Fast and Accurate Facial Landmark Localization in Depth Images for In-car Applications

Elia Frigieri, Guido Borghi, Roberto Vezzani, Rita Cucchiara
University of Modena and Reggio Emilia, Italy
name.surname@unimore.it

Abstract

A correct and reliable localization of facial landmark enables several applications in many fields, ranging from Human Computer Interaction to video surveillance. For instance, it can provide a valuable input to monitor the driver physical state and attention level in automotive context.

→ We tackle the problem of **facial landmark localization** through a deep learning-based approach.

→ It is more reliable than state-of-the-art competitors specially in presence of light changes and poor illumination, thanks to the use of **depth images** as input.

→ The developed system runs in **real time**.

→ We also collected and shared a **new realistic dataset** inside a car, called **MotorMark**, to train and test the system. In addition, we exploited the public *Eurecom Kinect Face Dataset* [1] for the evaluation phase.

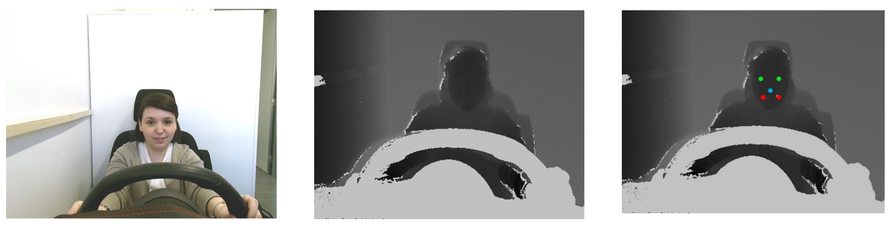


Figure 1. RGB, depth and facial landmarks

MotorMark dataset

We collected a dataset that includes **RGB** and the corresponding **depth images**, annotated with facial landmark coordinates on RGB and depth images. Frames are acquired through a *Microsoft Kinect One*.

Main features:

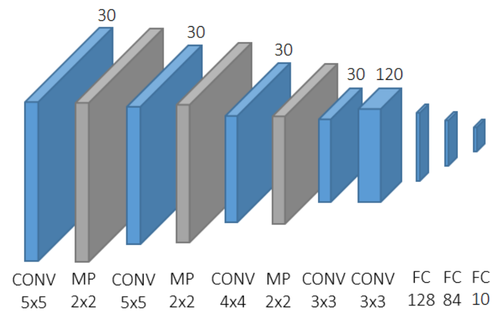
→ **Deep oriented**: is composed by more than 30k frames. A variety of subjects is guaranteed (35 subjects in total);

→ **Automotive context**: we recreate an automotive context. The subject is standing in a real car dashboard and performs real inside-car actions;

→ **Variety**: subjects are asked to follow a constrain path to rotate their head in fixed position or to freely move their head. Besides, subjects can wear glasses, sun glasses and a scarf, to generate partial face and landmark occlusions;

→ **Landmark annotations**: the annotation of 68 landmark positions on both RGB and depth frames is available, following the ISO MPEG-4 standard. The ground truth has been manually generated.

Proposed Method



INPUT

Depth images with this pre-processing:

- Contrast Limited Adaptive Histogram Equalization algorithm;
- Values scaled, mean and variance are 0 and 1;
- A fixed window containing the head is cropped and all the cropped images are resized to 64x64 pixels;

DEEP MODEL

Shallow model (real time performance)

- 5 convolutional layers;
- 3 Max-pooling layers ;
- 3 fully connected layers;
- Activation function: *tanh;*
- L_2 loss;

OUTPUT

10 coordinates (*x,y*) of 5 facial landmarks

- Nose tip, eye centers and mouth corners;
- Ground truth coordinates are normalized in the range [-1, 1], accordingly to the specific activation function of the output network layer;

Results

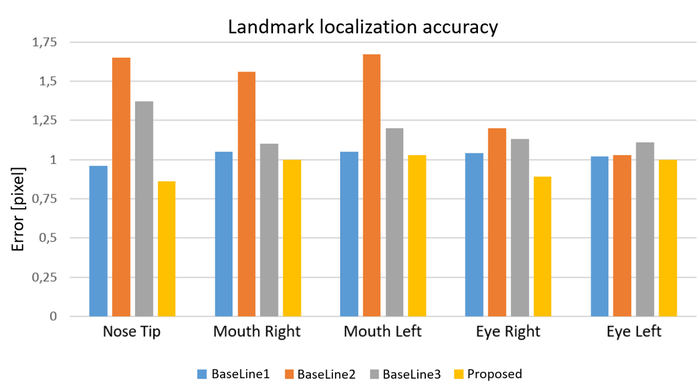
Method	Nose Tip	Mouth Right	Mouth Left	Eye Right	Eye Left	Avg Err
Zhao *et al.*	4.4±2.2	5.4±3.2	5.4±3.2	4.2±2.1	4.2±2.2	4.7±2.6
Our	3.3±4.5	3.5±3.7	3.4±3.9	3.5±4.1	3.4±4.0	3.4±4.0

Table 1 Results on *Eurecom Kinect Face dataset*, expressed as the mean error and the standard deviation in pixels w.r.t the ground truth, normalized by the interpupillary distance. (Zhao *et al.* [2])

Results on *MotorMark* dataset expressed as the mean error and the std in pixels w.r.t the ground truth, normalized by the interpupillary distance.

Different tests have been carried out:

1. Smaller window cropped from head center
2. Bigger input images (128x128)
3. Background suppression applied
4. The proposed method



Acknowledgements

This work has been carried out within the projects “Citta educante” (CTN01-00034-393801) of the National Technological Cluster on Smart Communities funded by MIUR and “FAR2015 - Monitoring the car drivers attention with multi-sensory 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] Min, Rui, Neslihan Kose, and Jean-Luc Dugelay. "Kinectfacedb: A kinect database for face recognition." IEEE Transactions on Systems, Man, and Cybernetics (2014)

[2] Zhao, Xi, et al. "Automatic 2.5-D facial landmarking and emotion annotation for social interaction assistance." IEEE transactions on cybernetics (2016)

19th International Conference on Image Analysis and Processing
11—15 september 2017
Catania—Italy