A Bayesian Hierarchy for Gaze Following

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

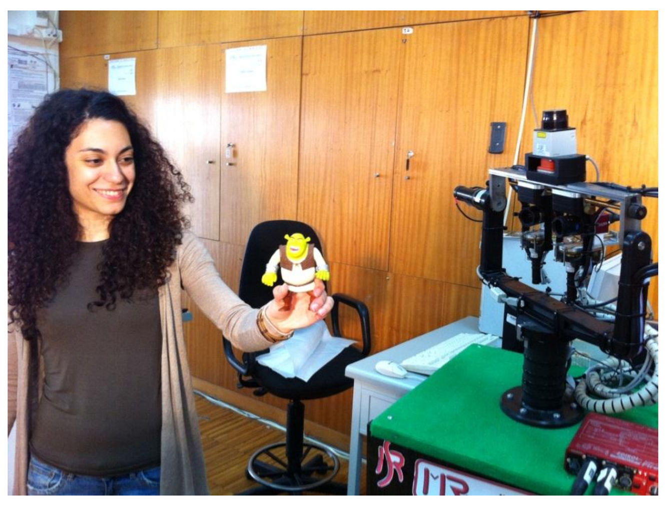
[1]gigel.macesanu@unitbv.ro, [2]{jfilipe,jorge}@isr.uc.pt

[1]Automation, Electronics and Computers Department, *Transilvania* University of Brașov, Romania

[2] Institute of Systems and Robotics, University of Coimbra, Portugal

Introduction

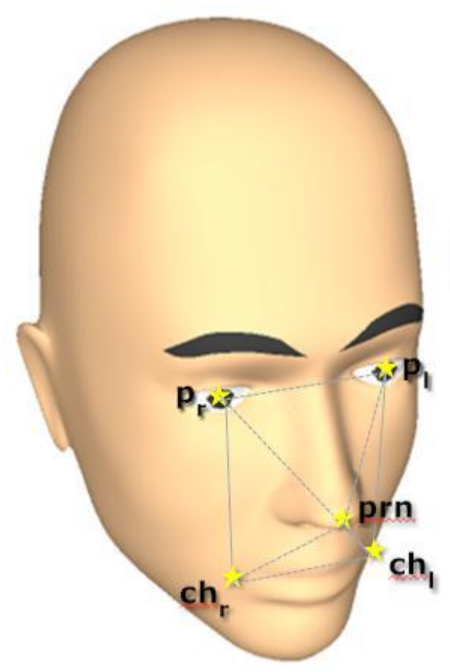
- In this work we propose a robust solution based on a probabilistic approach capable of dealing with perceptual uncertainty and incomplete data for gaze following
- The outcome of this project is an autonomous system, with the ability of robustly estimating gaze direction of interlocutors within the context of joint attention in Human Robotic Interface



- The proposed system, compared with others state of the art systems (e.g. see [1]), is able to return an estimate for gaze following even if only a partial set of features is available and integrate the dynamics of the evolution of the perceived scene

Head Pose Estimation and Features Detection

- It is believed that the human brain predicts the gaze of others as a combination of their head pose and their eye gaze direction [2]
- Human perception of gaze seems to rely on geometrical cues, such as the deviation of the nose angle and the deviation of the head from bilateral symmetry [1]
- Geometrical methods consist in obtaining orientation estimates from facial features, such as:
  - left and right pupil ($p_{r,l}$)
  - nose tip (prn)
  - mouth corners ($ch_{r,l}$)
- Features detection is performed as a combination of Haar classifier and color based segmentation



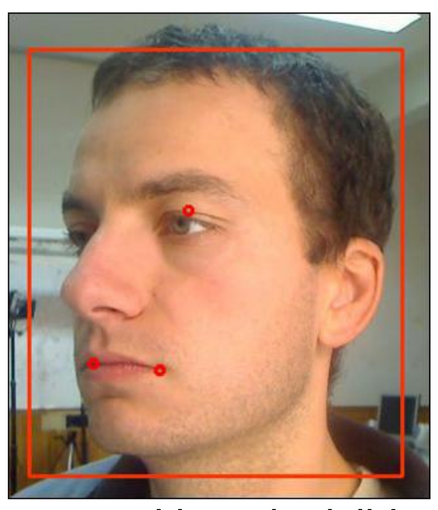
Probabilistic Framework

- probabilistic framework implementing a geometrical approach, inherently dealing with perceptual uncertainty and incomplete data

