;
2. a binaural setup;
3. a motorised head platform, with inertial sensors emulating the vestibular system.
An artificial multimodal perception system with gaze control capabilities for image stabilisation and perceptual attention purposes has been constructed, consisting of:

1. a stereovision setup;
2. a binaural setup;
3. a motorised head platform, with inertial sensors emulating the vestibular system.
Outline

1 Introduction
  - Motivations
  - Expected Scientific Contributions
  - Expected Practical Contributions

2 Bayesian Volumetric Maps for Multimodal Integration
  - Occupancy Grids and Spatial Configuration
  - Bayesian Sensor Models
  - Bayesian Multimodal Sensor Fusion

3 Ongoing and Future Work
  - What Next?
  - Entropy-Based Active Exploration

4 Bibliography
Outline

1 Introduction
   - Motivations
   - Expected Scientific Contributions
   - Expected Practical Contributions

2 Bayesian Volumetric Maps for Multimodal Integration
   - Occupancy Grids and Spatial Configuration
   - Bayesian Sensor Models
   - Bayesian Multimodal Sensor Fusion

3 Ongoing and Future Work
   - What Next?
   - Entropy-Based Active Exploration

4 Bibliography
Bayesian Models for Active Multimodal Perception
Bayesian Volumetric Maps for Multimodal Integration
Occupancy Grids and Spatial Configuration

**Spatial Perception Maps**

**Binding Sensation to Space**

**Problem**

How do we relate sensorial data with a probabilistic spatial representation of the perception of the three-dimensional scene?
**Metric maps** are very intuitive, yield a rigorous model of the environment and help to register measurements taken from different locations.

**Grid-based maps** are the most popular metric maps in mobile robotics applications:

- Metric maps allow storage of a high level of detail of the environment.
- Extremely useful to promote direct action, where precise size and time-to-collision estimates are needed.
- **However**, they are not well suited for exploration, since they do not commonly scale gracefully with surveying region size [Rocha et al.(2005)].
One of the most popular grid-based maps is the **occupancy grid**.

When considered in 3D, a grid divides the workspace into equally sized voxels, with each edge aligned with one of the axes of a reference coordinate frame.

Occuancy (and any other property) of each voxel is modelled using **probability values**.

When considering the dynamics of the environment (i.e. local motion), an extension of the occupancy grid has been devised called the **Bayesian Occupancy Filter (BOF)** — see [Coué et al.(2006)] and [Tay et al.(2007)].
Bayesian Volumetric Maps

Binding Sensation to Space — The Proposed Solution

Bayesian Volumetric Map

- It is egocentric (log-spherical and head-centred), metric (volumetric grid) and Bayesian in nature;
- Allows high-detail for close objects, while retaining regular grid (i.e. spatial multiresolution compression);
- (like BOF) it allows for the representation of dynamical spatial occupation (i.e. local motion), postponing data association to higher-level processing.
Bayesian Volumetric Maps

Binding Sensation to Space — The Proposed Solution

Bayesian Volumetric Map — Parametrisation

- **Domain:** \( \mathcal{Y} \equiv [\log_b \rho_{\text{Min}}; \log_b \rho_{\text{Max}}] \times [\theta_{\text{Min}}; \theta_{\text{Max}}] \times [\phi_{\text{Min}}; \phi_{\text{Max}}] \).

- **Distance Boundaries:**
  - *Egocentric gap* \( \rho_{\text{Min}} \), which accounts for sensors’ physical space.
  - *Maximum reach* \( \rho_{\text{Max}} \).

- **Cells:**
  - **Partitioning:** \( \mathcal{Y} \supset \mathcal{C} \equiv [\log_b \rho_{\text{min}}; \log_b \rho_{\text{max}}] \times [\theta_{\text{min}}; \theta_{\text{max}}] \times [\phi_{\text{min}}; \phi_{\text{max}}] \).
  - **Indexing:** each BVM cell is indexed by the coordinates of its far corner — \( \mathcal{C} = (\log_b \rho_{\text{max}}, \theta_{\text{max}}, \phi_{\text{max}}) \in \mathcal{C} \subset \mathcal{Y} \).

