Head Pose Estimation For a Domestic Robot

David van der Pol, Raymond H. Cuypers and James F. Juola
E-mail: d.v.d.pol@tue.nl

Introduction
Head pose estimation is an important gestural cue for a social robot because it signals turn-taking cues, attention, likeability etc. However, state of the art (Voit, Nickel and Stiefelhagen’s, 2005) mechanisms either have difficulties with different lighting conditions and low resolution images, or they are too slow for small computers.

Method
We extended Voit et al. (2005) method. We made it faster by (1) decreasing the resolution of the input image and (2) using just the edge filtered image. We averaged over multiple Neural Networks (NN) to obtain the head pose estimate. Each network was trained with a different selection of two databases. Validation was done with a third database.

Results
We found that individual networks differ in performance, and that taking the average over all networks results in the best estimate of head pose (Figure 1, NN Avg. is the average).

![Average Error, SE of the mean](image)

The average pixel intensity of the facial region determines the performance of the individual networks (see Figure 2). Edge filtering removes the intensity information. By adding the intensity as an input we increased the accuracy.

![Average Pixel Intensity of Face](image)

Conclusion
Our method is fast because it does not rely on high resolution images. Accuracy is improved by averaging over individual NN’s. It is therefore highly suitable for implementation in an autonomous robot.

References