Proceedings of the 14th IASTED International Conference
Robotics and Applications (RA 2009)
November 2 - 4, 2009 Cambridge, MA, USA

MULTIMODAL ACTIVE EXPLORATION USING A BAYESIAN APPROACH

João Filipe Ferreira
Institute of Systems and Robotics,
FCT-University of Coimbra
Coimbra, Portugal
email: jfilipe@isr.uc.pt

José Augusto Prado
Institute of Systems and Robotics,
FCT-University of Coimbra
Coimbra, Portugal
email: jaugusto@isr.uc.pt

Jorge Lobo
Institute of Systems and Robotics,
FCT-University of Coimbra
Coimbra, Portugal
email: jlobo@isr.uc.pt

Jorge Dias
Institute of Systems and Robotics,
FCT-University of Coimbra
Coimbra, Portugal
email: jorge@isr.uc.pt

ABSTRACT
In this text, we use a Bayesian framework for active multimodal perception of 3D structure and motion — which, while not strictly neuromimetic, finds its roots in the role of the dorsal perceptual pathway of the human brain — to implement a strategy of active exploration based on entropy. The computational models described in this text support a robotic implementation of multimodal active perception to be used in real-world applications, such as human-machine interaction or mobile robot navigation.

KEY WORDS
Active Perception, Multisensory, Bayesian, Bioinspired, Exploration, Realtime.

1 Introduction

Perception has as of recently been regarded as a computational process of unconscious, probabilistic inference. Aided by developments in statistics and artificial intelligence, researchers have begun to apply the concepts of probability theory rigorously to problems in biological perception and action [1]. One striking observation from this work is the myriad ways in which human observers behave as near-optimal Bayesian observers. Several authors even argue that the brain codes even complex patterns of sensory uncertainty in its internal representations and computations — see for example [1, 2].

Active perception has been an object of study in robotics for decades now, specially active vision, which was first introduced by [3] and later explored by [4]. Many perceptual tasks tend to be simpler if the observer is active and controls its sensors [4]. Active perception is thus an intelligent data acquisition process driven by the measured, partially interpreted scene parameters and their errors from the scene. The active approach has the important advantage of making most ill-posed perception tasks tractable [4].

Active multisensory perception using spatial maps has, however, been the object of study since only much recently — an example of this research would be the work of [5] in visuoauditory-driven gaze shift generation.

The availability of a probabilistic framework to implement spatial mapping of the environment allows the use of the concept of *information entropy*, which can be used to promote an exploratory behaviour of areas of the environment corresponding to cells on the volumetric map associated to high uncertainty.

Our work will contribute in providing a rather complete framework for active multimodal perception — introducing a novel approach which, while not strictly neuromimetic, finds its roots in the role of the dorsal perceptual pathway of the human brain and its egocentric trait — which will support the construction of a simultaneously flexible and powerful robotic implementation to be used in real-world applications, such as human-machine interaction or mobile robot navigation. A realtime implementation of all the processes of the framework has been developed, capitalising on the potential for parallel computing of most of its algorithms.

Its main strength lies on the fact that it offers a framework which is naturally fitting for acting upon the environment and also for the integration of readings from multiple sensors, since both processes inherently depend on egocentric reference frames.

2 The Integrated Multimodal Perception Experimental Platform

2.1 Platform description

To support our research work, an artificial multimodal perception system (IMPEP — Integrated Multimodal Perception Experimental Platform) has been constructed at the

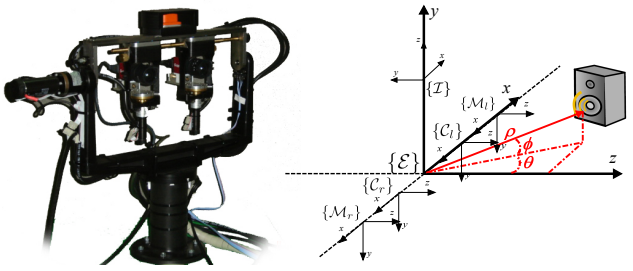


Figure 1: View of the current version of the Integrated Multimodal Perception Experimental Platform (IMPEP), on the left. The active perception head mounting hardware and motors were designed by the Perception on Purpose (POP - EC project number FP6-IST-2004-027268) team of the ISR/FCT-UC, and the sensor systems mounted at the Mobile Robotics Laboratory of the same institute, within the scope of the Bayesian Approach to Cognitive Systems project (BACS - EC project number FP6-IST-027140). On the right, the IMPEP perceptual geometry is shown: $\{\mathcal{E}\}$ is the main reference frame for the IMPEP robotic head, representing the egocentric coordinate system;$\{\mathcal{C}_{l,r}\}$ are the stereovision (respectively left and right) camera referentials; $\{\mathcal{M}_{l,r}\}$ are the binaural system (respectively left and right) microphone referentials; and finally $\{\mathcal{I}\}$ is the inertial measuring unit's coordinate system.

