A 3D Morphable Eye Region Model for Gaze Estimation

Erroll Wood[1] Tadas Baltrušaitis[2] Louis-Philippe Morency[2] Peter Robinson[1] Andreas Bulling[3]

[1]University of Cambridge, United Kingdom, {erroll.wood, peter.robinson} @cl.cam.ac.uk
[2]Carnegie Mellon University, United States, {tbaltrus, morency} @cs.cmu.edu
[3]Max Planck Institute for Informatics, Germany, bulling@mpi-inf.mpg.de

Abstract

Morphable models are a powerful tool, but have so far failed to model the eye accurately. We present a multi-part model that includes a new 3DMM of the facial eye region, and an anatomy-based eyeball. By fitting this model to an image, we can estimate 3D gaze robustly.

Model parameters Φ

We parameterize our multi-part model with

$$\Phi = \{ \beta, \tau, \theta, \iota, \kappa \}$$

Shape β and texture τ are controlled with linear models. Pose θ is defined by model transforms and procedural animation. We use ambient + directional illumination ι. We assume knowledge of camera parameters κ.

Fitting our model

We fit our model with *analysis-by-synthesis*: given an observed image I_{obs}, we produce a synthesized image I_{syn} that best matches it. We search for optimal parameters Φ^* as follows:

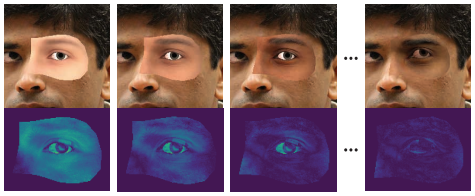
$$\Phi^* = \underset{\Phi}{argmin}\ E(\Phi)$$

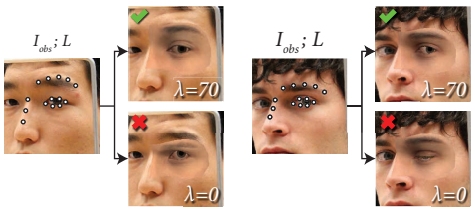
Our energy includes E_{img} that measures the dense pixel-wise similarity between I_{obs} and I_{syn}, and E_{ldmks} that regularizes our model against tracked facial feature landmarks [1]. λ controls their relative importance.

$$E(\Phi) = E_{img}(\Phi) + \lambda \cdot E_{ldmks}(\Phi, L)$$

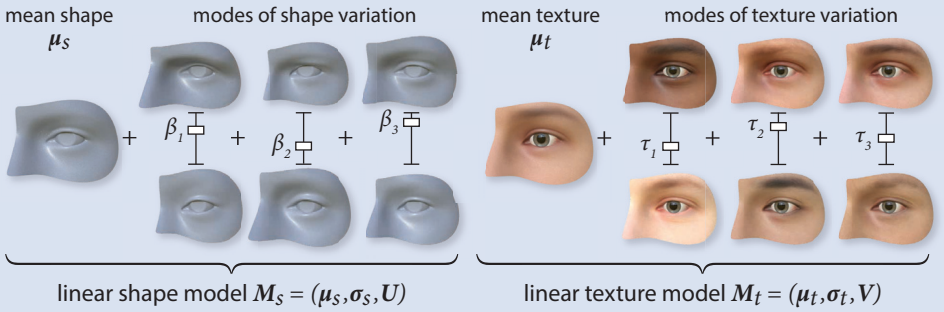
We minimze $E(\Phi)$ using gradient descent with numerical central derivatives, solving for all parameters simultaneously.

Reduction in E_{img} over 60 GD iterations





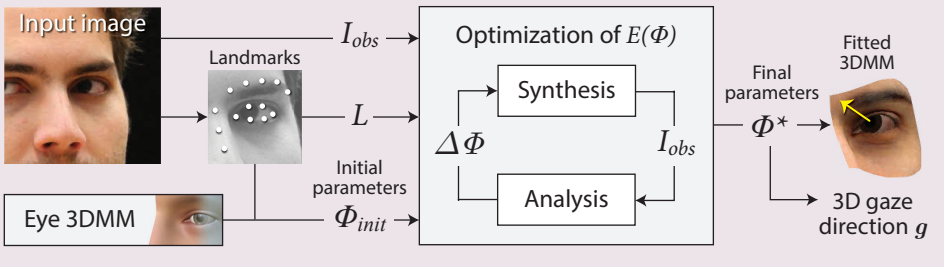
Eye Region 3D Morphable Model (β, τ)



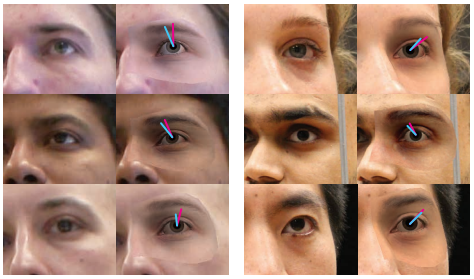
We built linear PCA models of shape M_S and texture M_t from a collection of high resolution 3D face scans (see bottom right). These scans were carefully re-parameterised to maintain detailed correspondences around the eyelids and interior margin. A similar PCA linear model M_{iris} was constructed for iris texture.

We used 22 high quality head scans as source data
5M vertices, >1mm accuracy, http://www.3dscanstore.com

System overview



Eyediap [3] Columbia [2]

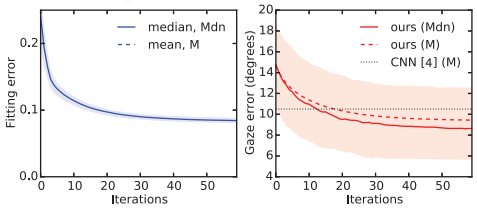


true gaze in **blue**, estimated gaze in **magenta**

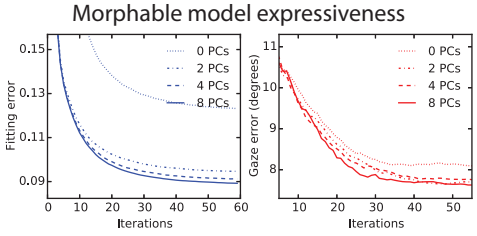
Gaze estimation results

We evaluated on the Columbia [2] and Eyediap [3] datasets. We (**9.44°** gaze error) outperformed a state of the art convolutional neural network method [4] (10.5°). Our model based approach also generalizes to extreme gaze angles not covered in their limited training set.

Eyediap (VGA) gaze error



Morphable model expressiveness



Top: a comparison against a state-of-the-art CNN method [4] Bottom: more PCs lets our model fit better, and gives better gaze error.

[1] T. Baltrusaitis, P. Robinson, and L.-P. Morency, Constrained local neural fields for robust facial landmark detection in the wild, ICCVW, 2013
[2] Smith, B., Yin, Q., Feiner, S., Nayar, S. Gaze Locking: Passive Eye Contact Detection for HumanObject Interaction, UIST 2013
[3] Funes Mora, K.A., Monay, F., Odobez, J.M. EYEDIAP, ETRA 2014
[4] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, Appearance-Based Gaze Estimation in the Wild, CVPR, 2015.