COMBINING DEEP AND DEPTH: DEEP LEARNING AND FACE DEPTH MAPS FOR DRIVER ATTENTION MONITORING

Guido Borghi, Roberto Vezzani, Rita Cucchiara
University of Modena and Reggio Emilia, Italy

Introduction

The human face is one of the richest source of information to monitor the **visual driver distraction**, *i.e.* when the driver's eyes are not looking at the road [5].

We assume the concept of attention as:

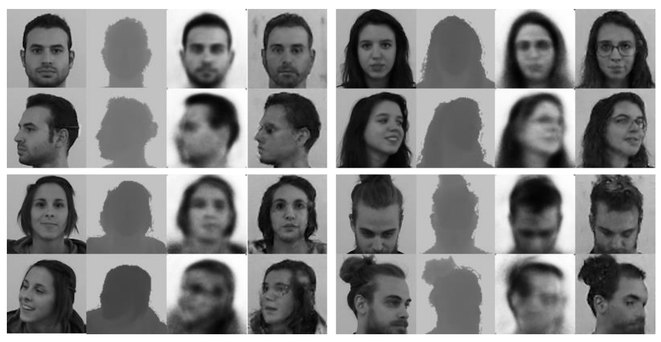
- **Head Pose Estimation**
- **Facial Landmark Detection** (for the analysis of salient elements belonging to the face).

We use only **depth images** as input data in order to develop methods that work even during the night or with bad light source conditions.

Summarizing, we present a study about algorithms that satisfy automotive requirements (light invariance, real time performance, occlusion reliability), focusing on Head Pose Estimation and Facial Landmark Detection tasks, based only on depth images

Face Generation

Face-from-Depth



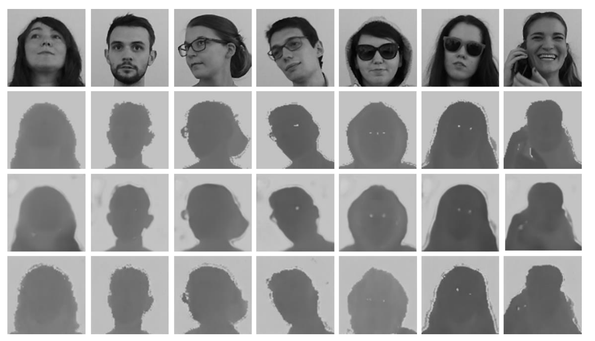
Is it possible to generate gray-level face images from the corresponding depth ones?

How: Deterministic Conditional GANs

Why: inspired by the *Privileged Information* approach, in which the main idea is to add knowledge at training time, we exploit the generated faces, in order to improve the performance of the presented systems at testing time.

Boost of performance for the Head Pose Estimation task

Depth-from-Face



Is it possible to generate depth face maps from the corresponding gray-level ones?

How: Deterministic Conditional GANs

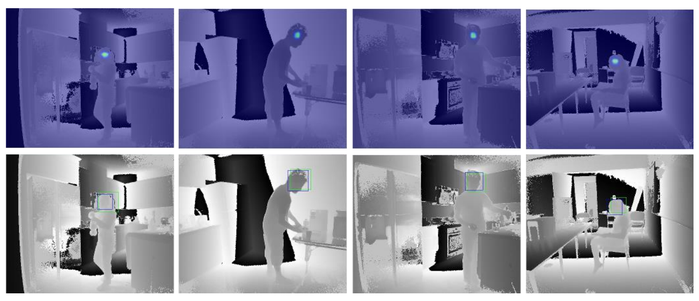
We show that the model is capable of predicting distinctive facial details by testing the generated depth maps through a deep model trained on authentic depth maps for the *Face Verification* task [4].

Head Detection

Head detection is the ability to detect and localize one or more heads in a given input image and is a key element for applications based on the analysis of the head.

Most of the current research approaches are based on images taken by conventional visible-light cameras – *i.e.* RGB or intensity cameras

We propose a *Fully Convolutional Network* that is able to output a probability map based on head locations, given a depth input map [1].



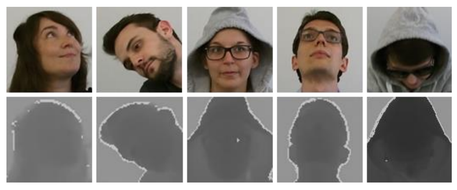
Experimental results show the good accuracy, the reliability and the speed performance (more than 30 fps) of the framework. **This method is currently the state-of-art for head detection in the wild with only depth images**.

