HOG ACTIVE APPEARANCE MODELS

Epameinondas Antonakos* , Joan Alabort-i-Medina*,
Georgios Tzimiropoulos¬*, Stefanos Zafeiriou*

* Imperial College London, Department of Computing, U.K.
¬ University of Lincoln, School of Computer Science, U.K.

CONTRIBUTIONS

We propose a facial landmark points localization technique in-the-wild that combines:
1) *Dense* Histogram of Oriented Gradients (HOG) descriptors
2) with the Inverse Compositional optimization of Active Appearance Models AAMs)

This results in a generic facial model that outperforms current state-of-the-art techniques.

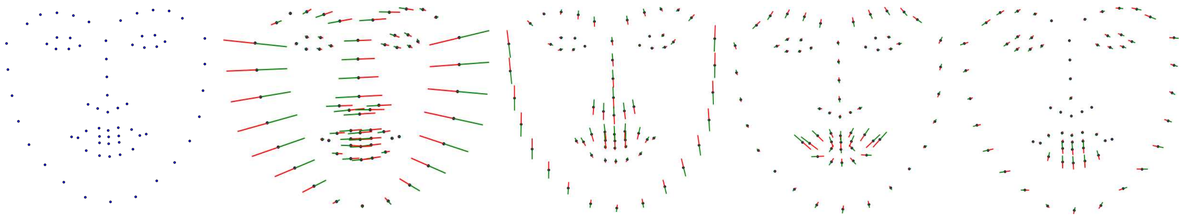
DENSE HOG DESCRIPTORS

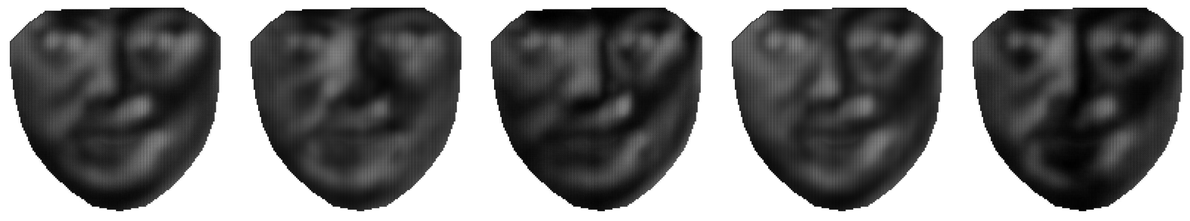
For each pixel location of the image we apply the following:
1) Create a histogram of the gradient's orientations for a rectangular neighbourhood around the pixel, weighted by the gradient magnitude.
2) Apply contrast normalization to the histogram based on the Euclidean norm.

Thus, for an input image of size HxW, the output image has size HxWxC where C is the number of channels.

ACTIVE APPEARANCE MODELS

AAMs are generative, statistical, parametric models of an object's shape and appearance.
- The shape model is built by aligning the training shapes wrt their similarity transform and applying PCA.
- The appearance model is built by extracting HOG features from the training images, warping the multichannel texture onto a common reference shape (i.e. mean shape) and apply PCA.



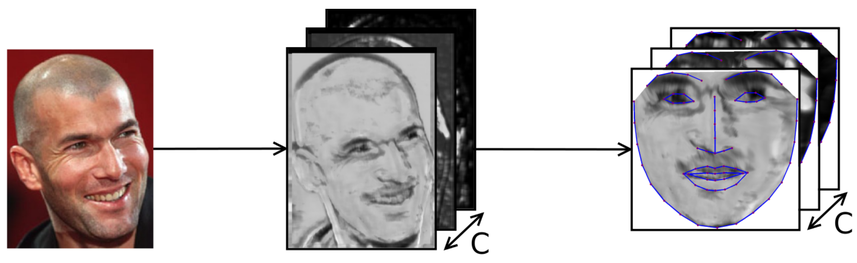


We employ two Gauss-Newton optimization techniques:

Alternating Inverse Compositional
- Optimizes alternatingly wrt the shape and appearance parameters
- Large parametric space
- Fairly fast and very accurate

Project-Out Inverse Compositional
- Only uses the mean apperance vector
- Small parametric space (shape parameters only)
- Very fast but poor accuracy