- **Resolution:** log-distance base \( b = a \frac{\log a \rho_{\text{Max}} - \rho_{\text{Min}}}{N}, \forall a \in \mathbb{R} \), for \( N \) partitions, and angular ranges \( \Delta \theta = \theta_{\text{Max}} - \theta_{\text{Min}} \) and \( \Delta \phi = \phi_{\text{Max}} - \phi_{\text{Min}} \).
Outline

1 Introduction
   - Motivations
   - Expected Scientific Contributions
   - Expected Practical Contributions

2 Bayesian Volumetric Maps for Multimodal Integration
   - Occupancy Grids and Spatial Configuration
   - Bayesian Sensor Models
   - Bayesian Multimodal Sensor Fusion

3 Ongoing and Future Work
   - What Next?
   - Entropy-Based Active Exploration

4 Bibliography
Our motivations suggest data structure analogous to neuronal population activity patterns [Pouget et al.(2000)].

A spatially organised 2D grid has each cell (corresponding to a photoreceptor) associated to a population code encoding a probability distribution function or pdf (on the right).
Vision sensors in IMPEP yield **stereoscopic depth** and **local motion** readings using probabilistic stereovision and local motion algorithms in real-time;

Readings are given in population code data structures in retinotopic fashion and referred to the cyclopean view (i.e. **egocentric head-centred coordinates**);

Each photoreceptor is described by its respective projection line’s direction in **spherical angles** $(\theta, \phi)$, and can thus be treated as a separate sensor.
Visual occlusion implies:

- All voxels between the camera and the voxel corresponding to the depth estimate are assumed as not occupied.
- The voxel corresponding to the depth estimate has a high probability of being occupied.
- The occupancy state of the remaining voxels should be taken as unknown.
This corresponds to an adaptation of the original probabilistic 1D range sensor model by [Elfes(1992)] introduced by [Yguel et al.(2007)] for use with the BOF formalism.

The pdfs from the population code data structure are converted from disparity ($\delta$) to depth ($\rho$) in egocentric, log-spherical coordinates → easy, due to cyclopean, polar configuration of vision sensors! → the relation between vision sensor measurements $Z$ and the corresponding readings $\delta$ and $\lambda$ is thus given by $L(\rho, \hat{\rho}(\hat{\delta}), \sigma_\rho(\lambda))$ for each sensor.
The resulting pdfs are used as soft evidence input for the Gaussian elementary sensor model described in [Yguel et al.(2007)]:

$$P_k([Z = z]) =$$

$$\begin{cases} 
\int_{-\infty}^{1} \mathcal{N}(k - 0.5, \sigma_{\log_b}(k))(u)du, & z \in [0; 1] \\
\int_{[z]}^{[z]+1} \mathcal{N}(k - 0.5, \sigma_{\log_b}(k))(u)du, & z \in ]1; N] \\
\int_{N;+\infty} \mathcal{N}(k - 0.5, \sigma_{\log_b}(k))(u)du, & z = "No Detection" 
\end{cases}$$

$$\sigma_{\log_b}(k) = \log_b(b^{(k-0.5)} + \sigma_\rho) - (k - 0.5), \text{ and } k = \lceil \mu_\rho \rceil \text{ is the index of the only occupied cell in the line-of-sight in log-space, which represents the coordinate interval } [k - 1; k].$$
Vision Sensor Model
Bayesian Program of the Direct Vision Sensor Model

Relevant variables:
- $C \in \mathcal{C} \subset \mathcal{Y}$: BVM cell identifier,
- $Z \in \{"No Detection"\} \cup \mathcal{Z}_{\text{VisDepth}}$: sensor depth measurement along line-of-sight ($\theta, \phi$);
- $O_C \in \mathcal{O}$: binary value describing the occupancy of cell $C$;
- $G_C \in \mathcal{G}_C \equiv \mathcal{O}^{N-1}$: state of all cells in the line-of-sight except for $C$.

Decomposition:
$$P(Z \mid O_C \cdot G_C) = P(C)P(O_C \mid C) \cdot \frac{P(G_C \mid O_C \cdot C)P(Z \mid G_C \cdot O_C \cdot C)}{\sum_{G_C} P(G_C \mid O_C \cdot C)}.$$ 

Gives $P(Z \mid O_C \cdot C)$ through $\sum_{G_C}$.

