BACS

Bayesian Models for Active Multimodal Perception

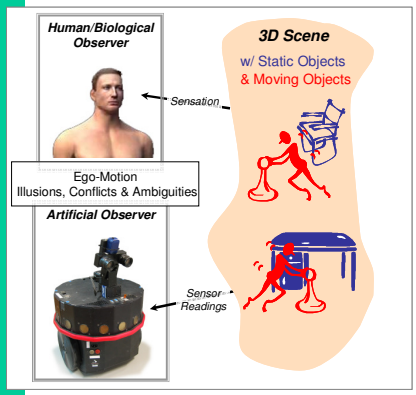
ISR/FCT-UC, University of Coimbra

João Filipe Ferreira, Cátia Pinho, Jorge Dias

BACS

Motivations

- A moving observer is presented with a non-static 3D scene containing several moving objects:



- How does the observer perceive his own motion (***egomotion***), the 3D structure of all objects in the scene and the 3D trajectory and velocity of moving objects (***independent motion***)?

The Bayesian Framework – Uncertainty and Ambiguity

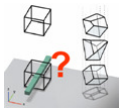
Biological uncertainties:

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

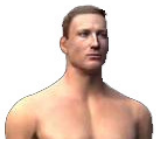
Artificial uncertainties:

- 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 – Justification



- Recently, several hypothesis regarding biological perception have surfaced [1]:

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.

BACS

The Bayesian Framework – Biological and Artificial Perception

BACS

Goals

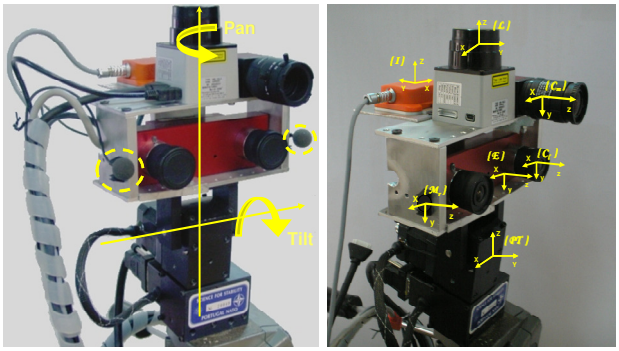
- We mainly expect to contribute in developing computational models which:

- *are based on the perceptual modalities of **vision**, **audition** and **vestibular sensing**;*
- *perform **perceptual fusion within a Bayesian framework**;*
- *do not involve processes such as scene interpretation and classification as in the perceptual brain's dorsal pathway(s);*
- *will serve as a framework for implementing **short-term egocentric spatial memory for active perception**.*

BACS

IMPEP 1 (Integrated Multimodal Perception Experimental Platform)

Platform description



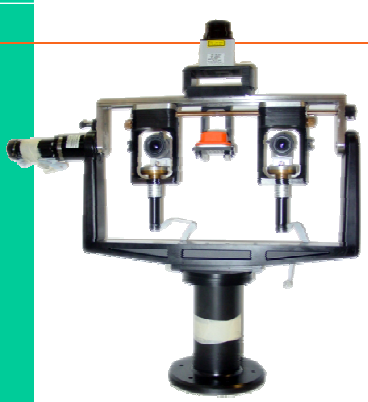
$\{ I \}$	Inertial frame	$\{ C_m \}$	Camera (mono setup)
$\{ L \}$	Laser frame	$\{ C_r \}, \{ C_l \}$	Cameras (stereo setup)
$\{ PT \}$	PanTilt frame	$\{ M_r \}, \{ M_l \}$	Binaural microphones
		$\{ E \}$	Egocentric frame

- IMPEP 1 consists 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**.**

BACS

IMPEP 2 (Integrated Multimodal Perception Experimental Platform)

New Platform



- The IMPEP (v2.0) active perception head adds **vergence** capabilities to the stereovision system besides **improved motor control and conditioning.**

- The design of the new robotic vision head (IMPEP2) is from our colleagues at FCT-UC working on the European project POP - Perception On Purpose - FP6-IST-2004-027268. [http://perception.inrialpes.fr/POP/]

BACS

Egocentric Spatial Maps – The Dorsal Pathway
Egocentric Directional Coding

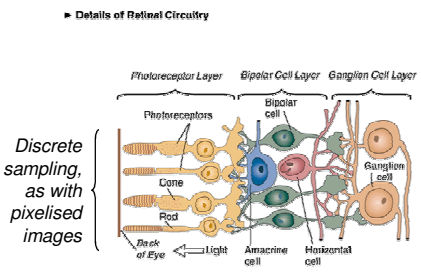
- *Importance of egocentric directional angles for action, namely the issuing of motor commands for manipulation and head turns for gaze control:*

egocentric direction

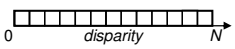
BACS

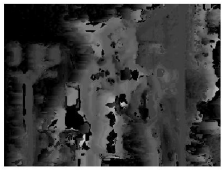
Egocentric Spatial Maps – The Dorsal Pathway
Just Noticeable Difference (JND) and the Logarithmic Scale

- ***Weber's Law of JND:*** *minimum amount by which a stimulus parameter must be changed in order to produce a noticeable variation in sensory experience [2].*
  - *Example*: ***binocular disparities***



Discrete sampling, as with pixelised images



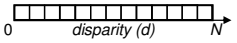


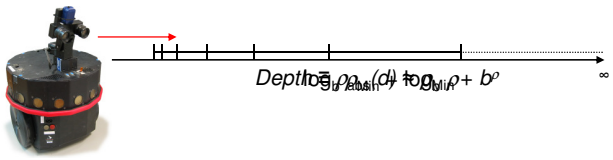
BACS

Egocentric Spatial Maps – The Dorsal Pathway

Just Noticeable Difference (JND) and the Logarithmic Scale

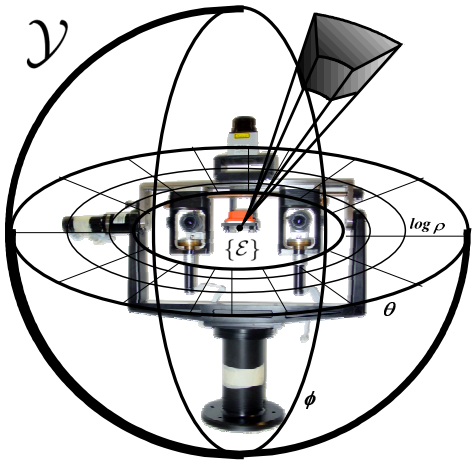
- ***Weber's Law of JND:*** *minimum amount by which a stimulus parameter must be changed in order to produce a noticeable variation in sensory experience [2].*
- *Example*: ***binocular disparities***





