Hierarchical Log-Spherical Inference Grid – An Unconventional Approach to Robotic Perception and Action

João Filipe Ferreira[1] and Jorge Dias[1,2]

Abstract— **In this text, an unconventional paradigm for robotic multisensory perception and action will be presented in the form of a generalisation of a framework devised in previous work by the authors, the Bayesian Volumetric Map (BVM). The BVM, a log-spherical inference grid providing an egocentric and probabilistic representation of spatial information, was designed to bridge multisensory perception and actuation by means of a unified framework. The underlying paradigm represents a fundamental departure from traditional outlooks on robotic perception and spatial reasoning, in that it embodies a non-Cartesian and egocentric approach as opposed to the conventional Cartesian and allocentric perspective.**

I. Introduction

Sensor data are derived from physical entities, each of which placed in precise locations in the observer's surroundings. The most important and immediate associations that humans and other animals make when trying to make sense of the incoming sensory data are precisely spatial associations, since these generally have imminent significance.

On the other hand, humans and robots alike have to deal with the unavoidable reality of sensory uncertainty. Consider the following scenario – a stationary or moving observer is presented with a dynamic 3D scene containing several stationary and moving entities, probably generating some kind of sound: how does this observer solve the *symbol grounding problem* [1], while taking into account the ambiguities and conflicts inherent to the perceptual process?

Natural evolution is currently believed to have imprinted an integrated solution to these problems in the human brain:

1) Several authors (for example [2], [3]) have presented evidence that the human perceptual system is supported by two processing streams: a fast lane, that seems to be associated to phylogenetically older brain sites, such as those composing the visual dorsal stream, which are committed to producing a quick egocentric description of the environment in terms of *where* objects are placed to support immediate action, postponing recognition for later processing stages; and a slow lane, including sites such as those composing the visual ventral stream, which implement object recognition within an allocentric frame of reference*. Additionally, direction and distance in egocentric representations in fast processing streams are believed to be separately specified by the brain [4]. Considering distance in particular, just-discriminable depth thresholds have been usually plotted as a function of the log of distance from the observer, with analogy to contrast sensitivity functions based on Weber's fraction [5].
2) Numerous other authors have presented research supporting that the human brain, in particular in what concerns perception, implements what seems to be a probabilistic approach from the neural to the functional level – see [6] for a detailed discussion on this matter.

In this paper, an unconventional bioinspired paradigm for robotic multisensory perception and action based on the aforementioned premises will be presented in the form of a generalisation of a framework devised in previous work [7], [8], [9], the Bayesian Volumetric Map (BVM). The BVM, a log-spherical inference grid providing an egocentric and probabilistic representation of spatial information, was designed to bridge multisensory perception and actuation by means of a unified framework.

II. The Bayesian Volumetric Map

A. Dense representation model of space using an egocentric, log-spherical tesselation

The tesselation of the BVM is primarily defined by its range of azimuth and elevation angles, and by its maximum reach in distance ρ_{Max}, which in turn determines its log-distance base through $b = a^{\frac{\log_a(\rho_{\text{Max}} - \rho_{\text{Min}})}{N}}$, $\forall a \in \mathbb{R}$, where ρ_{Min} defines the *egocentric gap*, for a given number of partitions N, chosen according to application requirements. The BVM space is therefore effectively defined by

$$\mathcal{Y} \equiv \left]\log_b \rho_{\text{Min}}; \log_b \rho_{\text{Max}}\right] \times \left]\theta_{\text{Min}}; \theta_{\text{Max}}\right] \times \left]\phi_{\text{Min}}; \phi_{\text{Max}}\right] \qquad (1)$$

In practice, the BVM is parametrised so as to cover the full angular range for azimuth and elevation. This configuration virtually delimits a *horopter* for sensor fusion around the egocentric origin $\{\mathcal{E}\}$.