Parametric forms:
- $P(C)$: uniform;
- $P(O_C \mid C)$: uniform or prior estimate;
- $P(G_C \mid O_C \cdot C)$: unknown, apart from dependency on number of occupied cells;
- $P(Z \mid G_C \cdot O_C \cdot C)$: probability of a measurement by sensor, knowing first occupied cell is $[C = k] \equiv \text{elementary sensor model } P_k(Z)$.

Identification:
Calibration for $P_k(Z) \Rightarrow P(Z \mid G_C \cdot O_C \cdot C)$.

Questions:
Direct model: $P(Z \mid o_C \cdot c)$; Inverse model: $P(O_C \mid z \cdot c)$
The “multiple looks” hypothesis has been previously proposed to explain monaural detection and discrimination performance with increasing signal duration [Viemeister and Wakefield(1991)].

**Multiple looks hypothesis:** The auditory system has a short-term memory (i.e. buffer) of “looks” at the signal, which can be accessed and processed selectively.
The first step in auditory sensing on IMPEP is **figure-ground segregation**, which is done while binaural audio is recorded in the “multiple looks” memory buffer.

Also known as foreground-background segmentation, this process segments “interesting” (i.e. loud), “foreground” auditory snippets from the audio stream in each “look” (that will be labelled as “background” noise otherwise) for further processing.
Auditory perception begins with monaural cochlear and auditory periphery processing, producing an *auditory image model* (AIM) [Patterson et al. (1995)].

The spectral analysis stage converts the sound wave of each monaural component of the binaural signal through a model of *basilar membrane motion* (BMM) into $M$ separate frequency channels.

The neural encoding stage stabilises the BMM in level and sharpens features like vowel formants, to produce a simulation of the *neural activity pattern* (NAP) yielded by sound in the auditory nerve.
Binaural perception is accomplished by using the “multiple looks” hypothesis together with interaural coherence (IC), as described in [Faller and Merimaa(2004)].

Interaural time differences (ITDs) are estimated, which are then used to compute interaural level differences (ILDs) and IC values, yielding ITD-ILD-IC triplets for each frequency channel $f_c$ given by \{$\tau(n), \Delta L(n), c_{12}(n)$\}.
IC values $c_{12} > c_0$ are used for selecting triplets most similar to free-field cues for individual sound-sources (i.e. auditory object segmentation).

ITDs are stable across frequencies for a sound source at a given azimuth $\rightarrow$ ITDs are summed over all frequencies (similar to what is believed to occur in the ICx brain area), resulting in the summary cross-correlogram.
Summary crosscorrelogram $\equiv$ auditory saliency map $\rightarrow i$ largest peaks were most probably effected by the most important sound-sources represented in the auditory image.

For each sound-source $i$, auditory localisation cue vectors $A_i = [\tau^i, \Delta L^i(f^1_c) \cdots \Delta L^i(f^M_c)]$ are constructed.

ITDs $= f(\theta)$, while ILDs $= f(\rho, \theta, \phi)$ for close-range ($< 2$ m) sources.
Audition sensor readings are by nature head-centred and in spherical coordinates.

The solution proposed for vision sensors assumes the notion of visual occlusion — however, for audition the notion of occlusion does not apply.

While reducing the richness of sensory information when comparing to vision sensors, on the other hand this makes it much easier to formulate the direct model for the audition sensor.
The (direct) auditory sensor model can be easily described by the decomposition:

\[ P(\tau \Delta L(f_c^1) \cdots \Delta L(f_c^M) O_C) = \]

\[
\text{Uniform} \quad P(\tau \Delta L(f_c^1) \cdots \Delta L(f_c^M)) \quad \text{From calibration} \quad \prod_{k=1}^{M} P(\Delta L(f_c^k)|\tau O_C) \]

\[ = P(Z|O_C) \]
A set $M_C$ of measurement vectors $[\tau, \Delta L(f^1_c) \cdots \Delta L(f^M_c)]$ is collected per cell $C$.