facial features are occluded

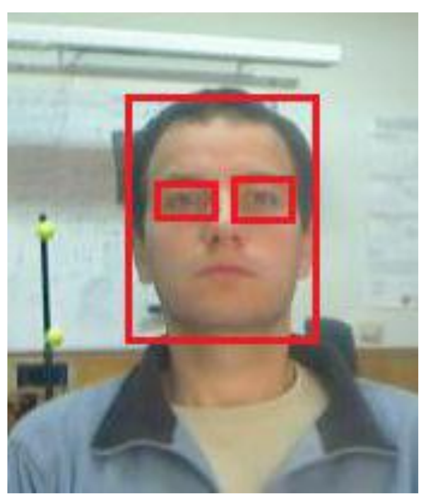
right eye occluded by nose

feature detection fails



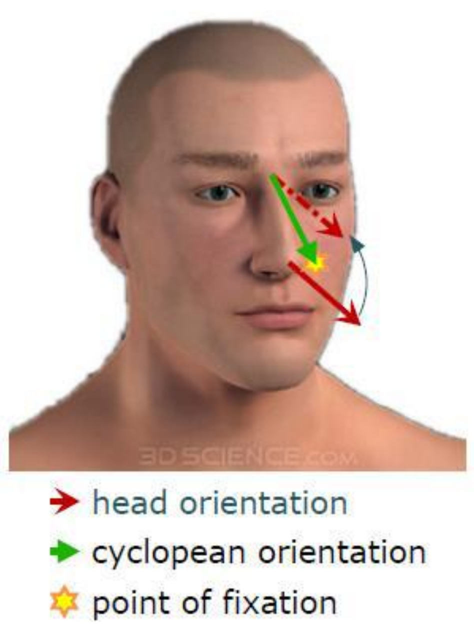
nose, although visible, is not detected

iris detection fails



eyes are detected, but iris detection fails do to distance

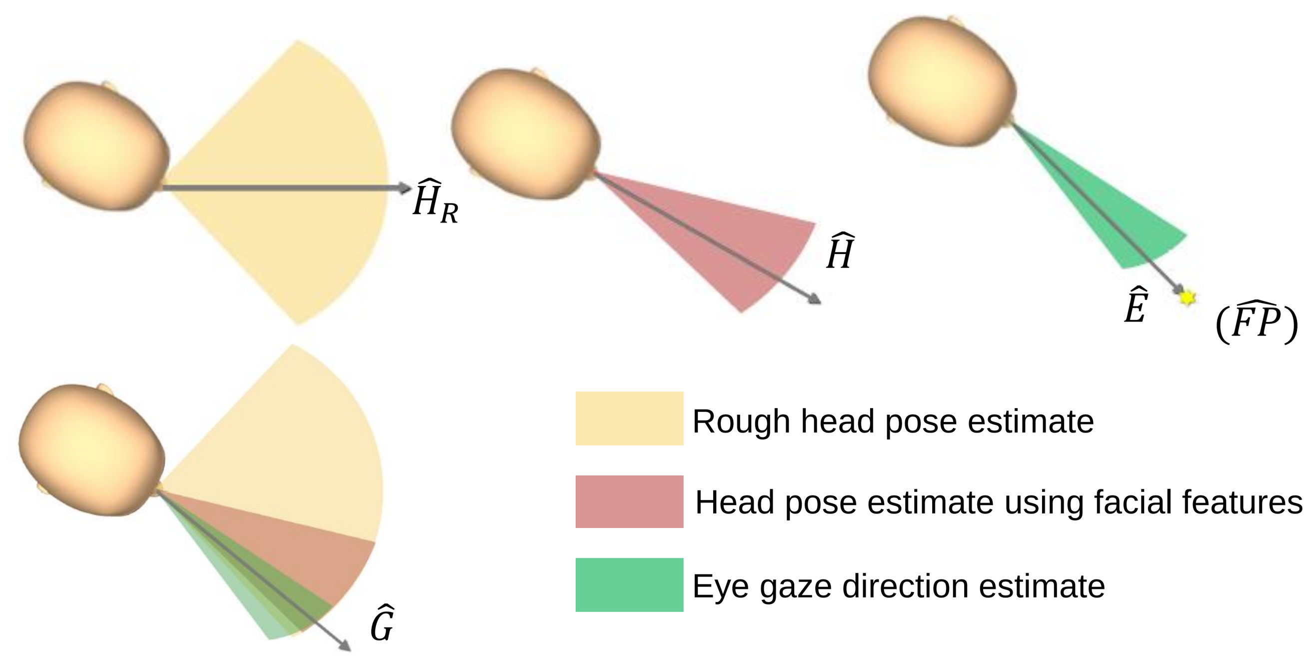
- To deal with incompleteness, the probabilistic framework takes all available data and produces the best possible estimate, taking into account the uncertainty caused by the missing information and by sensor model reliabilities
- This is done through ***probabilistic fusion*** of probability distributions associated with an incremental refining of the gaze estimate: head orientation, cyclopean orientation, prior distribution from previous time-steps.