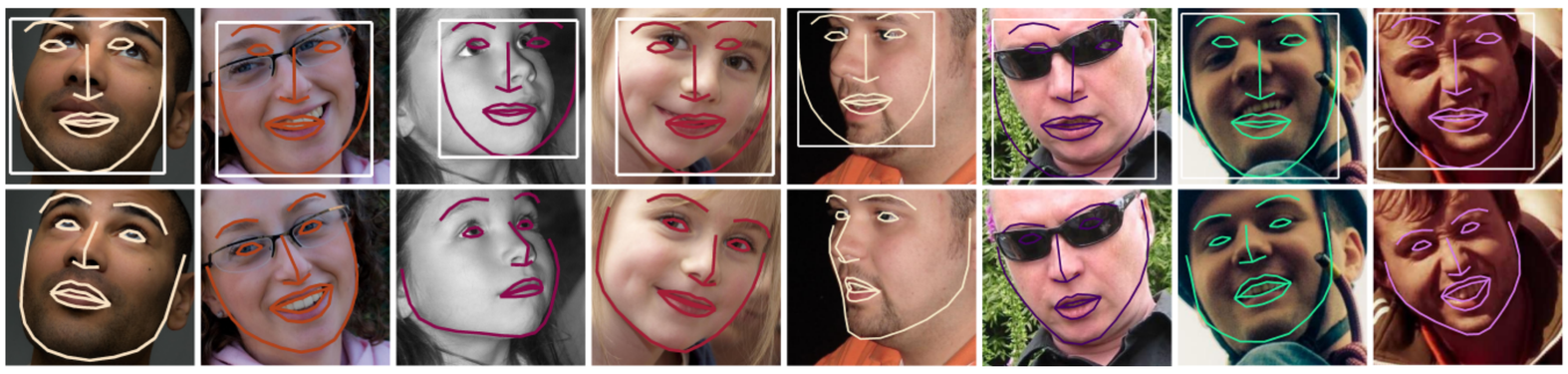
During fitting, we extract the HOG features *once* and then warp the multichannel appearance at each iteration.
This is much faster than extracting features at each iteration.

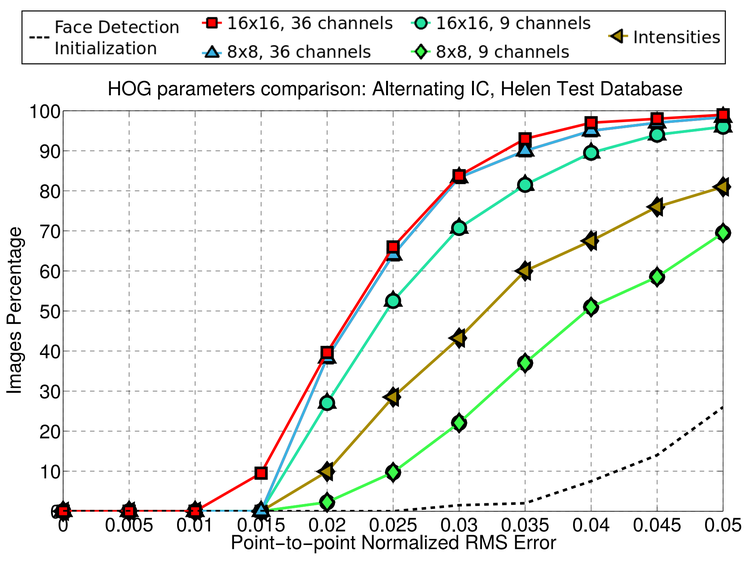


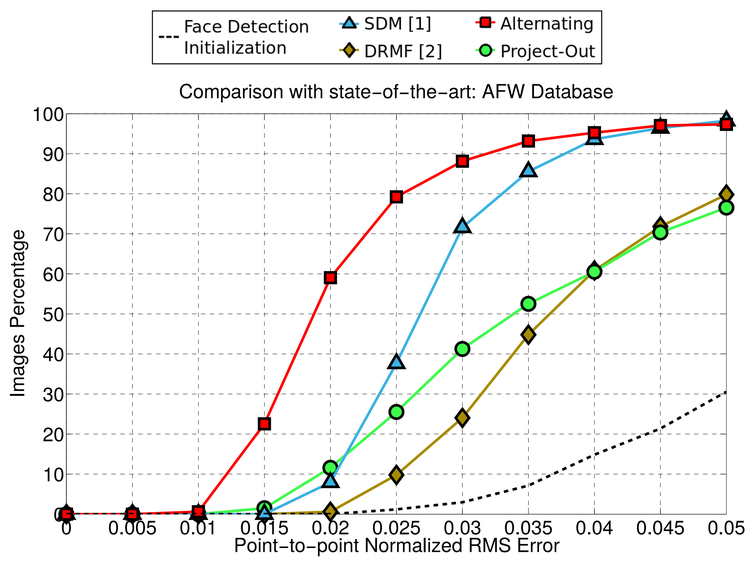
EXPERIMENTAL RESULTS

- Training on 811 images of LFPW database
- 15 eigenshapes, 100 eigentextures
- Initialization using method in [3].

The proposed methods proves to be accurate even with challenging initializations!







ACKNOWLEDGMENTS

The work of E. Antonakos and S. Zafeiriou was partially funded by the EPSRC project EP/J017787/1 (4D-FAB). The work of J. Alabort-i-Medina was funded by the Qualcomm Innovation Fellowship and by a European DTA from Imperial College London.

REFERENCES

[1] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment", CVPR 2013.
[2] A. Asthana, S. Zafeiriou, S. Cheng and M. Pantic, "Robust discriminative response map fitting with constrained local models", CVPR 2013.
[3] J. Orozco, B. Martinez and M. Pantic, "Empirical analysis of cascade deformable models for multi-view face detection", ICIP 2013.