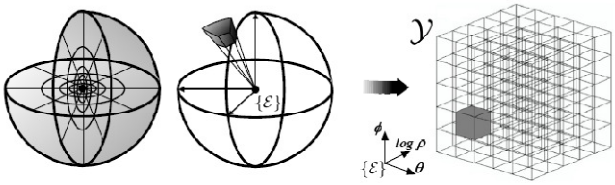
BACS

BVM - Bayesian Volumetric Map



BACS

BVM - Bayesian Volumetric Map



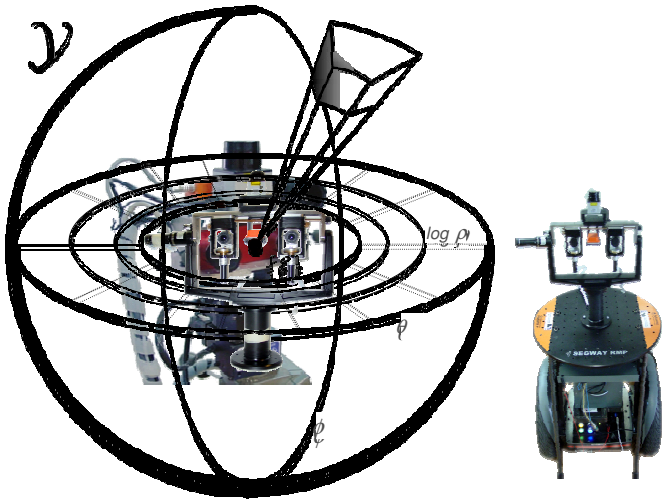
- An egocentric, log-spherical spatial memory map has been devised (in collaboration with INRIA Rhône-Alpes, CNRS-Grenoble and Probayes) as a framework for **multimodal sensor fusion**, named the ***Bayesian Volumetric Map*** (BVM).
- This map stores the independent probabilistic states for each cell C in a volumetric grid with log-spherical configuration, defined by the domain:

$$\mathcal{Y} \equiv \left] \log_b \rho_{Min} ; \log_b \rho_{Max} \right] \times \left] \theta_{Min} ; \theta_{Max} \right] \times \left] \phi_{Min} ; \phi_{Max} \right]$$

BACS

BVM - Bayesian Volumetric Map

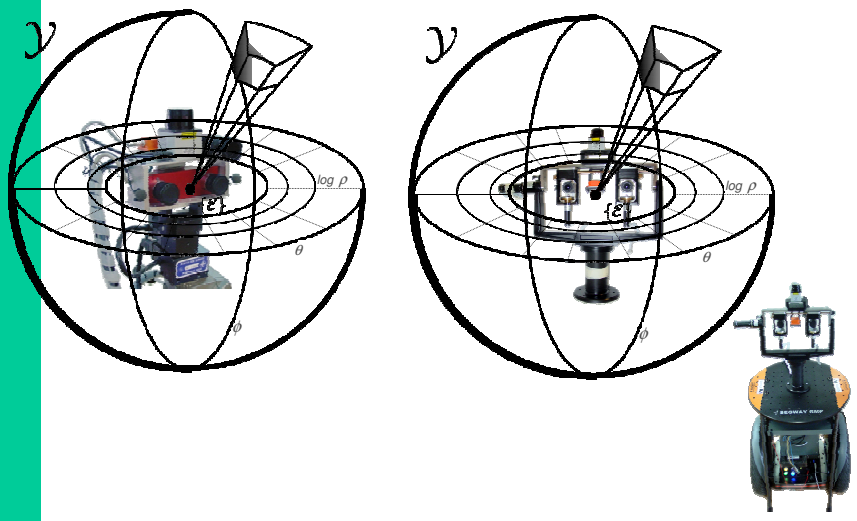
- **IMPEP 2**



- The IMPEP 2 system will be mounted on a SEGWAY RMP mobile platform at the Mobile Robotics Laboratory, ISR/FCT-UC, supported by the ***BACS European Project*** (EC-contract number FP6-IST-027140, Action line: Cognitive Systems).

BACS

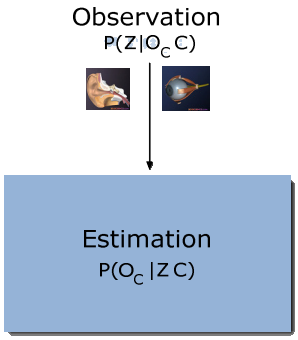
BVM - Bayesian Volumetric Map

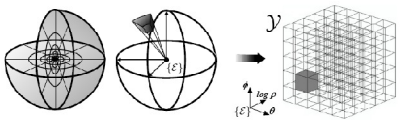


BACS

BVM - Bayesian Volumetric Map

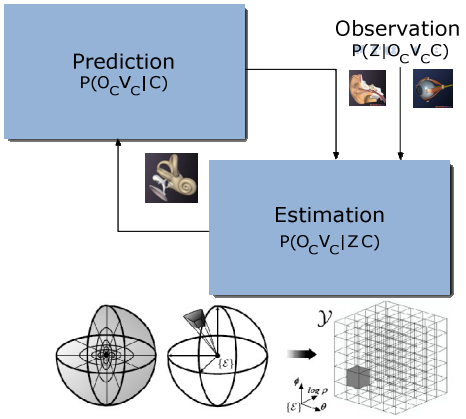
- Visual and auditory observations may be fused and registered using an occupancy map built upon the log-spherical volumetric configuration:



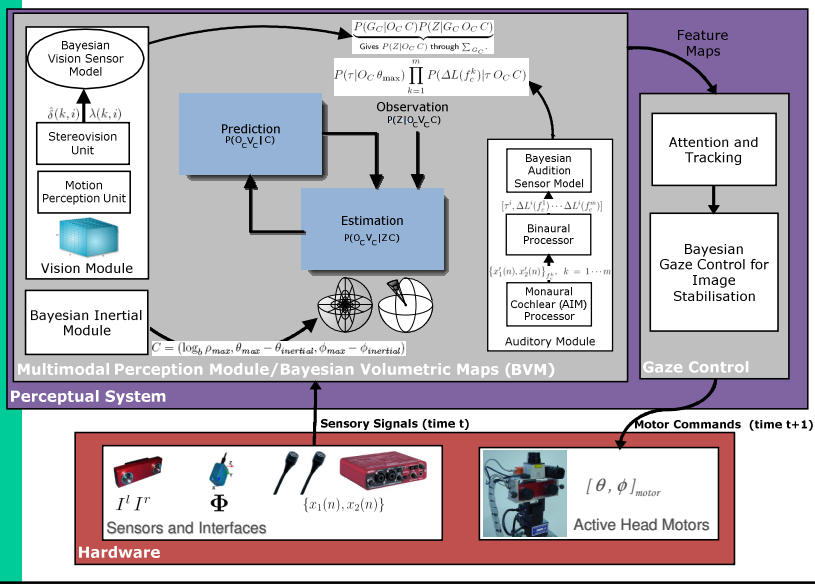


BVM - Bayesian Volumetric Map

- Denoting "cell antecedent" A_C as the cell where the object that occupies cell C was probably positioned in the previous estimation step, and assuming a high probability constant velocity model, a ***prediction*** step can be added.
- A ***Bayesian Occupancy Filter*** can then be used as on [3, 4] (Tay et al., Coué et al.), for both occupancy O_C and local motion V_C.
- ***Promotion*** of observations into integration is achieved by ***estimating egomotion using the inertial sensors*** as with the human vestibular system [5] (Laurens and Droulez).