The full set of collected measurement vectors for all cells in auditory sensor space $\mathcal{Y}'$ is expressed as $M = \bigcup M_C$.

The set of measurements collected for all cells other than $C$ is denoted as $M_{\bar{C}} = M \setminus M_C$.

$\mu_{\tau, \Delta L(f^k_c)}(M_C)$ and $\sigma_{\tau, \Delta L(f^k_c)}(M_C)$ are used to characterise the family of distributions for $[O_C = 1]$.

$\mu_{\tau, \Delta L(f^k_c)}(M_{\bar{C}})$ and $\sigma_{\tau, \Delta L(f^k_c)}(M_{\bar{C}})$ are used to characterise the family of distributions for $[O_C = 0]$.
A broadband audio stimulus is presented through a loudspeaker.

The loudspeaker is positioned in the geometric centre of each cell $C \in C'$ so as to sample space according to the auditory sensor space $\mathcal{Y}'$.

Different angles are obtained by rotating the active head; different distances are achieved by hand.

Example above: 2 calibration sessions of 3 h.
To process the inertial data, we follow the Bayesian model proposed by [Laurens and Droulez(2006)] of the human vestibular system, adapted here to the use of inertial sensors. The aim is to provide an estimate for the current angular position and angular velocity of the perceptual system, thus mimicking human vestibular perception.
Outline

1 Introduction
   - Motivations
   - Expected Scientific Contributions
   - Expected Practical Contributions

2 Bayesian Volumetric Maps for Multimodal Integration
   - Occupancy Grids and Spatial Configuration
   - Bayesian Sensor Models
   - Bayesian Multimodal Sensor Fusion

3 Ongoing and Future Work
   - What Next?
   - Entropy-Based Active Exploration

4 Bibliography
There is evidence that humans perform sensor fusion mainly through [Ernst and Bülthoff(2004)]:

- **Combination** → interactions between sensory signals that are not redundant, maximising information coming from different cues.
- **Integration** → interactions between sensory signals that are redundant, minimising variance in sensory estimates to increase their reliability.

Combination $\xrightarrow{Promotion}$ Integration [Ernst and Bülthoff(2004)].
Bayesian Volumetric Maps — Implementing Promotion

- Vestibular sensor models yield measurements of angular velocity and position.
- These are used to promote visual and auditory observations for BVM updates when the active head turns;
- Instead of rotating the BVM, the most effective solution is to perform the equivalent index shift: $C = (\log_b \rho_{\text{max}}, \theta_{\text{max}} - \theta_{\text{inertial}}, \phi_{\text{max}} - \phi_{\text{inertial}}) \in C \subset Y$.
- This process obviously relies on the assumption that inertial precision on angular measurements is greater than the chosen resolution parameters for the BVM.
Bayesian Volumetric Maps
Bayesian Program for Cell State Estimation

Relevant variables:
- \( C \in \mathcal{V} \): indexes a cell on the BVM;
- \( A_C \): identifier of the antecedents of cell \( C \) (stored as with \( C \));
- \( Z_1, \ldots, Z_S \in \{ \text{"No Detection"} \} \cup \mathcal{Z} \): independent measurements taken by \( S \) sensors;
- \( O_C, O_C^{-1} \in \mathcal{O} \equiv \{0, 1\} \): binary values describing the occupancy of cell \( C \), for current and preceding instants, respectively;
- \( V_C \): velocity of cell \( C \), discretised into \( n \) possible cases \( \in \mathcal{V} \equiv \{v_1, \ldots, v_n\} \).

Decomposition:
\[
P(C A_C O_C O_C^{-1} V_C Z_1 \cdots Z_S) =
\]
\[
P(A_C) P(V_C|A_C) P(C|V_C A_C) P(O_C^{-1}|A_C) P(O_C|O_C^{-1}) \prod_{i=1}^{S} P(Z_i|V_C O_C C)
\]

Parametric forms:
- \( P(A_C) \): uniform;
- \( P(V_C|A_C) \): histogram;
- \( P(C|V_C A_C) \): Dirac (constant velocity assumption);
- \( P(O_C^{-1}|A_C) \): probability of preceding state of occupancy given set of antecedents;
- \( P(O_C|O_C^{-1}) \): defined through transition matrix \( T = \begin{bmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{bmatrix} \);
- \( P(Z_i|V_C O_C C) \): direct measurement model for each sensor \( i \), given by respective sub-BP.

