Pose-Robust 3D Facial Landmark Estimation from a Single 2D Image

Brandon M. Smith
http://www.cs.wisc.edu/~bmsmith
Charles R. Dyer
http://www.cs.wisc.edu/~dyer

Department of Computer Sciences
University of Wisconsin-Madison
Madison, WI USA

Despite much research interest in facial landmark estimation in recent years, relatively little work has been done to handle the full range of head poses encountered in the real world (*e.g.*, beyond $\pm 45^\circ$ rotation). As a result, the large majority of face alignment algorithms are limited to near fronto-parallel faces, and break down on profile faces. We propose an approach to face alignment that can handle 180° of head rotation.

The foundation of our approach is cascaded shape regression (CSR), which has emerged as the leading strategy (see, *e.g.*, [2]). To better handle a wide range of head poses, we extend the 2D CSR approach to 3D. That is, instead of fitting a 2D face model to single 2D images, we fit a 3D face model to single 2D images (3D-to-2D). Intuitively, as the range of head poses increases, the 3D geometry of the face becomes increasingly important in explaining its 2D image projection.

Recent facial landmark estimation methods, including 3D-to-2D approaches [3], employ *local* optimization algorithms at each cascade level, which can fail on face collections with large head pose variation. It is unlikely that a single cascade of generic domain maps (from input features to output landmark updates) will consistently find the true solution. We therefore partition the shape regression problem into a set of simpler *viewpoint domains*, and learn a separate cascade of regressors for each. Each viewpoint domain corresponds to an automatically learned range of camera viewpoints/head poses, as shown in Figure 1. At test time our algorithm adaptively chooses which CSR to apply.

Despite a recent trend toward modeling face shape nonparametrically (*e.g.*, directly updating landmark coordinates), we adopt a parametric model and show empirically that there are no significant differences in accuracy between parametric and nonparametric shape models.

CSR methods commonly use off-the-shelf feature mapping functions (*e.g.*, SIFT) to produce features from the image. Instead, we employ regression random forests [1] to *learn* local binary features that predict ideal shape parameter updates.

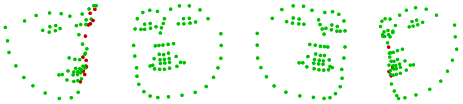

Figure 1: The first four modal viewpoints found for $V = 8$ viewpoint domains. The modal occlusion state is stored for each viewpoint domain (green is visible, red is occluded).

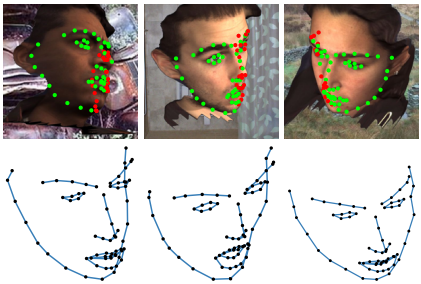

Figure 2: Qualitative results on faces from BU-4DFE [4]. Top row: estimated visibility of each landmark (green is visible, red is occluded). Bottom row: estimated 3D shape.

Results demonstrate quantitatively that the proposed approach is significantly more accurate than recent work. Figure 2 shows a sample of qualitative results.

[1] Leo Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

[2] Shaoqing Ren, Xudong Cao, Yichen Wei, and Jian Sun. Face alignment at 3000 fps via regressing local binary features. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[3] Sergey Tulyakov and Nicu Sebe. Regressing a 3D face shape from a single image. In *IEEE International Conference on Computer Vision*, 2015.

[4] Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, and Michael Reale. A high-resolution 3d dynamic facial expression database. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2008.