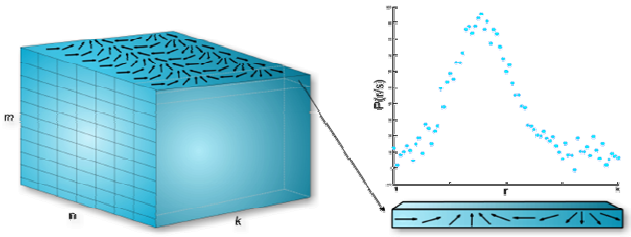
IMPEP/BVM Framework Overview



- Slide with animation showing relations between the cyclopean view, the elementary sensor model and the population code-like data structure.

BACS

Vision Sensor Description

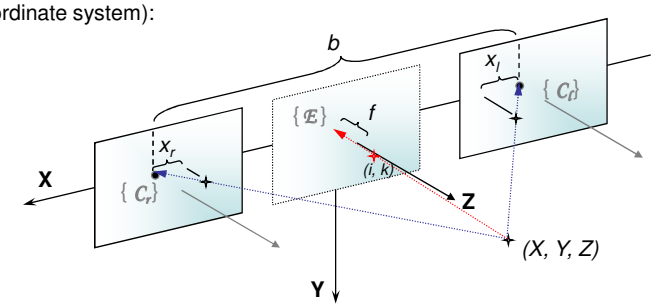


- The IMPEP stereovision system gives a probabilistic depth map referred to the cyclopean view (i.e. the egocentric coordinate system).
- This depth map is stored in a neuronal population code-like data structure, such as the one depicted above.
- Each virtual photoreceptor on the cyclopean view is related to a probability distribution that is used as the so-called ***elementary sensor model*** $P_k(Z)$, which gives the probability of a given depth measurement when only cell [C = k] is known to be occupied by a reflecting surface in the sensor's line-of-sight.

BACS

Egocentric Depth Mapping Using Stereovision

- The IMPEP stereovision system yields a disparity map with associated confidence values λ, which can then be referred to the ***cyclopean view*** (i.e. the egocentric coordinate system):



$(x_l, x_r) \Rightarrow (u_l, u_r)$ *in pixels*

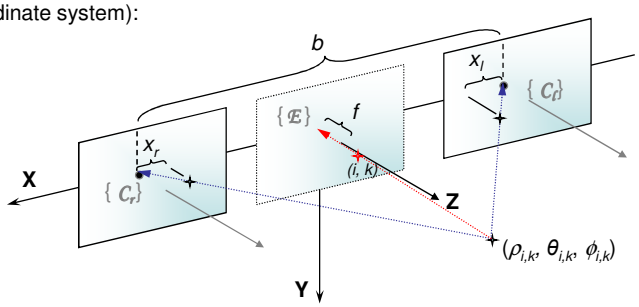
$\delta = u_l - u_r$

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

BACS

Egocentric Depth Mapping Using Stereovision

- The IMPEP stereovision system yields a disparity map with associated confidence values λ, which can then be referred to the ***cyclopean view*** (i.e. the egocentric coordinate system):



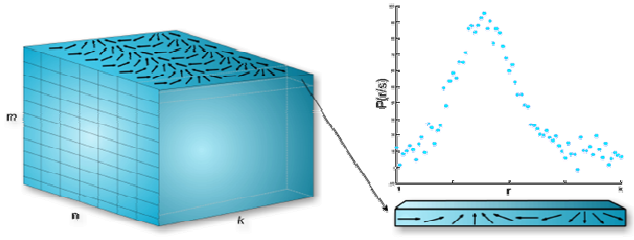
For each pixel (i, k) on the cyclopean image corresponding to $(\theta_{i,k}, \phi_{i,k})$ *in spherical coordinates, a disparity estimate* $\hat{\delta}$ *yields a depth estimate* $\hat{\rho}$

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

BACS

Egocentric Depth Mapping Using Stereovision

- The resulting depth map is stored in a ***neuronal population code-like data structure,*** analogous to the one depicted below:

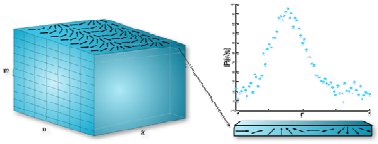


- Each virtual photoreceptor on the cyclopean view is related to a probability distribution that is used as the so-called ***elementary sensor model*** $P_k(Z)$.

BACS

Soft Evidence: Elementary Sensor Model

- The (Gaussian) ***elementary sensor model*** $P_k(Z)$, which yields the probability of a given depth measurement $Z = \log \rho$ when only cell $[C = k]$ in the sensor's line-of-sight is known to be occupied by a reflecting surface, is given by:



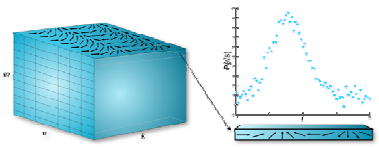
$$P_k([Z = z]) = \begin{cases} \int_{]-\infty;0]} \mathcal{N}(\mu(k-0.5), \sigma(\sigma_\rho))(u)du, & z \in [0;1] \\ \int_{\lceil z \rceil - 1}^{\lceil z \rceil} \mathcal{N}(\mu(k-0.5), \sigma(\sigma_\rho))(u)du, & z \in]1;N] \\ \int_{]N;+\infty]} \mathcal{N}(\mu(k-0.5), \sigma(\sigma_\rho))(u)du, & z = \text{“No Detection”} \end{cases} \qquad \begin{cases} \mu_\rho(k) &= \hat{\rho}(\hat{\delta}) \\ \sigma_\rho(k) &= \frac{1}{\lambda}\sigma_{min} \end{cases}$$

where μ(•) and σ(•) are the operators that perform the required spatial coordinate transformations, and $k = \lceil \mu_\rho \rceil$ is assumed to be the log-space index of the only occupied cell in the line-of-sight, taken from the depth estimate.

BACS

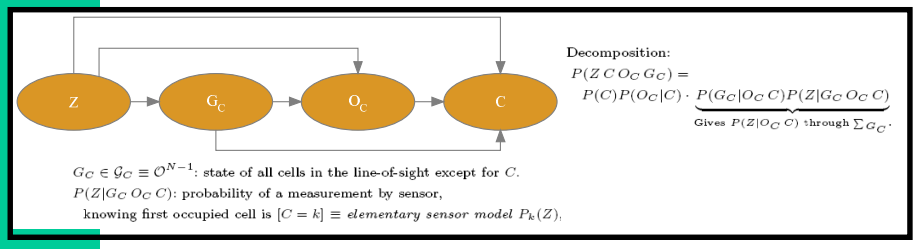
Stereovision Calibration:
Estimating Elementary Sensor Model Parameters

$$\begin{cases} \mu_\rho(k) &= \hat{\rho}(\hat{\delta}) \quad (1) \\ \sigma_\rho(k) &= \frac{1}{\lambda}\sigma_{min} \quad (2) \end{cases}$$



- The elementary sensor model depends of two sets of intrinsic parameters:
  - The intrinsic parameters of both cameras, which are used to compute (1).
  - σ_{min}, the smallest error in depth yielded by the stereovision system, used to compute (2).
- These can be estimated using standard camera calibration techniques.