ISR/FCT-UC consisting of a stereovision, binaural and inertial measuring unit (IMU) setup mounted on a motorised head, with gaze control capabilities for image stabilisation and perceptual attention purposes — see Fig. 1. This solution will enable the implementation of an active perception system with great potential in applications as diverse as social robots or even robotic navigation.

The stereovision system is implemented using a pair of Guppy IEEE 1394 digital cameras from Allied Vision Technologies (`http://www.alliedvisiontec.com`), the binaural setup using two AKG Acoustics C417 linear microphones (`http://www.akg.com/`) and an FA-66 Firewire Audio Capture interface from Edirol (`http://www.edirol.com/`), and the miniature inertial sensor, Xsens MTi (`http://www.xsens.com/`), provides digital output of 3D acceleration, 3D rate of turn (rate gyro) and 3D earth-magnetic field data for the Inertial Measurement Unit (IMU).

2.2 Sensory processing

As mentioned before, several authors argue that current evidence strongly suggests that the brain codes even complex patterns of sensory uncertainty in its internal representations and computations. One such representation is believed to be neural population coding (e.g., average firing rate) — see for example [1, 2].

For stereovision sensing, our motivations suggest a tentative data structure analogous to neuronal population ac-

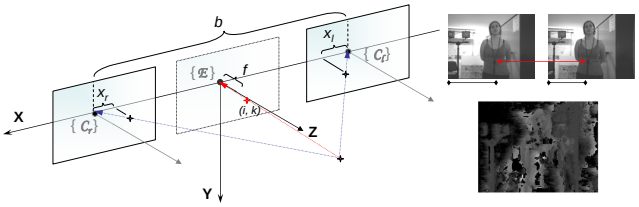


Figure 2: Cyclopean geometry for stereovision. The use of cyclopean geometry (pictured on the left for an assumed frontoparallel configuration) allows direct use of the egocentric reference frame for depth maps taken from the disparity maps yielded by the stereovision system (of which an example is shown on the right).

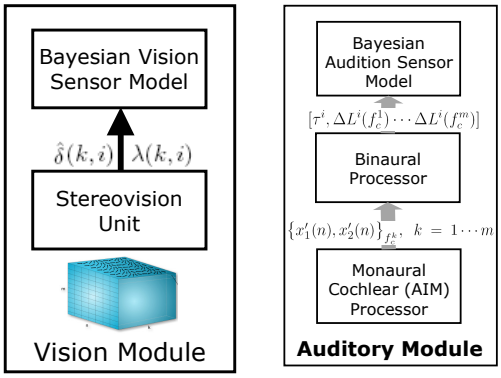


Figure 3: The IMPEP Bayesian sensor systems.

tivity patterns to represent uncertainty in the form of probability distributions [2]. Thus, a spatially organised 2D grid may have each cell (corresponding to a virtual photoreceptor in the cyclopean view — see Fig. 2) associated to a "population code" extending to additional dimensions, yielding a set of probability values encoding a N-dimensional probability distribution function or pdf. This information is consequently used as soft evidence by a Bayesian sensor model previously presented in [6, 7] (Fig. 3).

The Bayesian binaural system, which was fully described in [8, 9], is composed of three distinct and consecutive processors (Fig. 3): the monaural cochlear unit, which processes the pair of monaural signals $\{x_1, x_2\}$ coming from the binaural audio transducer system by simulating the human cochlea, so as to achieve a tonotopic representation (i.e. a frequency band decomposition) of the left and right audio streams; the binaural unit, which correlates these signals and consequently estimates the binaural cues and segments each sound-source; and, finally, the Bayesian 3D sound-source localisation unit, which applies a Bayesian sensor model so as to perform localisation of sound-sources in 3D space.

Finally, a Bayesian inertial module was devised, as

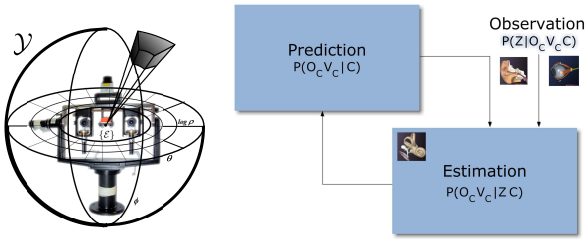


Figure 4: The Bayesian Volumetric Map.

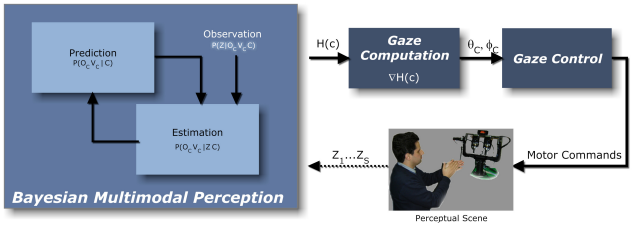


Figure 5: Active multimodal perception using entropy-based exploration. Gaze control module is described on Fig. 6.