FACE INPUT IMAGES

Given the head center, the input of the following methods can be computed.
The user's head is cropped with a *dynamic* window (*w, h*) to include smaller part of the background:

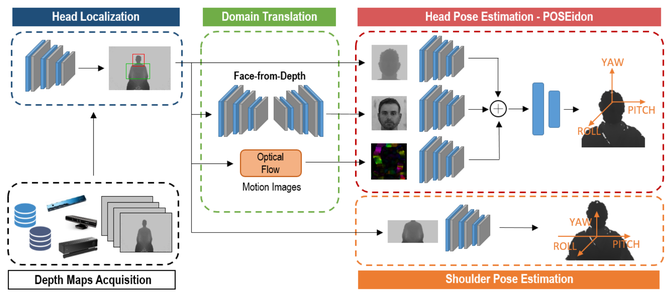
$$w, h = \frac{f_{x,y} \cdot R}{Z}$$

$f_{x,y}$: focal length , R: width of a generic face, Z: distance between the subject's face and the device.



Head Pose Estimation

Head Pose Estimation is the ability to infer the orientation of a person's head relating to the view of the acquisition device.



We propose a framework, called **POSEidon**, to estimate **head and shoulder poses**, measured as continuous rotation angles. A new triple regressive CNN architecture is introduced, that combines raw depth maps, Motion Images and generated faces from the corresponding depth images [2, 6].

One of the most innovative contribution is the *Face-from-Depth* network, that is able to reconstruct gray level faces directly from depth images (see the "Face Generation" section).

PANDORA DATASET

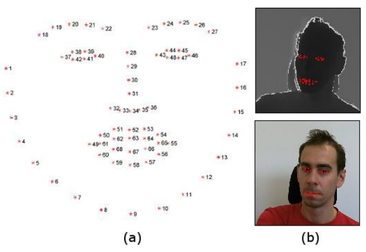
We also introduce a new dataset, namely **Pandora**:

- **110 annotated sequences** using 10 male and 12 female actors. Each subject has been recorded five times;
- The dataset contains **more than 250k** full resolution RGB (1920x1080) and depth (512x424) images;
- **Shoulder angles** included (*yaw*, *pitch* and *roll*);
- Occlusions generated by the **camouflage** (glasses, scarves, smartphones, tablets and so on).

http://imagelab.ing.unimore.it/pandora

Facial Landmarks

A reliable localization of facial landmarks – i.e. the ability to infer the position of prominent face elements relative to the view of the acquisition device – is one of the basic components to conduct driver physical state investigation, through eyes or mouth direct monitoring and facial expressions recognition, as reported in literature.



(a) (b)

We propose a deep-based approach specifically designed for real time facial landmarks localization in the automotive context, through a regression manner approach [3].

The model architecture is designed to deal with two main issues: **low memory** requirements and **real time** performance.

MOTORMARK DATASET

We introduce a new dataset called **MotorMark**:

- More than **30k** frames;
- **Both RGB and depth** images, acquired through the *Microsoft Kinect One*, are included;
- Annotation of **68 landmark positions** on both RGB and depth frames, following the ISO MPEG-4 standard.

http://imagelab.ing.unimore.it/motormark

Acknowlegments

This work has been carried out within the project "*FAR2015 - Monitoring the car driver's attention with multisensory 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.

References

[1] D. Ballotta, G. Borghi, R. Vezzani, and R. Cucchiara. Head detection with depth images in the wild. In 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2017
[2] G. Borghi, M. Venturelli, R. Vezzani, and R. Cucchiara. Poseidon: Face-from-depth for driver pose estimation. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017
[3] E. Frigieri, G. Borghi, R. Vezzani, and R. Cucchiara. Fast and accurate facial landmark localization in depth images for in-car applications. In International Conference on Image Analysis and Processing, pages 539–549. Springer, 2017.
[4] S. Pini, F. Grazioli, G. Borghi, R. Vezzani, and R. Cucchiara. Learning to generate facial depth maps. arXiv preprint arXiv:1805.11927, 2018.
[5] M. Venturelli, G. Borghi, R. Vezzani, and R. Cucchiara. Deep head pose estimation from depth data for in-car automotive applications. In International Workshop on Understanding Human Activities through 3D Sensors, pages 74–85. Springer, 2016.
[6] M. Venturelli, G. Borghi, R. Vezzani, and R. Cucchiara. From depth data to head pose estimation: a siamese approach. In 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP), 2016