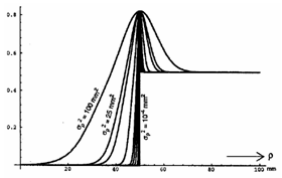
Vision Sensor Model (Adapted from Yguel et al. [6])



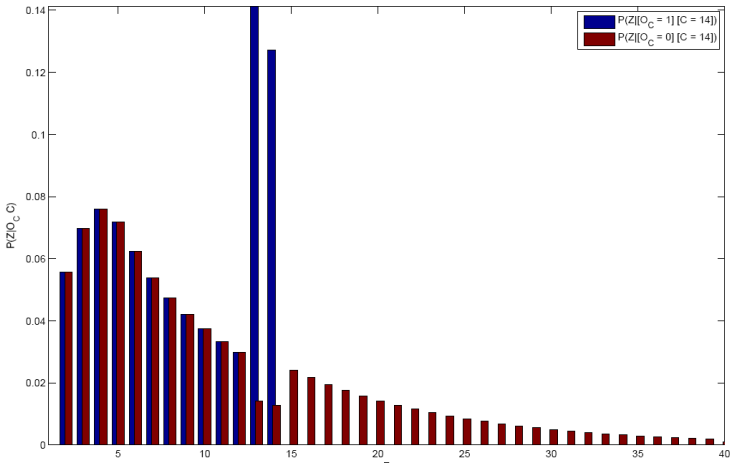
C – *Cell Identifier (log$_b$ ρ_{max}, θ_{max}, ϕ_{max})*

Z – *Measurement taken by the vision sensor (1D range-sensing photoreceptor)*

O_C – *Occupancy of cell C (0 = unnocupied; 1 = occupied by reflective surface)*



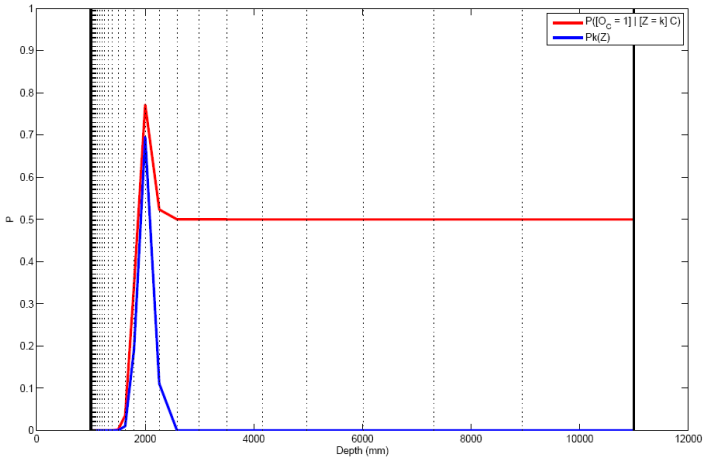
Vision Sensor Model – Simulation Results



Simulation for [C = 14] (total of 40 cells), considering both occupied and unoccupied states. Note that the results reflect the assumption that when [O_C = 1] cells farther from the origin than [C = 14] are occluded, and hence do not yield visual readings.

BACS

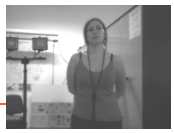
Inference Using Vision Sensor Model – Simulation Results

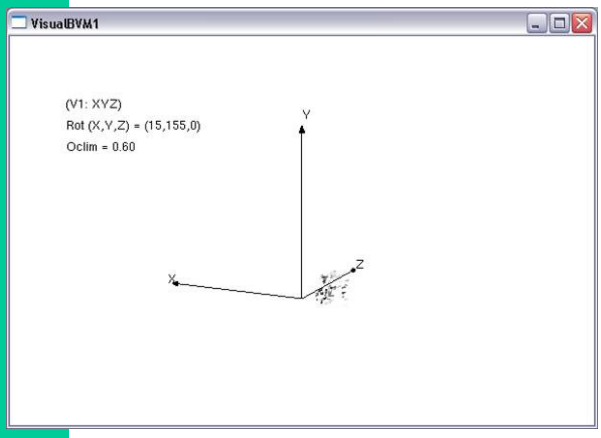


Simulation results with number of cells $N = 40$, and $\rho_{Min} = 1000$mm and $\rho_{Max} = 11000$mm. The full red trace corresponds to the result of inference and the full blue trace corresponds to the Gaussian elementary sensor model. Note the effects of the logarithmic partitioning of depth and of the soft evidence conveyed by the elementary sensor model.

BACS

Inference Using Vision Sensor Model – Results





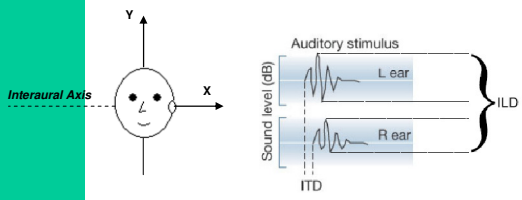
- Occupancy-only (i.e. no prediction/dynamics) results, showing 22-frame movies of BVM *(left)* corresponding to each of the left-camera images *(right),* taken at consecutive instants in time.
- Depicted cells have occupancy probabilities higher than 60%, varying from transparent to opaque with increasing P ([$O_C = 1$] | Z C).

BACS

3D Localisation using Binaural Cues

Binaural Cues

- The ears' spatial disparity and the mass between them → Each ear receives a different version of the arriving sound



*Differences in **time**:*

Interaural Time Difference (ITD)

*Differences in **level**:*

Interaural Level Difference (ILD)

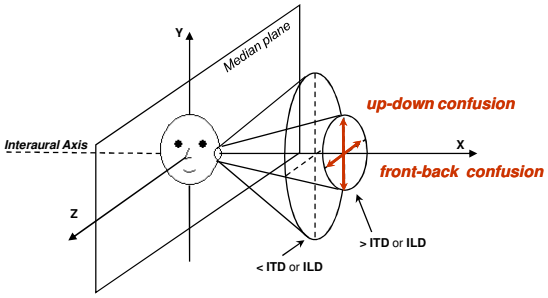
- **ITDs:**
  - vary systematically with the angle of incidence of the sound wave relative to the interaural axis;
  - are virtually independent of frequency, representing the most important localization cue for low frequency signals (<1500Hz in humans).
- **ILDs** vary much more with frequency. Low frequency sounds easily travel around the head and produce negligible ILDs.

BACS

3D Localisation using Binaural Cues

Binaural Isosurfaces – Symmetry Effects

- ***For distances > 2m***, iso-ITD/ILD surfaces form hollow cones of confusion with a specific thickness extending from each ear in a symmetrical configuration relatively to the medial plane:



  - **up-down confusion**
  - **front-back confusion**
- In this case, only azimuth (θ) can be estimated, thus ***monaural cues are needed to fully localise a sound source in 3D***.