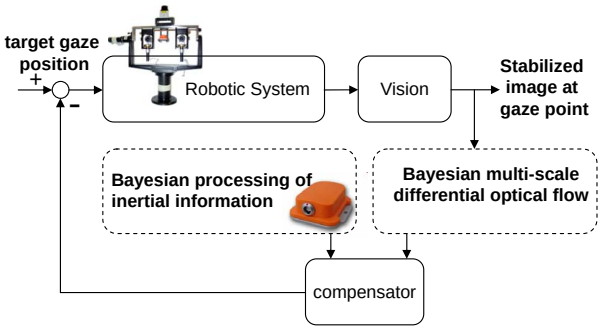


Figure 6: System block diagram for the implementation of gaze control and image stabilisation — for more information see [10].

fully described in [10], to simulate human's vestibular sensing.

3 Active Exploration Using Bayesian Models for Multimodal Perception of 3D Structure and Motion

3.1 Multimodal Sensor Fusion Using Log-Spherical Bayesian Volumetric Maps

A spatial representation framework for multimodal perception of 3D structure and motion, the Bayesian Volumetric Map (BVM), was presented in [6, 7, 8, 9], characterised by an *egocentric*, *log-spherical spatial configuration* to which the Bayesian Occupancy Filter (BOF), as formalised by Tay *et al.* [11], has been adapted. It effectively provides a computational means of storing and updating a perceptual spatial map in a short-term working memory data-structure, representing both 3D structure and motion without the need for any object segmentation process (see Fig. 4). In this model, cells of a partitioning grid on the BVM log-spherical space $\mathcal{Y}$ are indexed through $C \in \mathcal{C} \subset \mathcal{Y}$, where $\mathcal{C}$ represents the subset of positions in $\mathcal{Y}$ corresponding to the "far corners" of each cell C, O_C is a binary variable representing the state of occupancy of cell C (as in the commonly used occupancy grids — see [12]), and V_C is a finite vector of random variables that represent the state of all local motion possibilities used by the prediction step of the Bayesian filter associated to the BVM for cell C, assuming a constant velocity hypothesis, as depicted on Fig. 4.

The BVM is extendible in such a way that other properties characterised by additional random variables and corresponding probabilities might be represented, other than the already implemented occupancy and local motion properties, by augmenting the hierarchy of operators through Bayesian subprogramming [13].

3.2 Active Exploration Using the Bayesian Volumetric Map

Information in the BVM is stored as the *probability of each cell being in a certain state*, defined as $P(V_c\, O_c | z\, c)$. The state of each cell thus belongs to the state-space $\mathcal{O} \times \mathcal{V}$. The *joint entropy* of the random variables V_C and O_C that compose the state of each BVM cell $[C = c]$ is defined as follows:

$$H(c) \equiv H(V_c, O_c) = -\sum_{\substack{o_c \in \mathcal{O} \\ v_c \in \mathcal{V}}} P(v_c\, o_c | z\, c) \log P(v_c\, o_c | z\, c) \tag{1}$$

The joint entropy value $H(c)$ is a sample of a continuous joint entropy field $H : \mathcal{Y} \to \mathbb{R}$, taken at log-spherical positions $[C = c] \in \mathcal{C} \subset \mathcal{Y}$. Let $c_{\alpha-}$ denote the contiguous cell to C along the negative direction of the generic log-spherical axis α, and consider the edge of cells to be of unit length in log-spherical space, without any loss of generality. A reasonable first order approximation to the joint entropy gradient at $[C = c]$ would be

$$\overrightarrow{\nabla} H(c) \approx [H(c) - H(c_{\rho-}), H(c) - H(c_{\theta-}), H(c) - H(c_{\phi-})]^T \tag{2}$$

with magnitude $\|\overrightarrow{\nabla} H(c)\|$.

A great advantage of the BVM over Cartesian implementations of occupancy maps such as the one presented on [14] is the fact that the log-spherical configuration avoids the need for time-consuming ray-casting techniques when computing a gaze direction for active exploration, since the log-spherical space is already defined based on directions

(θ, ϕ). Hence, the active exploration algorithm is simplified to the completion of the following steps:

1. Find the last non-occluded, close-to-empty (i.e. $P([O_C = 1]|[C = c]) < .5$) cell for the whole span of directions $(\theta_{\max}, \phi_{\max})$ in the BVM — these are considered to be the so-called *frontier cells* as defined on [14]; the set of all frontier cells will be denoted here as $\mathcal{F} \subset \mathcal{C}$.