Identification: None.

Question: \( P(O_C V_C|z_1 \cdots z_S c) \)
The inference process of multimodal fusion using the BVM leads to the Bayesian filtering formulation as used in the BOF grids by [Tay et al. (2007)]:

- Prediction propagates cell occupancy probabilities for each velocity and cell in the grid — $P(O_{C V_{C}} | C)$.
- During estimation, $P(O_{C V_{C}} | C)$ is updated using the observations yielded by the set of $S$ vision and audition sensors $\prod_{i=1}^{S} P(Z_{i} | V_{C} O_{C} C)$ to obtain the final state estimate $P(O_{C V_{C}} | Z_{1} \cdots Z_{S} C)$.
- This estimate is used for the prediction step in the next iteration.
Outline

1 Introduction
   - Motivations
   - Expected Scientific Contributions
   - Expected Practical Contributions

2 Bayesian Volumetric Maps for Multimodal Integration
   - Occupancy Grids and Spatial Configuration
   - Bayesian Sensor Models
   - Bayesian Multimodal Sensor Fusion

3 Ongoing and Future Work
   - What Next?
   - Entropy-Based Active Exploration

4 Bibliography
Outline

1. Introduction
   - Motivations
   - Expected Scientific Contributions
   - Expected Practical Contributions

2. Bayesian Volumetric Maps for Multimodal Integration
   - Occupancy Grids and Spatial Configuration
   - Bayesian Sensor Models
   - Bayesian Multimodal Sensor Fusion

3. Ongoing and Future Work
   - What Next?
   - Entropy-Based Active Exploration

4. Bibliography
Regarding the BVM itself:

- Study of the development of a velocity vision sensor model using stereoscopic optical flow population codes (with C. Braillon, INRIA).
- Study of the use of a tracking solution such as used on [Tay et al.(2007)] for independent motion segmentation.

BVM integration in a full-fledged active perception system:

- BVM for entropy-based active exploration (as in [Rocha et al.(2005)] — ongoing work);
- Saliency map-based “saccadic movement” generation;
- Behaviour/action selection for active perception (like [Koike(2005)]?).
Regarding the BVM itself:
- Study of the development of a velocity vision sensor model using stereoscopic optical flow population codes (with C. Braillon, INRIA).
- Study of the use of a tracking solution such as used on [Tay et al.(2007)] for independent motion segmentation.

BVM integration in a full-fledged active perception system:
- BVM for entropy-based active exploration (as in [Rocha et al.(2005)] — ongoing work);
- Saliency map-based “saccadic movement” generation;
- Behaviour/action selection for active perception (like [Koike(2005)]?).
Regarding the BVM itself:
- Study of the development of a velocity vision sensor model using stereoscopic optical flow population codes (with C. Braillon, INRIA).
- Study of the use of a tracking solution such as used on [Tay et al.(2007)] for independent motion segmentation.

BVM integration in a full-fledged active perception system:

- BVM for entropy-based active exploration (as in [Rocha et al.(2005)] — ongoing work);
- Saliency map-based “saccadic movement” generation;
- Behaviour/action selection for active perception (like [Koike(2005)]?).
Regarding the BVM itself:

- Study of the development of a velocity vision sensor model using stereoscopic optical flow population codes (with C. Braillon, INRIA).
- Study of the use of a tracking solution such as used on [Tay et al. (2007)] for independent motion segmentation.

BVM integration in a full-fledged active perception system:

- BVM for entropy-based active exploration (as in [Rocha et al. (2005)] — ongoing work);
- Saliency map-based “saccadic movement” generation;
- Behaviour/action selection for active perception (like [Koike (2005)]?).
Regarding the BVM itself:
- Study of the development of a velocity vision sensor model using stereoscopic optical flow population codes (with C. Braillon, INRIA).
- Study of the use of a tracking solution such as used on [Tay et al.(2007)] for independent motion segmentation.