BACS

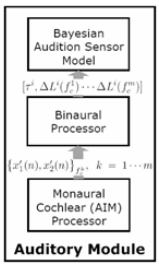
3D Localisation using Binaural Cues

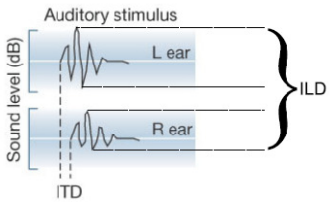
Binaural Isosurfaces – Proximity Effect on ILDs

- However, for sound sources within 1 – 2 meters of the listener, iso-**ILD** surfaces delimit hollow spherical volumes with different shapes for each frequency, symmetrically placed about the medial plane and centred on a point on the interaural axis [8].
- Thus, for sources within 2 meters range, the intersection of the ILD and ITD volumes is a torus-shaped volume [8].
- In this case, ***binaural cues alone can be used to fully localise the source in 3D space*** (i.e. azimuth θ, elevation ϕ and distance ρ).
- If the source is more than 2 meters away, the change in ILD with source position is too gradual to provide spatial information (at least for an acoustically transparent head), and the source can only be localised to a volume around the correct cone of confusion [8].

BACS

Binaural Sensor Description

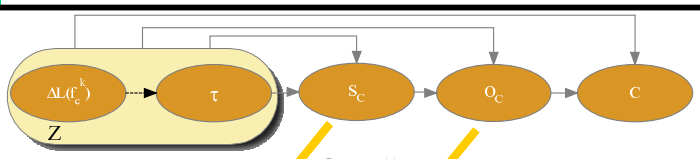




- The binaural system grabs a stereophonic signal and analyses it by applying a model of the monaural processing performed by the middle ear and *cochlea* [9, 10, 11] followed by binaural processing (Faller and Merimaa [7]) of the resulting signals so as to derive binaural cues related to each frequency channel.
- This results on binaural cue measurement vectors of the form $[\tau, \Delta L(f_c^1) \cdots \Delta L(f_c^M)]$, where τ is the frequency-independent ITD, and ΔL are the ILDs.

BACS

Audition/Binaural Sensor Model



S_C: binary value describing the presence of a sound-source in cell C, $[S_C = 1]$ if a sound-source is present at C, $[S_C = 0]$ otherwise;
$P(\Delta L(f_c^k)|\tau\, S_C\, O_C\, C) \approx P(\Delta L(f_c^k)|S_C\, O_C\, C)$
$P(Z|O_C\, C)$: probability of a measurement $[\tau, \Delta L(f_c^1) \cdots \Delta L(f_c^m)]$

Decomposition:
$P(Z\, C\, S_C\, O_C) =$
$P(C)P(O_C|C)\, P(S_C|O_C\, C)P(\tau|S_C\, O_C\, \theta_{\max}) \prod_{k=1}^{m} P(\Delta L(f_c^k)|\tau\, S_C\, O_C\, C)$
Gives $P(Z|O_C\, C)$ through $\sum_{S_C}$

C – *Cell Identifier* $(\log_b \rho_{max},\ \theta_{max},\ \phi_{max})$

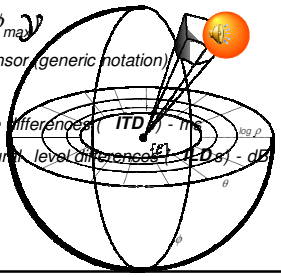
Z – *Measurement taken by the audio sensor (generic notation)*

O_C – *Occupancy of cell C*

τ – *frequency invariant interaural time differences (**ITD**s) - ms*

$\Delta L(f_c^k)$ – *frequency dependent interaural level differences (**ILD**s) - dB*

$P(S_C \mid O_C, C)$	$[O_C=0]$	$[O_C=1]$
$[S_C=0]$	1	0.5
$[S_C=1]$	0	0.5
$\sum$	1	1



Cell occupied by a sound source

$\rightarrow\rightarrow$ $S_C = 1$

… meaning that S_c *is a special case of* O_c

- Slide with animation showing the difference between an occupied cell and cell occupied by a sound-source (i.e., the latter is a special case of the former).

BACS

Audition/Binaural Sensor Model

S_C: binary value describing the presence of a sound-source in cell C, $[S_C = 1]$ if a sound-source is present at C, $[S_C = 0]$ otherwise;
$P(\Delta L(f_c^k)|\tau\, S_C\, O_C\, C) \approx P(\Delta L(f_c^k)|S_C\, O_C\, C)$
$P(Z|O_C\, C)$: probability of a measurement $[\tau, \Delta L(f_c^1) \cdots \Delta L(f_c^m)]$

Decomposition:
$P(Z\, C\, S_C\, O_C) =$
$P(C)P(O_C|C)\, P(S_C|O_C\, C)P(\tau|S_C\, O_C\, \theta_{\max}) \prod_{k=1}^{m} P(\Delta L(f_c^k)|\tau\, S_C\, O_C\, C)$
Gives $P(Z|O_C\, C)$ through $\sum_{S_C}$

C – *Cell Identifier* $(\log_b \rho_{max},\ \theta_{max},\ \phi_{max})$

Z – *Measurement taken by the audio sensor (generic notation)*

O_C – *Occupancy of cell C*

τ – *frequency invariant interaural time differences (**ITD**s) - ms*

$\Delta L(f_c^k)$ – *frequency dependent interaural level differences (**ILD**s) - dB*

BACS

Binaural System Calibration

- The auditory calibration's purpose is to characterise the normal distributions of the ***Direct Auditory Sensor Model (DASM)***, defined as:

$$P(Z \mid O_C C) = \sum_{S_C} P(S_C \mid O_C\ C) P(\tau \mid S_C O_C\ \theta_{\max}) \prod_{k=1}^{m} P\left(\Delta L\left(f_c^k\right) \mid \tau\ S_C O_C\ C\right)$$

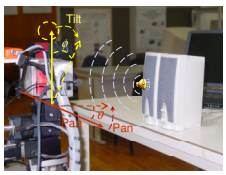
- This will allow the full localisation of sound-sources in three-dimensional space:
  - Azimuth (θ)
  - Elevation (ϕ)
  - Distance (ρ)

- Replace the following static slides with animations concerning the auditory calibration procedure.

BACS

Binaural System Calibration: Experimental Plan

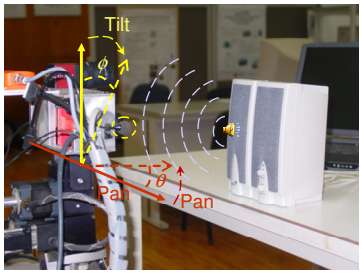
- ***Goal:*** to capture binaural readings using the stereo microphones of the IMPEP Active Perception Head for each cell in the auditory sensor space of a broadband noise sound-source (1s) – similarly to standard Head-Related Transfer Function (HRTF) measurements.