2. Compute the joint entropy gradient for each of the frontier cells and select $c_s = \arg\max_{c \in \mathcal{F}} \left[(1 - P([O_C = 1]|[C = c])) \|\overrightarrow{\nabla} H(c)\| \right]$ as the best candidate cell to direct gaze to. In case there is more than one global maximum, choose the cell corresponding to the direction closest to the current heading (i.e. $(\theta_{\max}, \phi_{\max}) = (0, 0)$, so as to ensure minimum gaze shift rotation effort.

3. Compute gaze direction as being (θ_C, ϕ_C), where θ_C and ϕ_C are the angles that bisect cell $[C = c_s]$ (i.e. which pass through the geometric centre of cell c_s in Cartesian space).

The full BVM entropy-based active perception system is described by the block diagram presented in Fig. 5 (see also Fig. 6).

4 System Implementation and Calibration

4.1 System implementation

The BVM-IMPEP framework, of which an implementation diagram is presented on Fig. 7, was implemented as follows:

- **Vision sensor system**: With the OpenCV toolbox and David Gallup's implementation of a basic binocular stereo algorithm on GPU using CUDA (please refer to `http://www.cs.unc.edu/~gallup/stereo-demo` for more information). The algorithm reportedly runs at 40 Hz on 640×480 images at 50 disparities, computing left and right disparity maps and performing left-right consistency validation (which in our adaptation is used to produce the stereovision confidence maps).

- **Binaural sensor system**: Using an adaptation of the realtime software kindly made available by the Speech and Hearing Group at the University of Shefield [15] to implement binaural cue analysis as described in [8, 9].

- **Bayesian Volumetric Map, Bayesian sensor models and active exploration**: using our proprietary, parallel processing, GPU implementation developed with NVIDIA's general purpose parallel computing architecture CUDA `http://www.nvidia.com/object/cuda_home.html`).

4.2 System calibration

4.2.1 Vision system calibration

Accurate camera calibration can greatly simplify solutions to many important vision problems such as the stereo vision problem, the three-dimensional visual tracking problem, the mobile-robot visual guidance problem, the 3D reconstruction problem, the 3D visual information registration problem, etc. For example, it is well known that a well-calibrated stereo vision system would not only dramatically reduce the complexity of the stereo correspondence problem but also significantly reduce the 3D estimation error [16].

Camera calibration can be performed using a standard stereovision calibration software to estimate left and right camera *intrisic parameters* (i.e. focal length and distortion parameters for undistorting images for processing) and *extrinsic parameters* (i.e. transformation between camera local coordinate systems — in the case of an ideal frontoparallel setup, the estimation of baseline b) that allow the application of the reprojection equation:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & f \\ 0 & 0 & \frac{1}{b} & 0 \end{bmatrix} \begin{bmatrix} u_l - \frac{\hat{\delta}}{2} \\ v_l \\ \hat{\delta} \\ 1 \end{bmatrix} = \begin{bmatrix} WX \\ WY \\ WZ \\ W \end{bmatrix} \quad (3)$$

where u_l is the horizontal coordinate and v_l is the vertical coordinate of a point on the left camera, and $\hat{\delta}$ is the disparity estimate for that point, all of which in pixels, f and b are the estimated focal length and baseline, respectively, both of which in metric distance, and X, Y and Z are 3D point coordinates respective to the egocentric/cyclopean referential system $\{\mathcal{E}\}$.

Using reprojection error measurements given by the calibration procedure, parameter σ_{min} as defined in [6] is taken as being equal to the maximum error exhibited by the stereovision system.

Finally, to determine $(\theta_{i,k}, \phi_{i,k})$ and $\hat{\rho}_{i,k}(\hat{\delta})$ (i.e. to perform the cartesian-to-spherical transformation) for each projection line (i, k) to use with the vision sensor model given in [6], the following relations are built from equation (3),

$$\begin{cases} \theta_{i,k} & = 2 \arctan\left(\frac{X}{2f}\right) \\ \phi_{i,k} & = 2 \arctan\left(\frac{Y}{2f}\right) \\ \hat{\rho}_{i,k}(\hat{\delta}) & = \sqrt{X^2(\hat{\delta}) + Y^2(\hat{\delta}) + Z^2(\hat{\delta})} \end{cases} \quad (4)$$

