Face alignment assisted by Head Pose Estimation

Heng Yang[1]
heng.yang@cl.cam.ac.uk

Wenxuan Mou[2]
w.mou@qmul.ac.uk

Yichi Zhang[3]
yichizhang@fas.harvard.edu

Ioannis Patras[2]
i.patras@qmul.ac.uk

Hatice Gunes[2]
h.gunes@qmul.ac.uk

Peter Robinson[1]
peter.robinson@cl.cam.ac.uk

[1] Computer Laboratory
University of Cambridge
Cambridge, UK

[2] School of EECS
Queen Mary University of London
London, UK

[3] Faculty of Arts & Sciences
Harvard University
Cambridge, MA, US

Both head pose estimation and face alignment have been well studied in recent years given their wide application in human computer interaction, avatar animation, and face recognition/verification. However, even the most sophisticated face alignment methods also show some failures when they are applied on face images collected in the wild. In this paper, we show how face alignment can be improved by explicit head pose estimation. In summary, we make the following contributions:

- We investigate the failure cases of several state of the art face alignment approaches and find that the head pose variation is a common issue across those methods. See Fig. 1.
- Based on the above observation, we propose a ConvNet framework for explicit head pose estimation (Fig. 2). It is able to achieve an accuracy of 4° absolute mean error of head pose estimation for face images acquired in unconstrained environment.
- We propose two initialisation schemes based on reliable head pose estimation. They enable baseline face alignment method (RCPR [1]) perform better and reduce large head pose failures by 50% when using only one initialisation.

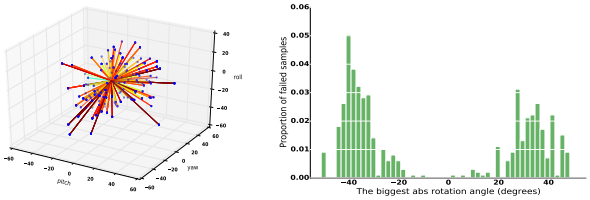

Figure 1: Distribution of the most erroneous samples of various state of the art face alignment methods.

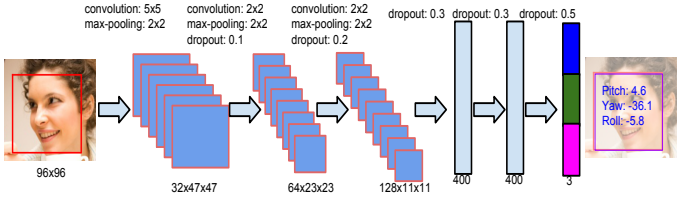

Figure 2: ConvNet model for head pose estimation.

3D face shape based initialisation scheme Given a 3D mean face shape, represented by 68 3D facial landmark locations, as shown in Fig. , we first project this shape under the estimated head pose to a set of canonical 2D locations. More specifically we use constant translation and focus length in order to get a reasonable projection for all images. Then we re-scale the canonical 2D projection by the face bounding box scale of the test image to get the initialisation. We can represent the initialisation process by function $\mathcal{F}$ as follows.

$$S_0 = \mathcal{F}(\theta, bb, \bar{S}^{3D}) \tag{1}$$

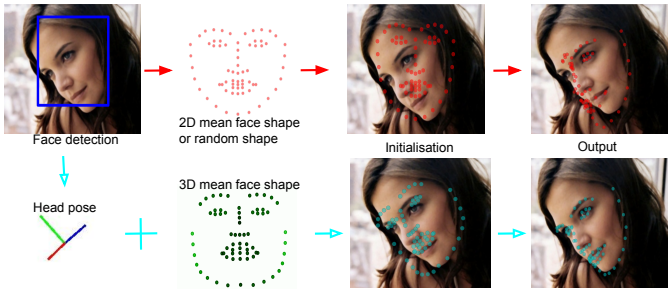

Figure 3: Our proposed head pose based cascaded face alignment procedure (path in cyan color) vs. conventional cascaded face alignment procedure (path in red color).

with bb the face bounding box, $\bar{S}^{3D}$, the 3D mean face shape, θ, the estimated head pose, which can be represented by:

$$\theta = \mathcal{G}(I, bb) \tag{2}$$

where $\mathcal{G}$ is the deep convolutional model.

Results

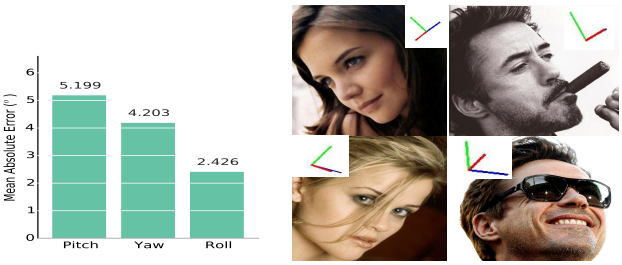

Figure 4: Head pose estimation result. Left, absolute mean error on test set; right, example results of head pose estimation.

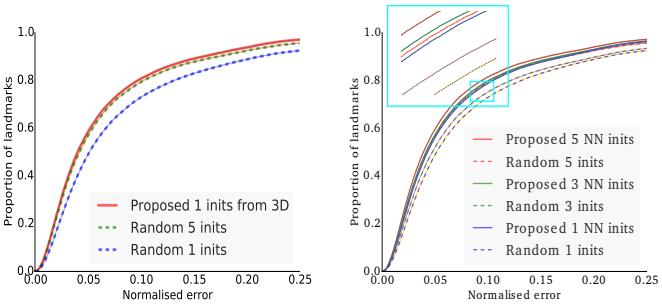

Figure 5: Comparison to baseline method with random initialisation.

[1] Xavier P Burgos-Artizzu, Pietro Perona, and Piotr Dollár. Robust face landmark estimation under occlusion. In *Proc. IEEE Int. Conf. Comput. Vis.*, pages 1513–1520, 2013.