Each BVM cell is defined by two limiting log-distances, $\log_b \rho_{\text{min}}$ and $\log_b \rho_{\text{max}}$, two limiting azimuth angles, θ_{min} and θ_{max}, and two limiting elevation angles, ϕ_{min} and ϕ_{max}, through:

$$\mathcal{Y} \supset \mathcal{C} \equiv \left]\log_b \rho_{\text{min}}; \log_b \rho_{\text{max}}\right] \times \left]\theta_{\text{min}}; \theta_{\text{max}}\right] \times \left]\phi_{\text{min}}; \phi_{\text{max}}\right] \qquad (2)$$

[1] J. F. Ferreira (jfilipe@isr.uc.pt) and J. Dias are with the Institute of Systems and Robotics and the Faculty of Science and Technology, University of Coimbra, Coimbra, Portugal

[2] J. Dias is also with the Robotics Institute from the Khalifa University of Science, Technology and Research (KUSTAR), Abu Dhabi, UAE

**egocentric* – related to point-of-view centred on the observer, usually directly or indirectly sensor-referred; *allocentric* – related to point-of-view centred on any concrete or abstract entity other than observer.

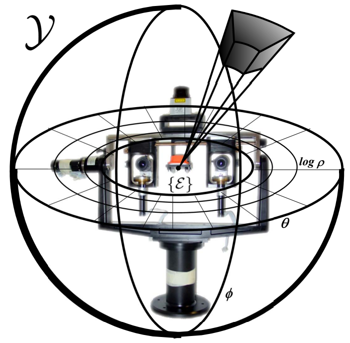

Fig. 1. The Bayesian Volumetric Map (BVM) referred to the egocentric coordinate frame of a robotic active perception system [7]. Hardware and motors were mounted within the scope of the Perception on Purpose (EC project number FP6-IST-2004-027268) project, and sensors were installed within the scope of the Bayesian Approach to Cognitive Systems project (EC project number FP6-IST-027140).

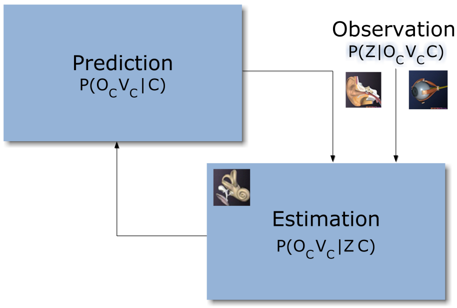

Fig. 2. Bayesian Occupancy Filter (BOF) diagram for the estimation of the current cell state [7]. In this diagram, Z denotes a sensor measurement (a single measurement is represented for simpler reading, with no loss of generality), and O_C and V_C represent the occupancy and local motion states of cell C (which, besides indexing a cell on the BVM, is also used as a random variable within the model), respectively.

where constant values for log-distance base b, and angular ranges $\Delta\theta = \theta_{\text{max}} - \theta_{\text{min}}$ and $\Delta\phi = \phi_{\text{max}} - \phi_{\text{min}}$, chosen according to application resolution requirements, ensure BVM grid regularity. Finally, each BVM cell is formally *indexed* by the coordinates of its *far corner*, defined as $C = (\log_b \rho_{\text{max}}, \theta_{\text{max}}, \phi_{\text{max}})$.

The BVM spatial representation model is shown in Fig. 1 within the egocentric context of the IMPEP (Integrated Multimodal Perception Experimental Platform), an active robotic head fitted with multisensory capabilities, namely a stereovision and binaural setup, used to experimentally validate the framework in previous publications.

B. Inference grid model – the Bayesian Occupancy Filter

Perhaps the most successful probabilistic approach for metric mapping has been the *occupancy map or grid*, first introduced by Moravec and Elfes in their seminal work [10]. The state of each cell C in the grid maps is originally defined by the authors as given by its *occupancy*, represented by the binary random variable O_C – there is either an object partially or completely present within the spatial boundaries of each cell, in which case the cell is said to be occupied, a state denoted as $[O_C = 1]$, or the cell is empty, a state denoted as $[O_C = 0]$. If, instead of just occupancy, more properties are added to the state, thereby forming a random vector, the occupancy grid generalises to the notion of *inference grid*.