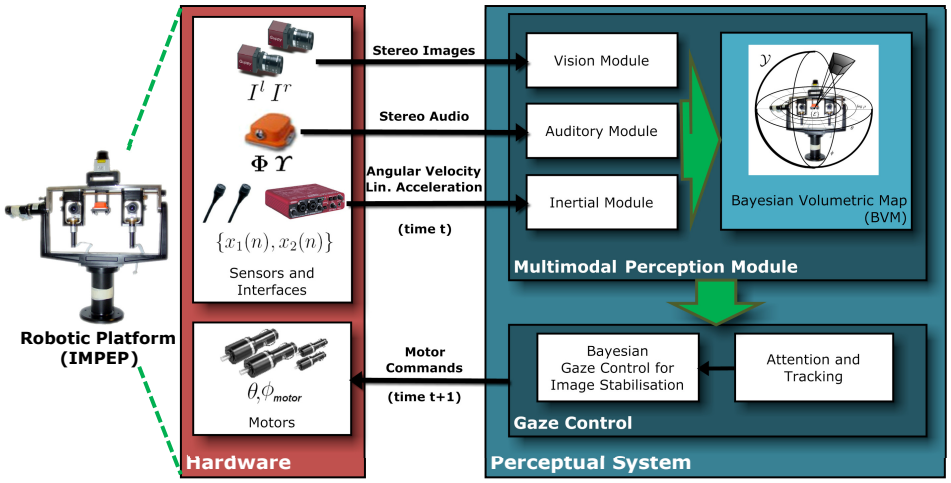


Figure 7: Implementation diagram for the BVM-IMPEP multimodal perception framework.

Given $\theta_{i,k}$ and $\phi_{i,k}$, it becomes possible at any moment to compute depth from a given disparity estimate by substitution of the two first expressions onto the last in Equation 4, yielding

$$\hat{\rho}_{i,k}(\hat{\delta}) = f\sqrt{4\left(\tan^2\frac{\theta_{i,k}}{2} + \tan^2\frac{\phi_{i,k}}{2}\right) + \left(\frac{b}{\hat{\delta}}\right)^2} \quad (5)$$

4.2.2 Binaural system calibration

As described in [9], calibration of the binaural system involves the characterisation of the families of normal distributions $P(\tau|S_C\, O_C\, \theta_{\max})$ and $P(\Delta L(f_c^k)|\tau\, S_C\, O_C\, C) \approx P(\Delta L(f_c^k)|S_C\, O_C\, C)$ through descriptive statistical learning of their central tendency and statistical variability, where S_C is an intermediate binary variable signalling the occupation of a cell with a sound-source (i.e. the occupancy of a cell does not reflect if the object that occupies it is a sound-source) that cancels out through marginalisation during inference, τ (ITD — interaural time difference) and ΔL (ILD — interaural level difference) are the binaural cues as defined in [8, 9], and f_c^k is the central frequency of each band k of the tonotopic representation. This is done in an equivalent manner as with commonly used head-related transfer function (HRTF) calibration processes (see, for example, [17]) and is described in the following paragraphs.

A set M_c of n-dimensional measurement vectors such as defined in [9] is collected per cell $c \in \mathcal{C}$. The full set of collected measurement vectors for all cells in auditory sensor space $\mathcal{Y}$ is expressed as $M = \bigcup M_c$. Denoting $M_{\bar{c}} = M \setminus M_c$ as the set of measurements for all cells other than c, the statistical characterisation process of each family of distributions is effected for each cell c through

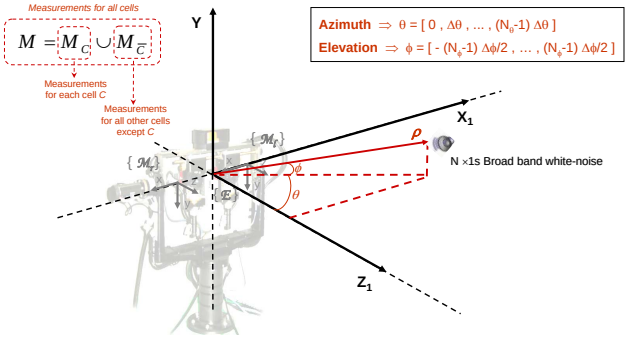


Figure 8: Experimental setup for the binaural system calibration procedure.

$$P(\tau|[S_c = 1]\, O_c\, \theta_{\max}) \equiv \mathcal{N}(\tau, \mu_\tau(M_c), \sigma_\tau(M_c)) \quad (6a)$$
$$P(\tau|[S_c = 0]\, O_c\, \theta_{\max}) \equiv \mathcal{N}(\tau, \mu_\tau(M_{\bar{c}}), \sigma_\tau(M_{\bar{c}})) \quad (6b)$$
$$P(\Delta L(f_c^k)|[S_c = 1]\, O_c\, c) \equiv \mathcal{N}(\Delta L(f_c^k), \mu_{\Delta L(f_c^k)}(M_c), \sigma_{\Delta L(f_c^k)}(M_c)) \quad (6c)$$
$$P(\Delta L(f_c^k)|[S_c = 0]\, O_c\, c) \equiv \mathcal{N}(\Delta L(f_c^k), \mu_{\Delta L(f_c^k)}(M_{\bar{c}}), \sigma_{\Delta L(f_c^k)}(M_{\bar{c}})) \quad (6d)$$

Auditory calibration is performed by presenting a broadband audio stimulus through a loudspeaker positioned in well-known spatial coordinates corresponding to the geometric centre of each cell $c \in \mathcal{C}$ so as to sample space according to the auditory sensor space $\mathcal{Y}$.