- To cover the complete auditory sensor space, the sound-source must be positioned at each grid cell on the BVM
- ***Methodology:*** sound-source is positioned at a specific distance ***(ρ)*** from the IMPEP head, directly facing the front of the Pan & Tilt Unit (PTU), and the corresponding relative rotation is performed by the PTU;
- The different distances (ρ) between the sound-source and the IMPEP head are obtained manually.
- To avoid redundancies and to simplify the procedure ⇒ ***only one quadrant*** is used capitalising on:
  - *Symmetry from **front-back confusion phenomenon***
  - *Left-Right anti-symmetry (ITD = - ITD and ILD = - ILD)*

BACS

Experimental plan (example)



Resolution (Pan / Tilt):

- 2 degree of azimuth (θ) – performed by the **Pan** motor – **360º / 2º = 180 cells (θ)**
- 10 degrees of elevation (ϕ) – performed by the **Tilt** motor – **180º / 10º = 18 cells (ϕ)**

- Acquisition for $\mathbf{N_d = 2}$ different distances ⇒ d : d_1 ~ 0.32m ; d_2 ~ 3.2 m **= 2 cells (ρ)**

BACS

Experimental plan (example contd.)

To avoid redundancies ⇒ ***only one quadrant*** is used

- *N_d – number of distances*
- *N_m – numbers of measurements in each cell*

Because:

- *Symmetry from* ***front-back confusion phenomenon***
- *Left-Right anti-symmetry (ITD = - ITD e ILD = - ILD)*

Auditory sensor space angular ranges simplify to (including PTU spec limitations for elevation):

Azimuth (θ) : 90º / 2º **= 45 cells**

Elevation (ϕ) : (- 30º to 30º) / 10º **= 6 cells**

Consequently: *Azimuth meas.* *Elevation meas.*

- $[N_d \times (45 \times 6)] \times N_m = 2 \times 270 \times N_m =$ **$540 \times N_m$** sets of measurements
- (*N_m measurements = 20 stimulus*) in each place to perform a statistical description

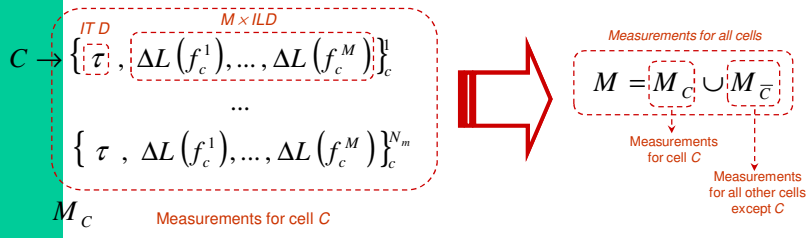
Becoming:

- $540 \times 20 = 10800$
- if each measurement takes 1s plus 1s of pause (play / record), calibration for each distance (i.e. the calibration process is conveniently divided into N_d sessions) will take: $10800 \times 2s / 2$ = ***3 h***

BACS

Binaural System Calibration

- Calibration is done in as with commonly used head-related transfer function (HRTF) calibration processes (see, for example, [11]), where the HRTF parameters are estimated by collecting binaural measurements through the positioning of a sound-source so as to uniformly sample sensor space.
- Measurement sets definitions (after applying processing binaural readings):



BACS

Binaural System Calibration

- *Normal probability distribution statistical characterisation process:*

$$P(\tau \mid S_C\ O_C\ \theta_{\max}) = \begin{cases} P(\tau \mid [S_C = 1]\ \theta_{\max}) \Rightarrow \mu_{M_C}(\tau)\ ,\ \sigma_{M_C}(\tau) \\ P(\tau \mid [S_C = 0]\ \theta_{\max}) \Rightarrow \mu_{M_{\bar{C}}}(\tau)\ ,\ \sigma_{M_{\bar{C}}}(\tau) \end{cases}$$

Sound-source at C; Measurements average; Standard deviation; No sound-source at C

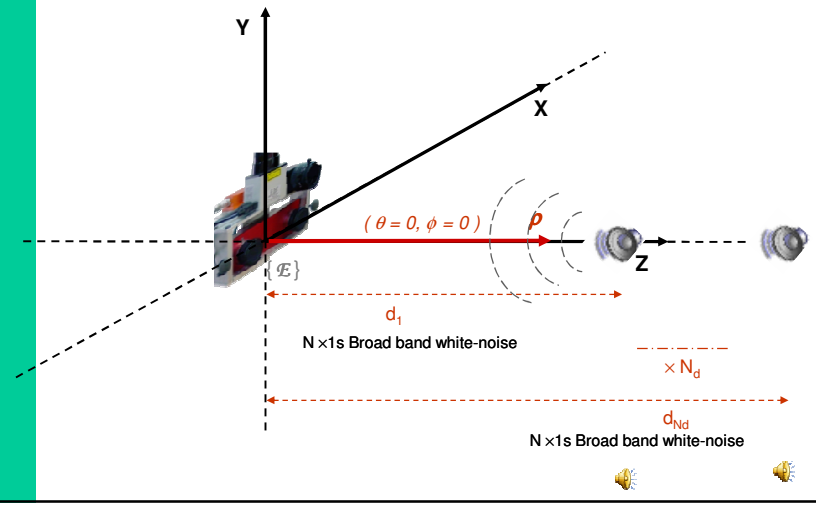
$$P(\Delta L \mid \tau\ S_C\ O_C\ C) \approx \begin{cases} P(\Delta L \mid [S_C = 1]\ C) \Rightarrow \mu_{M_C}(\Delta L)\ ,\ \sigma_{M_C}(\Delta L) \\ P(\Delta L \mid [S_C = 0]\ C) \Rightarrow \mu_{M_{\bar{C}}}(\Delta L)\ ,\ \sigma_{M_{\bar{C}}}(\Delta L) \end{cases}$$

Sound-source at C; Measurements average; Standard deviation; No sound-source at C

BACS

Binaural System Calibration Process

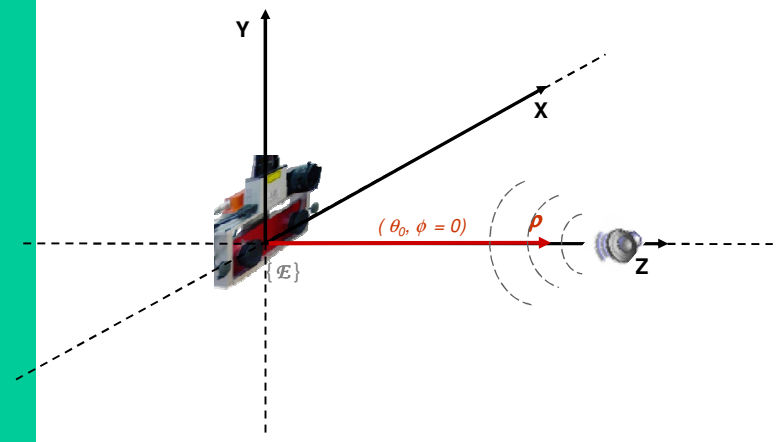
Distance calibration: for each (θ, ϕ)***, sound-source is positioned at*** N_d ***distances along Z axis at*** $(\theta = 0, \phi = 0)$