BVM integration in a full-fledged active perception system:
- BVM for entropy-based active exploration (as in [Rocha et al.(2005)] — ongoing work);
- Saliency map-based “saccadic movement” generation;
- Behaviour/action selection for active perception (like [Koike(2005)]?).
Regarding the BVM itself:
- Study of the development of a velocity vision sensor model using stereoscopic optical flow population codes (with C. Braillon, INRIA).
- Study of the use of a tracking solution such as used on [Tay et al.(2007)] for independent motion segmentation.

BVM integration in a full-fledged active perception system:
- BVM for entropy-based active exploration (as in [Rocha et al.(2005)] — ongoing work);
- Saliency map-based “saccadic movement” generation;
- Behaviour/action selection for active perception (like [Koike(2005)]).
Outline

1. Introduction
   - Motivations
   - Expected Scientific Contributions
   - Expected Practical Contributions

2. Bayesian Volumetric Maps for Multimodal Integration
   - Occupancy Grids and Spatial Configuration
   - Bayesian Sensor Models
   - Bayesian Multimodal Sensor Fusion

3. Ongoing and Future Work
   - What Next?
   - Entropy-Based Active Exploration

4. Bibliography
Problem: where to direct perception?

- The BVM allows the use of the concept of information entropy to promote an exploratory behaviour of areas of the environment corresponding to cells on the volumetric map associated to high uncertainty.

- The joint entropy of the state of cell \([C = c]\) on the BVM is given by

\[
H(c) \equiv H(V_c, O_c) = -\sum_{o_c \in O, v_c \in V} P(v_c, o_c | z_c) \log P(v_c, o_c | z_c).
\]

- High entropy means high uncertainty → low information.
Bayesian Models for Active Multimodal Perception

Ongoing and Future Work

Entropy-Based Active Exploration

**BVMs for Active Perception**

Using Uncertainty and Entropy for Active Exploration (II)

---

**Active Exploration Algorithm**

1. Find the last non-occluded, close-to-empty cell for the whole span of directions \((\theta_{\text{max}}, \phi_{\text{max}})\) in the BVM — these are the so-called *frontier cells* as defined on [Rocha et al. (2005)]; the set of all frontier cells is denoted as \(\mathcal{F} \subset \mathcal{C}\).

2. Compute the joint entropy gradient for each of the frontier cells and select \(c_s = \arg \max_{c \in \mathcal{F}} \left[ (1 - P([O_C = 1]| [C = c])) \| \nabla H(c) \| \right] \) as the best candidate cell to direct gaze to. If more than one global maximum, choose cell corresponding to the direction closest to current heading.

3. Compute gaze direction as being \((\theta_C, \phi_C)\), where \(\theta_C\) and \(\phi_C\) are the angles that bisect cell \([C = c_s]\).
Outline

1 Introduction
   - Motivations
   - Expected Scientific Contributions
   - Expected Practical Contributions

2 Bayesian Volumetric Maps for Multimodal Integration
   - Occupancy Grids and Spatial Configuration
   - Bayesian Sensor Models
   - Bayesian Multimodal Sensor Fusion

3 Ongoing and Future Work
   - What Next?
   - Entropy-Based Active Exploration

4 Bibliography
References I

ROCHA, R.

KOIKE, C.

FERREIRA, J. F. AND CASTELO-BRANCO, M.
References II

Rochar, R., Dias, J., and Carvalho, A.

Pouget, A., Dayan, P., and Zemel, R.
Information processing with population codes.
Review.

Landy, M. S., Maloney, L., Johnston, E. B., and Young, M.
Measurement and modeling of depth cue combination: in defense of weak fusion.

Ernst, M. O. and Bülthoff, H. H.
Merging the senses into a robust percept.
Spence, C. and Squire, S.

Knudsen, E. and Brainard, M.

Wagner, H.

Ingham, N. J., Hart, H. C., and McAlpine, D.


Viemeister, N. F. and Wakefield, G. H.
Temporal integration and multiple looks.

Faller, C. and Merimaa, J.
Source localization in complex listening situations: Selection of binaural cues based on interaural coherence.

Laurens, J. and Droulez, J.
Bayesian processing of vestibular information.
*Biological Cybernetics*, December 2006.
(Published online: 5th December 2006).