The acquisition method may be simplified by a factor of 4 by taking into account the spatial redundancies of auditory sensing, namely the symmetry enforced by the back-to-front ambiguity and the left-to-right antisymmetry for both

ITDs and ILDs, to reduce calibration space to the front-left quadrant.

A further simplification of the procedure consists in positioning the loudspeaker, for each of the N_d considered distances from the binaural system, precisely in front of the active perception head (i.e. $(\theta, \phi) = (0, 0)$) and to *rotate the active head* so that the whole range of azimuths and elevations of the auditory sensor space is covered. This replaces the several minutes taken to reposition the loudspeaker by hand (now only happening N_d times) by a few seconds of head motions for each cell. The full procedure is depicted in Fig. 8.

4.2.3 Visuoinertial calibration

Visuoinertial calibration can be performed using the InerVis toolbox (`http://www.deec.uc.pt/~jlobo/InerVis_WebIndex/InerVis_Toolbox.html`) [18]. The toolbox estimates the rotation quaternion between the Inertial Measurement Unit and a chosen camera, requiring a set of static observations of a standard checkerboard visual calibration target and of sensed gravity.

5 Results and Conclusion

The realtime implementation of all the processes of the framework was subjected to performance testing for each individual module. Processing times and rates for the sensory systems are as follows:

- **Stereovision unit** 15 Hz.
- **Binaural processing unit** Realtime processing for 44 KHz, 16-bit audio, with 16 frequency channels and 50 ms buffer for cue computation.
- **Inertial processing unit** 10 Hz.

The processing times for each individual module for a BVM space of size $360 \times 90 \times 10$ (azimuth $\times$ elevation $\times$ log-distance), with 500 runs of a BVM filter time-step in the processing of real-world scenarios, are the following:

- **Bayesian vision sensor model** — Approximately 50 ms processing time average for 640×480 images.
- **Bayesian audition sensor model** — Approximately 10 ms processing time average.
- **Bayesian vision and binaural sensor models running in parallel CUDA threads** — Approximately 50 ms processing time average for conditions described above.
- **Bayesian volumetric map filter** —- Approximately 55 ms processing time average.

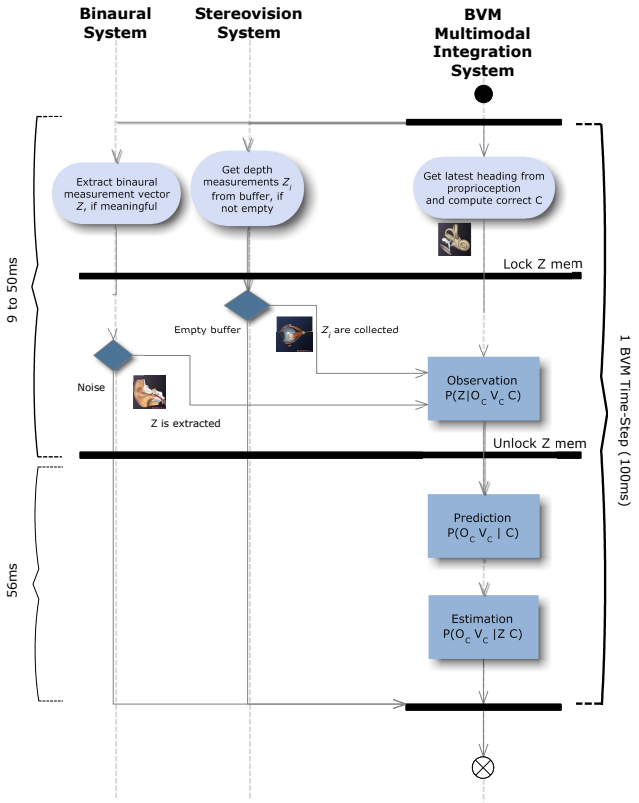


Figure 9: Activity diagram for an inference time-step at time t.

- **Entropy and gaze shift computation** — Approximately 20 ms processing time average.