BACS

Binaural System Calibration Process

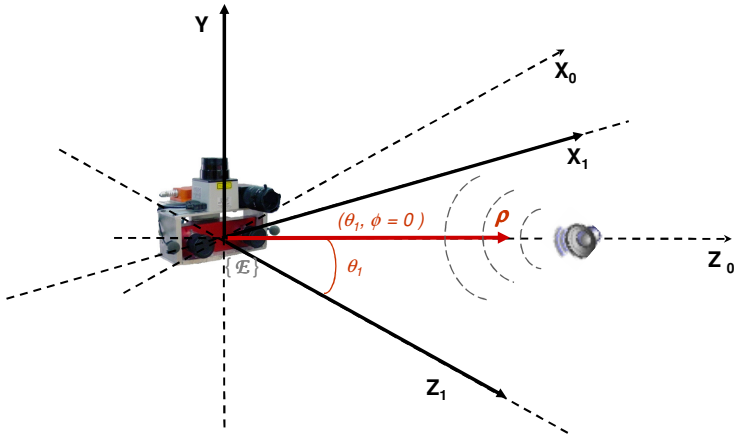
Rotating Sound-Source ⇔ Rotating Binaural System



BACS

Binaural System Calibration Process

Rotating Sound-Source ⇔ Rotating Binaural System



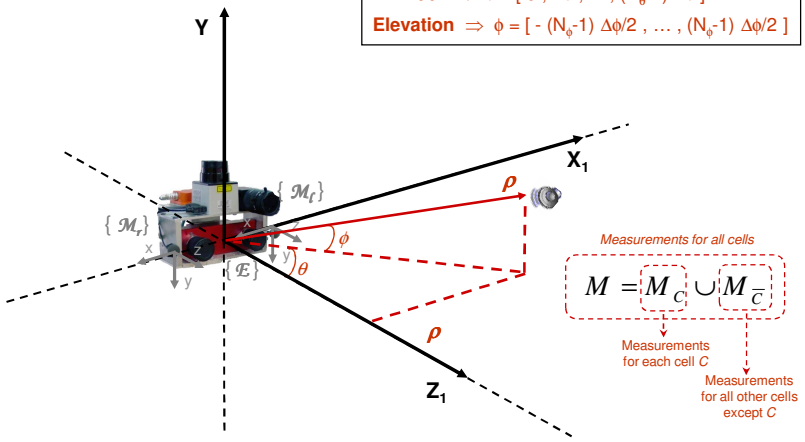
BACS

Binaural System Calibration Process

General Schematic:

Azimuth ⇒ $\theta = [\, 0 , \Delta\theta , \ldots , (N_\theta\text{-}1)\, \Delta\theta \,]$

Elevation ⇒ $\phi = [\, -(N_\phi\text{-}1)\, \Delta\phi/2 , \ldots , (N_\phi\text{-}1)\, \Delta\phi/2 \,]$



BACS

IASM - Inverse Auditory Sensor Model

IASM - Inverse Auditory Sensor Model

$$P([O_C=1]|ZC) = \frac{P([O_C=1]|C)P(Z|[O_C=1]C)}{\sum_{O_C=0,1} P(O_C|C)P(Z|O_C C)}$$

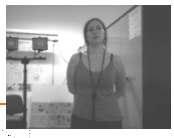
$$= \frac{P([O_C=1]|C)\,P(Z|[O_C=1]C)}{P([O_C=0]|C)\,P(Z|[O_C=0]C)+P([O_C=1]|C)\,P(Z|[O_C=1]C)}$$

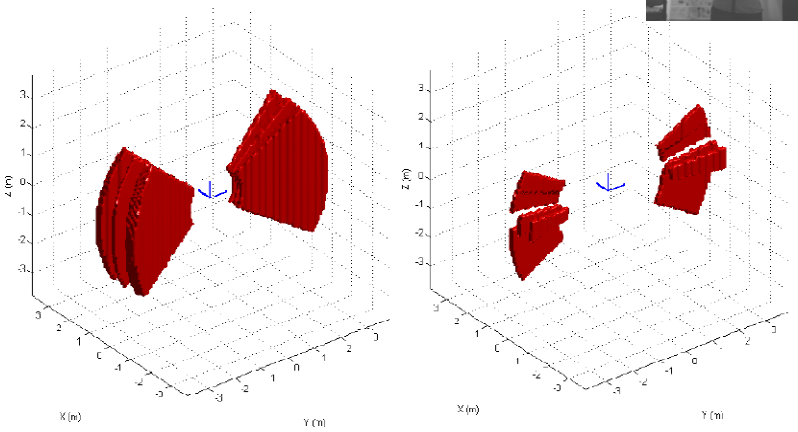
DASM

$$P(Z|[O_C=0,1]C) \begin{cases} O_C=0 \rightarrow P(\tau|[O_C=0]C)\prod_{k=1}^{m} P\left(\Delta L\left(f_c^k\right)|\tau[O_C=0]C\right) \\ O_C=1 \rightarrow P(\tau|[O_C=1]C)\prod_{k=1}^{m} P\left(\Delta L\left(f_c^k\right)|\tau[O_C=1]C\right) \end{cases}$$

BACS

Inference Using Audition Sensor Model – Results

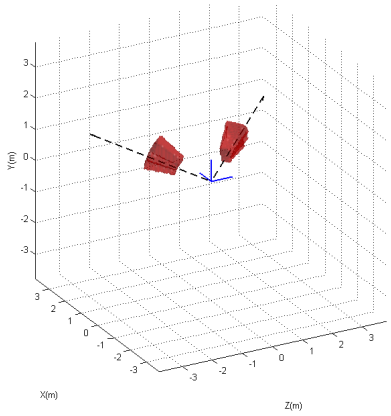




Results of Bayesian inference on the occupancy state of the BVM after processing an audio snippet of a human speaker placed in front of the binaural perception system — only BVM cells with probability of being occupied over .75 are represented. On the left, inference on the direct auditory sensor model using ITDs only; on the right, result of adding ILDs to the model. Note the effects of the front-to-back confusion, and of ILD dependence on distance and elevation, further improving inference precision.

BACS

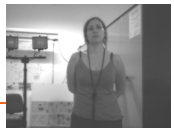
Inference Using Audition Sensor Model – Results



Occupancy results for the processing of an audio snippet of a sound-source placed at ρ = 1320 mm, θ = 36º, ϕ = 20º. Two dashed directional lines at (θ, ϕ) and (180º - θ, ϕ) have been additionally plotted to demonstrate the effect of front-to-back confusion. The fact that θ >> 0º means that precision in elevation and distance is improved as compared to the previous results.