This paper is supported by the Sectoral Operational Program Human Resources Development (SOPHRD), financed from the European Social Fund and by the Romanian Government under the projects POSDRU/88/1.5/S/59321

RObust VISion and Control Laboratory
Automation, Electronics and Computers Department,
Transilvania University of Brașov, Romania

Estimated Variables



Generative model for perception

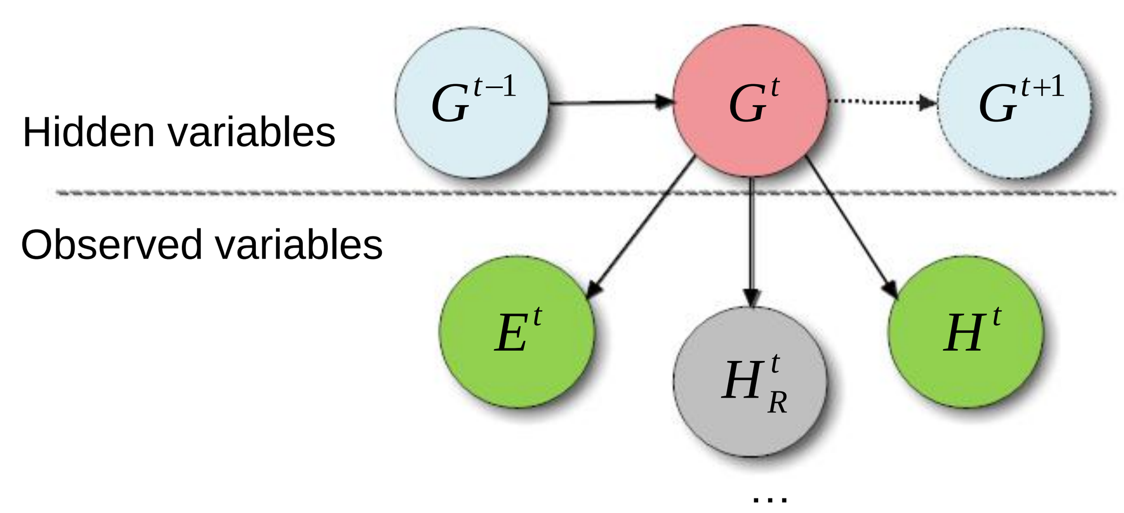
$$P(G,E,H,H_R)=P(G)P(E|G)P(H|G)P(H_R|G)$$

Perception is equivalent to answering the question: $P(G\,|\,e,h,h_R)$, which, using Bayes rule to perform inference becomes:

$$P(G|e,h,h_R)\propto P(G)P(e|G)P(h|G)P(h_R|G)$$

Bayesian framework

The Bayesian framework used, to perform the G^t estimation, is a hierarchical formulation of a **hHMM** (***hierarchical Hidden Markov Model***:



Experimental results

- During the training stage we use 120 images with different subjects that modify their poses from 0,5m to 2,2m. From these, distribution parameters for the feature detector sensor models (mean detection error in pixels σ' and the probability of a detection within acceptable error bounds) were obtained:

Learned Distribution Parameters		
Detector Name	P_{Hit}	σ'
Left/Right Iris	0,703	4,112
Left/Right Eye Left/Right Extremity	0,642	6,611
Nose tip	0,604	2,119
MouthLeftExtremity	0,824	1,272
MouthRightExtremity	0,865	1,599
Rough pose	0,5	-

- The mean detection errors are all added to the mean camera calibration reprojection errors to get the final σ.
- Estimation results show that our solution always yields an estimate for gaze, which is generally satisfactory, even in cases when only incomplete data is available.

Conclusions and Future Work

- We have implemented a robust solution to gaze following based on a probabilistic approach, even if only a partial set of features is available, with a clear indication of the uncertainty involved
- In the future, we expect to extend the framework for multiple interlocutors, and so as to deal with robot egomotion

[1] E. Chutorian and M. Trivedi, "Head pose estimation in computer vision: A survey," IEEE Transactions on Pattern Analyses and Machine Intelligence, vol. 31, no. 4, pp. 607–629, 2009.
[2] S. R. H. Langton, H. Honeyman, and E. Tessler, "The influence of head contour and nose angle on the perception of eye-gaze direction", Attention, Perception and Psychophysics, vol. 66, no. 5, pp. 752-771, 2004.

Mobile Robotics Laboratory
Institute of Systems and Robotics
ISR – Coimbra