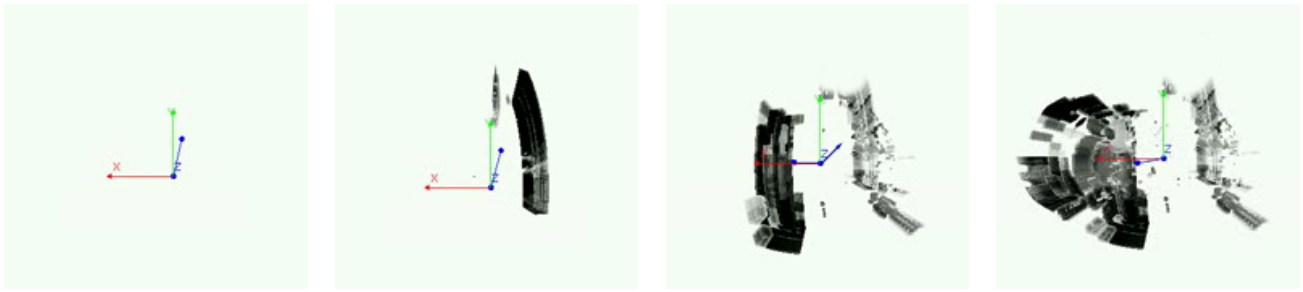
The activity diagram for the BVM Bayesian framework is presented on Fig. 9, depicting an inference step corresponding to time t and respective timeline.

As can be seen, the full active exploration system runs at about 6 Hz. This is ensured by forcing the main BVM thread to pause for each time-step when no visual measurement is available (i.e. during 40 ms for $N = 10, \Delta\phi = 2^{\text{o}}$ — see Fig. 9). This guarantees that BVM time-steps are regularly spaced, which is a very important requirement for correct implementation of prediction/dynamics, and also ensures that processing and memory resources are freed and unlocked regularly.

These performance ratings show that gaze shift reaction times to stimuli in full-fledged, multimodal operation are consistent with realtime standards.

The results of processing a scenario testing different aspects of the full system are presented on Fig. 10. A scene consisting of two male speakers talking to each other in a cluttered lab is observed by the IMPEP active perception system and processed online by the BVM Bayesian filter, using entropy-based active exploration as described earlier, in order to scan the surrounding environment.

(a) Left camera snapshots corresponding to chronologically ordered time-instants. Two male speakers are maintaining a dialog, at -22^{o} and 14^{o} azimuth respectively relatively to the Z axis, which defines the frontal heading respective to the IMPEP "neck". As can be seen on the first frame, both speakers are initially outside the stereovision region-of-interest for processing, being consecutively scanned as a result of active exploration-driven gaze-shifts.



(b) BVM results (frontal views, with Z pointing outward) corresponding to each of the snapshots in (a). The blue arrow depicted in each map denotes the current gaze orientation. Interpretation, from left to right (chronological evolution): 1) initial non-informative map; 2) sound coming from the speaker on the right triggers an estimate for occupancy from the binaural sensor model, and a consecutive exploratory gaze shift; 3) a few frames from the stereovision system trigger further evidence accumulation for occupancy by the vision sensor model at the gaze direction site, fusing readings from both sensory systems — higher spatial resolution from vision carves out the right speaker's sillouette from the first rough estimate from audition —, while sound coming from the speaker on the left triggers an estimate for occupancy from the binaural sensor model, and a consecutive exploratory gaze shift in the speaker's direction; 4) a few frames from the stereovision system trigger further evidence accumulation for occupancy by the vision sensor model at the gaze direction site, fusing readings from both sensory systems — again, higher spatial resolution from vision carves out the left speaker's sillouette from the first rough estimate from audition.

Figure 10: Results for the realtime prototype for multimodal perception of 3D structure and motion using the BVM. A scene consisting of two male speakers talking to each other in a cluttered lab is observed by the IMPEP active perception system and processed online by the BVM Bayesian filter, using the active exploration heuristics described in the main text, in order to scan the surrounding environment. The parameters for the BVM are as follows: $N = 10$, $\rho_{Min} = 1000$ mm and $\rho_{Max} = 2500$ mm, $\theta \in [-180^{o}, 180^{o}]$, with $\Delta\theta = 1^{o}$, and $\phi \in [-90^{o}, 90^{o}]$, with $\Delta\phi = 2^{o}$, corresponding to $10 \times 360 \times 90 = 648,000$ cells, approximately delimiting the so-called "personal space" (the zone immediately surrounding the observer's head, generally within arm's reach and slightly beyond, within 2 m range [19]).

The active exploration algorithm successfully drives the IMPEP-BVM framework to explore areas of the environment mapped with high uncertainty in realtime, with an intelligent heuristic that minimises the effects of local minima by attending to the closest regions of high entropy first.

Further details on ongoing work using these models can be found at `http://paloma.isr.uc.pt/~jfilipe/BayesianMultimodalPerception`.

6 Future Work

Extensions to the BVM operators are currently being implemented, so as to include other perceptual properties for each cell, an example of which would be sensory *saliency*, as presented by [20], an important feature of active perception in animals (also known as "automatic orienting"). Human studies using paradigms devised for the development of Bayesian models of active perception that are extensions of the BVM framework will also be performed soon, in order to train the artificial active perception framework. The framework will then be used in realistic scenarios, such as robotic navigation and human-robot interaction.