BACS

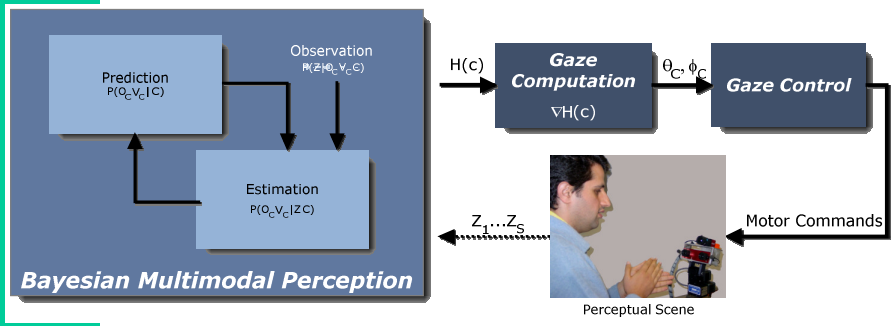
Inference Using BVM – Results



Ongoing Work…

BACS

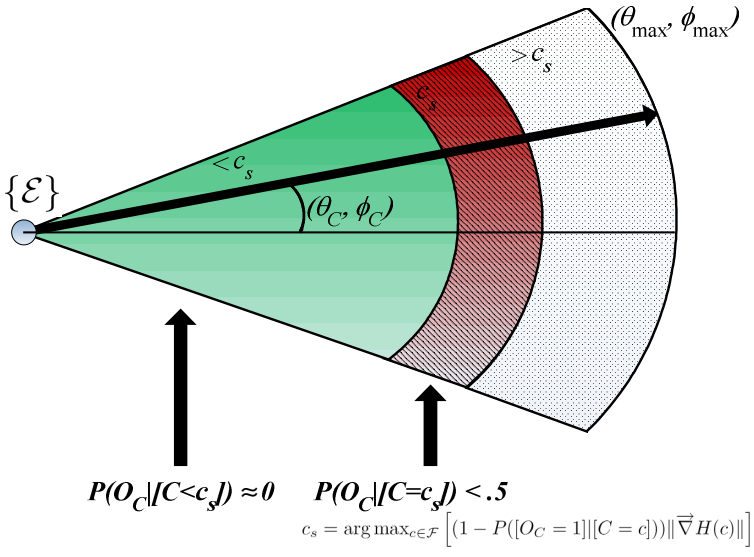
Using the BVM Framework for Entropy-Based Active Exploration



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

BACS

Using the BVM Framework for Entropy-Based Active Exploration

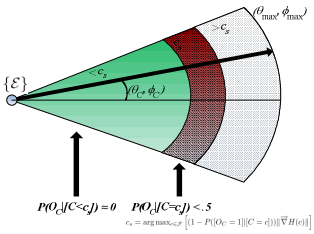


$$c_s = \arg\max_{c \in \mathcal{F}} \left[(1 - P([O_C = 1]|[C = c])) \|\vec{\nabla} H(c)\| \right]$$

BACS

BVM for Entropy-Based Active Exploration – Discussion

- Entropy-based exploration using a framework such as the BVM has ***biological support*** – e.g. most animals do not look at a scene in a steady way; it is believed that the eyes move in saccades so that small parts of a scene which have not been unambiguously resolved can be sensed with greater resolution, building up a mental "map" corresponding to the scene [1].
- Entropy-based exploration has been shown on previous work [13, 14] to be an efficient strategy for perception.
- The solution presented in this work, taking advantage of the log-spherical configuration provided by the BVM, allows for a much ***faster search algorithm*** for gaze orientation than the Cartesian solutions presented on [13, 14], thus improving on this concept.
- Moreover, the multimodal sensor fusion framework of the BVM allows for a ***robust perceptual solution***.



BACS

Conclusions and Future Work

- We have developed a Bayesian solution for multimodal perception implementing visual, auditory and vestibular/inertial sensory fusion.
- This solution will serve as a framework for implementing short-term egocentric spatial memory for entropy-based active multimodal perception and to support the modelling of other perceptual behaviours.
- An experimental platform has been built so as to provide a test bed for this framework.

- Website: http://paloma.isr.uc.pt/~jfilipe/BayesianMultimodalPerception
- Other selected publications:

- J. F. Ferreira, P. Bessière, K. Mekhnacha, J. Lobo, J. Dias, and C. Laugier, "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.
- C. Pinho, J. F. Ferreira, P. Bessière, and J. Dias, "A Bayesian Binaural System for 3D Sound-Source Localisation," in International Conference on Cognitive Systems (CogSys 2008), University of Karlsruhe, Karlsruhe, Germany, April 2008.
- J. F. Ferreira, C. Pinho, and J. Dias, "Active Exploration Using Bayesian Models for Multimodal Perception", A. Campilho and M. Kamel (Eds.): ICIAR 2008, LNCS 5112, pp. 369–378, 2008.

BACS

References (Selected Bibliography)

[1] - Ferreira, J. F., Castelo-Branco, M. 3D Structure and Motion Multimodal Perception, State-of-the-Art Report, Institute of Systems and Robotics and Institute of Biomedical Research in Light and Image, University of Coimbra, BACS, April 2007.

[2] - E. H. Weber. De pulsu, resorptione, audita et tactu. Annotationes anatomicae et physiologicae. Koehler, Leipzig, 1834. As cited by: Shen, Jianhong and Jung, Yoon-Mo, Weberized Mumford-Shah Model with Bose-Einstein Photon Noise, Applied Mathematics and Optimization, Springer, 53(2):331-358, May, 2006.

[3] - Tay, C., Mekhnacha, K., Chen, C., Yguel, M., and Laugier, C. An efficient formulation of the Bayesian occupation filter for target tracking in dynamic environments, International Journal of Autonomous Vehicles, 2008.

[4] - Coué, C., Pradalier, C., Laugier, C., Fraichard, T., and Bessière, P. Bayesian occupancy filtering for multitarget tracking: an automotive application. Int. Journal of Robotics Research, 25(1):19–30, 2006.

[5] - Laurens, J. and Droulez, J. Bayesian processing of vestibular information. Biological Cybernetics, December 2006.

[6] - Yguel, M. and Aycard, O. and Laugier, C. Efficient GPU-based Construction of Occupancy Grids Using several Laser Range-finders. International Journal Of Autonomous Vehicles, (Accepted) To be published.

[7] - Faller, C. and Merimaa, J. Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. J. Acoust. Soc. Am., 116:30753089, 2004.

[8] - B. G. Shinn-Cunningham, S. Santarelli, and N. Kopco, “Tori of confusion: Binaural localization cues for sources within reach of a listener,” J. of the Acoustical Society of America, vol. 107, no. 3, pp. 1627–1636, March 2000.

BACS

References (Selected Bibliography)

[9] - Akeroyd, M. A. A Binaural Cross-correlogram Toolbox for MATLAB. February 2001.

[10] - Johannesma, P. I. M. The pre-response stimulus ensemble of neurons in the cochlear nucleus. In Symposium on Hearing Theory, pages 5869. IPO, Eindhoven, The Netherlands, 1972.

[11] - Slaney, M. Auditory Toolbox: A MATLAB Toolbox for Auditory Modeling Work. Technical report, Interval Research Corporation, 1998.

[12] - P. T. Calamia, “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.

[13] - 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), 282–311 (2005) [2] - Tay, C., Mekhnacha, K., Chen, C., Yguel, M., and Laugier, C. An efficient formulation of the Bayesian occupation filter for target tracking in dynamic environments, International Journal of Autonomous Vehicles, 2008.

[13] - Rocha, R., Dias, J., Carvalho, A.: Exploring information theory for vision-based volumetric mapping. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005), Edmonton, Canada, August 2005, pp. 2409–2414 (2005).