Tay et al. [11] devised a version of an important extension to the occupancy grid approach by applying a Bayesian filter to capture the dynamics of object motion within the environment being represented – the Bayesian Occupancy Filter (BOF). These authors relaxed the common restriction that objects remain static through time by introducing an extra variable, V_C, to the inference framework, encoding the probability of (local) motion of an object occupying a neighbouring cell in instant $t-1$ to a particular reference cell C in the current time instant t, thus propagating the occupancy state from the former to the latter during that time interval. V_C therefore denotes the dynamics of the occupancy of cell C, assuming a *constant velocity model* associated to a quantified degree of plausibility – it is a vector signalling local motion to this cell from its antecedents, discretised into $N+1$ possible cases for velocities $\in \mathcal{V} \equiv \{v_0, \cdots, v_N\}$, with v_0 signalling that the most probable antecedent is C itself, i.e. no motion between two consecutive time instants. The authors managed to do this without compromising the feasibility of exact inference inherent to the original occupancy grid concept. The BOF model, supported by the log-spherical configuration introduced earlier, represents the basis of the BVM framework.

A diagram of the BOF in the context of the BVM is presented on Fig. 2, including a general overview of the main variables used by the framework – for more details on the formal derivation of this model, please refer to [8].

III. Building Up the Framework Hierarchy

A. Developing perceptual models

Developing perceptual models for the BOF-BVM framework involves defining observation models of the form

$$P(Z_i \mid O_C\, V_C\, C) \equiv \sum_{G_C \in \mathcal{G}_C} P(Z_i \mid G_C\, O_C\, V_C\, C), \qquad (3)$$

where $G_C \in \mathcal{G}_C \equiv \mathcal{O}^{N-1}$ represents the state of all cells in a subset comprising N cells of $\mathcal{Y}$ (including C), excepting the state the cell of C itself. Random variable Z_i denotes an M-dimensional sensor reading from an arbitrary sensor, with a measurable space or support defined as $\{\text{“No Detection”}\} \cup \mathcal{Z}$. Observation models must be defined under the assumption that all Z_i within a set of K ($i \in 1 \ldots K$) measurements taken at time t are *conditionally independent*. These observation models can then be used as probabilistic “subroutines” or “plug-ins” for the BOF-BVM framework, and their respective estimates and uncertainty combined explicitly and adequately.

Most models one might expect being used with the BOF-BVM framework would be occupancy observation models, which would specialise to $P(Z_i \mid O_C\, C)$. There are two particular forms of observation models within this subset we suggest have particular importance and will cover most (if not all) types of sensors to be used in a robotic perception scenario: the *beam model* and the *egocentric-related probabilistic transfer function*. The former, conceptually introduced in [12], is generally used for projection-based sensors or arrays of sensors, such as sonars, laser range-finders or cameras, for which the notions of transparency/opacity and occlusion are paramount in providing information on "emptiness" (i.e. $[O_C = 0]$) – an example of such an observation model can be found for a stereovision setup in [7]. The latter is an adaptation of the head-related transfer function (HRTF) concept commonly used to model auditory responses to audible stimuli – a maximum likelihood estimation method is applied to determine the free parameters of the observation model for each cell by occupying it with an object capable of generating a signal that can be read by the sensor – an example of such a model can also be found in [7] for a binaural setup. Additionally, observation models on motion can also be used with the BOF-BVM framework, which would then specialise to $P(Z_i \mid V_C\, C)$ – these would be, in fact, *local independent motion estimation models*.

The BOF-BVM is supposed to be used as the kernel of the hierarchical model supported by the log-spherical representation. However, the modeller is encouraged to extend the framework and use it to perform inference through random variables representing other perceptual properties besides the occupancy state and its propagation.

B. Going beyond detection and the "fast lane" – proto-objects

As with the human perceptual system, although postponed at first, data association in the form of object recognition may subsequently be performed supported by algorithms clustering neighbouring cells, such as presented in [11]. These, in turn, provide a rough object detection and discrimination process, allowing the formation of volatile perceptual units called proto-objects – see [13], [14].

Proto-object representations can then be used to infer percepts that relate to object identification/recognition (as in the human visual ventral pathway, relating to the *"what"* problem).

C. Developing models for actuation

Models for actuation may be hierarchically be superimposed on the perception framework, and goal-related properties may be assigned to random variables in order to prioritise action – an example of this approach using the BOF-BVM framework can be found in [8].

