POSEidon: Face-from-Depth for Driver Pose Estimation
Guido Borghi, Marco Venturelli, Roberto Vezzani, Rita Cucchiara
Imagelab - Dipartimento di Ingegneria "Enzo Ferrari " - University of Modena and Reggio Emilia - Italy

Motivations
- We aim at monitoring the driver attention, day and night
- Continuous head pose estimation provides useful cues
- Requirements:
  o Non-Invasive (no wearable devices)
  o High computation speed is mandatory (real-time processing)
  o Independency from external illumination
  o Embedded systems portability

Overall architecture of the system

Head Localization
- Input: depth frames
- Output: head center position (coordinates x,y)

The head size in pixels is estimated given the head center position and the depth (i.e., distance) values around it

Shoulder Pose Estimation
- Input: depth frames
- Output: 3D shoulder pose angles (yaw, pitch and roll)
A single network, with the same architecture of CNNs exploited for head pose estimation task
Combined with the head, shoulder pose helps to detect distractions

Head Pose Estimation
- Input: depth frames, Farneback Optical Flow images, Face-from-Depth images
- Output: 3D head pose angles (yaw, pitch and roll)
The overall POSEidon network is obtained as a fusion of 3 CNNs, individually trained for a regression on the 3D pose angles. Three additional fully connected layers are used to merge the contributions

Experimental results
1. public datasets exploited:
   - Biwi Kinect Head Pose: 15k images
   - ICT-3DHP database: 10k images
2. Pandora dataset
   - Annotation of shoulder angles
   - Wide angle ranges
   - Challenging camouflage and postures
   - Deep learning oriented (250k images)
   - High quality ToF data (Kinect v2)

Head Pose Estimation error (Angular accuracy)

Results on Biwi

Results on Pandora

The framework works at 30 fps on a desktop with GPU, while it processes around 10 fps on embedded devices.

ACKNOWLEDGMENTS - This work has been carried out within the project "HALES15 - Monitoring the car drivers attention with multimodality systems, computer vision and machine learning" funded by the University of Modena and Reggio Emilia. We also acknowledge the CINECA award under the ISCRA initiative, for the availability of high performance computing resources and support.