Acknowledgement

This publication has been supported by EC-contract number *FP6-IST-027140, Action line: Cognitive Systems*. The contents of this text reflect only the author's views. The

European Community is not liable for any use that may be made of the information contained herein.

References

[1] Knill, D.C., Pouget, A.: The Bayesian brain: the role of uncertainty in neural coding and computation. TRENDS in Neurosciences **27**(12) (December 2004) 712–719

[2] Pouget, A., Dayan, P., Zemel, R.: Information processing with population codes. Nature Reviews Neuroscience **1** (2000) 125–132 Review.

[3] Bajcsy, R.: Active perception vs passive perception. In: Third IEEE Workshop on Computer Vision, Bellair, Michigan (1985) 55–59

[4] Aloimonos, J., Weiss, I., Bandyopadhyay, A.: Active Vision. International Journal of Computer Vision **1** (1987) 333–356

[5] Koene, A., Morén, J., Trifa, V., Cheng, G.: Gaze shift reflex in a humanoid active vision system. In: 5th International Conference on Computer Vision Systems (ICVS 2007), Bielefeld University, Germany, Applied Computer Science Group (2007) ISBN 978-3-00-020933-8.

[6] Ferreira, J.F., Bessière, P., Mekhnacha, K., Lobo, J., Dias, J., Laugier, C.: Bayesian Models for Multimodal Perception of 3D Structure and Motion. In: International Conference on Cognitive Systems (CogSys 2008), University of Karlsruhe, Karlsruhe, Germany (April 2008) 103–108

[7] Ferreira, J.F., Pinho, C., Dias, J.: Bayesian Sensor Model for Egocentric Stereovision. In: 14ª Conferência Portuguesa de Reconhecimento de Padrões Coimbra (RECPAD 2008). (October 31 2008)

[8] Pinho, C., Ferreira, J.F., Bessière, P., Dias, J.: A Bayesian Binaural System for 3D Sound-Source Localisation. In: International Conference on Cognitive Systems (CogSys 2008), University of Karlsruhe, Karlsruhe, Germany (April 2008) 109–114

[9] Ferreira, J.F., Pinho, C., Dias, J.: Implementation and Calibration of a Bayesian Binaural System for 3D Localisation. In: 2008 IEEE International Conference on Robotics and Biomimetics (ROBIO 2008), Bangkok, Tailand (February, 21–26 2009)

[10] Lobo, J., Ferreira, J.F., Dias, J.: Robotic Implementation of Biological Bayesian Models Towards Visuo-inertial Image Stabilization and Gaze Control. In: 2008 IEEE International Conference on Robotics and Biomimetics (ROBIO 2008), Bangkok, Tailand (February, 21–26 2009)

[11] Tay, C., Mekhnacha, K., Chen, C., Yguel, M., Laugier, C.: An efficient formulation of the bayesian occupation filter for target tracking in dynamic environments. *International Journal of Autonomous Vehicles* (2007)

[12] Elfes, A.: Using occupancy grids for mobile robot perception and navigation. IEEE Computer **22**(6) (1989) 46–57

[13] Lebeltel, O.: Programmation Bayésienne des Robots. PhD thesis, Institut National Polytechnique de Grenoble, Grenoble, France (September 1999)

[14] Rocha, R., Dias, J., Carvalho, A.: Cooperative Multi-Robot Systems: a study of Vision-based 3-D Mapping using Information Theory. Robotics and Autonomous Systems **53**(3–4) (December 2005) 282–311

[15] Lu, Y.C., Christensen, H., Cooke, M.: Active binaural distance estimation for dynamic sources. In: Interspeech 2007, Antwerp, Belgium (2007)

[16] Shih, S.W., Hung, Y.P., Lin, W.S.: Calibration of an Active Binocular Head. IEEE Transactions on Systems, Man, and Cybernetics — Part A: Systems and Humans **28**(4) (July 1998) 426–442

[17] Calamia, P.T.: Three-dimensional localization of a close-range acoustic source using binaural cues. Master's thesis, Faculty of the Graduate School of The University of Texas at Austin (1998)

[18] Lobo, J., Dias, J.: Relative Pose Calibration Between Visual and Inertial Sensors. International Journal of Robotics Research, Special Issue 2nd Workshop on Integration of Vision and Inertial Sensors **26**(6) (June 2007) 561–575

[19] Cutting, J.E., Vishton, P.M.: Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In Epstein, W., Rogers, S., eds.: Handbook of perception and cognition. Volume 5; Perception of space and motion. Academic Press (1995)

[20] Niebur, E., Itti, L., Koch, C.: Modeling the "where´´ visual pathway. In Sejnowski, T.J., ed.: 2nd Joint Symposium on Neural Computation, Caltech-UCSD. Volume 5., La Jolla, Institute for Neural Computation (1995) 26–35