In the general case, however, a transformation between the common BVM representation and effector space will be necessary. This can be naturally performed using probabilistic approaches – refer to [15] for an example.

IV. Discussion

In this text, an unconventional paradigm for robotic multisensory perception and action was presented in the form of a generalisation of the Bayesian Volumetric Map framework. Its egocentric approach deals with the geometry of fusion in a natural fashion, in particular for the common case of energy projection-based sensors (either in a passive or active sense); on the other hand, spatial reasoning for actuation benefits from a direct way to infer useful information such as distance-to- or time-to-impact. Additionally, the proposed spatial configuration confers two important advantages: (i) a robustness advantage, since the spherical coordinate system avoids unnecessary ray tracing and therefore prevents undesired Moiré effects (in other words, aliasing); and (ii) an efficiency advantage, since log-partitioning of distance allows for the use of a lower tesselation resolution without compromising any detail of nearer objects. Therefore, this paper's contribution is to make this framework accessible to robotic perception system designers who would wish to make use of its advantages. Proof-of-concept of this paradigm has been presented in [7], its usefulness in bridging perception and action demonstrated in [8], and its potential for implementation using real-time exact inference through parallel programming shown in [9].

References

[1] S. Harnad, "The symbol grounding problem," *Physica D*, vol. 42, pp. 335–346, 1990.

[2] D. Kahneman, *Thinking, Fast and Slow*. Farrar, Straus and Giroux, October 2011.

[3] J. Bullier, "Integrated model of visual processing," *Brain Research Reviews*, vol. 36, pp. 96–107, 2001, review.

[4] J. McIntyre, F. Stratta, and F. Lacquaniti, "Short-Term Memory for Reaching to Visual Targets: Psychophysical Evidence for Body-Centered Reference Frames," *Journal of Neuroscience*, vol. 18, no. 20, pp. 8423–8435, October 15 1998.

[5] J. E. Cutting and P. M. Vishton, "Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth," in *Handbook of perception and cognition*, W. Epstein and S. Rogers, Eds. Academic Press, 1995, vol. 5; Perception of space and motion.

[6] K. Doya, S. Ishii, A. Pouget, and R. P. N. Rao, Eds., *Bayesian Brain – Probabilistic Approaches to Neural Coding*. MIT Press, January 2007.

[7] J. F. Ferreira, J. Lobo, P. Bessière, M. Castelo-Branco, and J. Dias, "A Bayesian Framework for Active Artificial Perception," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. PP, no. 99, pp. 1 –13, September 2012, Early Access Article.

[8] J. F. Ferreira, M. Castelo-Branco, and J. Dias, "A hierarchical Bayesian framework for multimodal active perception," *Adaptive Behavior*, vol. 20, no. 3, pp. 172–190, June 2012.

[9] J. F. Ferreira, J. Lobo, and J. Dias, "Bayesian real-time perception algorithms on GPU — Real-time implementation of Bayesian models for multimodal perception using CUDA," *Journal of Real-Time Image Processing*, vol. 6, no. 3, pp. 171–186, September 2011.

[10] H. Moravec and A. Elfes, "High resolution maps from wide angle sonar," in *IEEE International Conference on Robotics and Automation*, 1985.

[11] C. Tay, K. Mekhnacha, C. Chen, M. Yguel, and C. Laugier, "An efficient formulation of the Bayesian occupation filter for target tracking in dynamic environments," *International Journal of Autonomous Vehicles*, vol. 6, no. 1–2, pp. 155–171, 2008.

[12] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. MIT press Cambridge, MA, 2005, vol. 1.

[13] R. A. Rensink, "The Dynamic Representation of Scenes," *Visual Cognition*, vol. 2000, no. 7, pp. 17–42, 2003.

[14] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Networks*, vol. 19, pp. 1395–1407, 2006.

[15] E. Gilet, J. Diard, and P. Bessière, "Bayesian Action-Perception Computational Model: Interaction of Production and Recognition of Cursive Letters," *PLoS ONE*, vol. 6, no. 6, p. e20387, June 2011.