Inverse Rendering Best Explains Face Perception Under Extreme Illuminations

Bernhard Egger[1,2] Max Siegel[1,2] Riya Arora[1] Amir Arsalan Soltani[1,2] Ilker Yildirim[3] Josh Tenenbaum[1,2]

[1]Massachusetts Institute of Technology, [2] Center for Brains, Minds, & Machines, [3]Yale University

Overview

- Humans can successfully interpret images and recognize objects even under **significant image transformations**
- Such significantly transformed images could aid in **differentiating computational architectures** for perception
- We study two classes of degraded stimuli – **Mooney faces and silhouettes** of faces – as well as typical faces

- We find that our **top-down inverse rendering** model better matches human percepts than either an invariance-based account implemented in a deep neural network (VGG-face), or a neural network trained to perform approximate inverse rendering in a feedforward circuit (EIG).

Top-Down Inverse Rendering Model

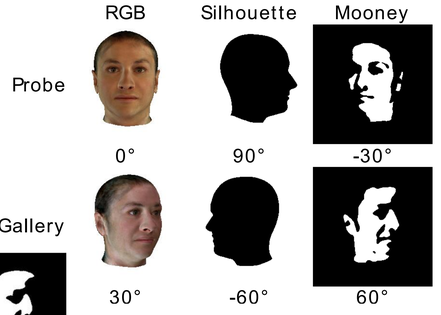
Our inverse rendering methods are based on a generative statistical model of human faces, namely the 3D Morphable Model [5]. In our experiments we estimate shape, albedo and illumination parameters using an MCMC inference strategy.

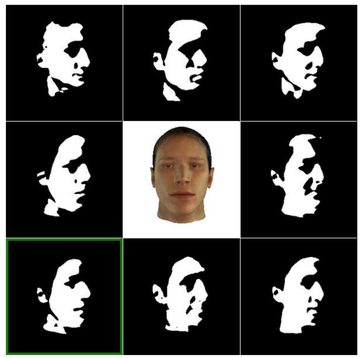
target | IlluminationOnly (IR-Illum.) | TransformLayer (IR-Trans.)

Here we show reconstructions of the target image for both our two computational models. IlluminationOnly uses the standard 3D Morphable Model approach and explains Mooney images via Illumination. TransformLayer actively renders a Mooney reconstruction as postprocessing.

Experiments

Full set of synthetic images per identity. We vary yaw-rotation and illumination across the different settings.





Example of a 8-AFC trial shown to participants. The RGB image is the probe and the Mooney images are the gallery.
We recruited 60 participants for each of Exps. 1 and 2, via Amazon Mechanical Turk, for a total of 120 participants

References

[1] Egger, B., Schonborn, S., Schneider, A., Kortylewski, A., Morel-Forster, A., Blumer, C., & Vetter, T. (2018). *Occlusion-aware 3d morphable models and an illumination prior for face image analysis*. IJCV

[2] Gerig, T., Morel-Forster, A., Blumer, C., Egger, B., Luthi, M.,Schonborn, S., & Vetter, T. (2018*). Morphable face models -an open framework*. FG

[3] Yildirim, I., Belledonne, M., Freiwald, W., & Tenenbaum, J. (2020). *Efficient inverse graphics in biological face processing*. ScienceAdvances,6(10).

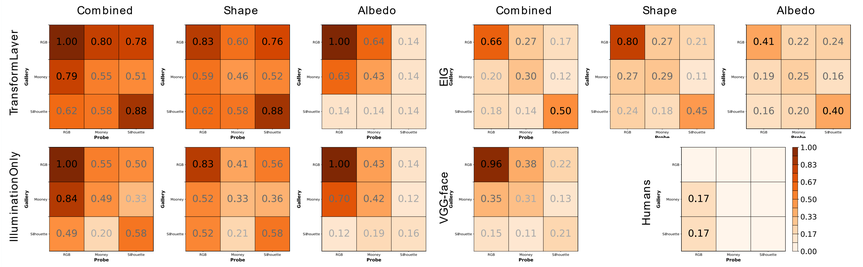
[4] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015*). Deep face recognition*. BMVC.

[5] Blanz, V., & Vetter, T. (1999). *A morphable model for the synthesis of 3D faces*. In TOG

Acknowledgment

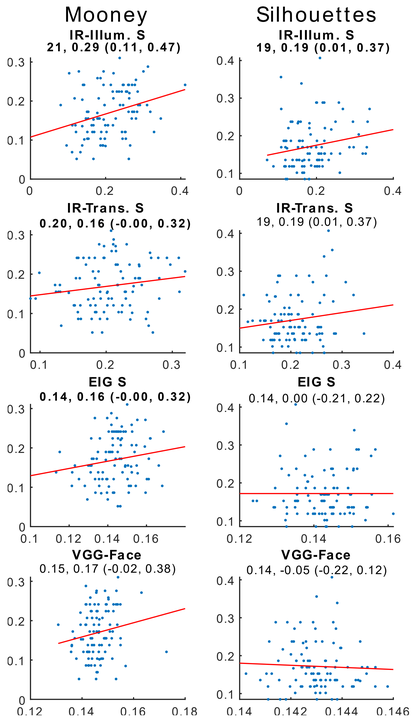
This project is supported by the Center for Brains, Minds and Machines (CBMM), funded by NSF STC award CCF-1231216

Results



Performance for all our computational models on all combinations of RGB, Mooney and Silhouette. For the inverse rendering based models we show separate performance based on shape and albedo latents, as well as combined shape-and-albedo latents. Chance success rate is 0.125.

Model (softmax $\beta = 2$)	**Experiment 1: Mooney** acc., corr. w/ humans	**Experiment 2: Silhouette** acc., corr. w/ humans
IR-Illum. C	0.21, 0.21 (-0.04, 0.45)	0.17, 0.13 (-0.08, 0.35)
IR-Illum. S	**0.21, 0.29 (0.11, 0.47)**	**0.19, 0.19 (0.01, 0.37)**
IR-Illum. A	0.21, 0.07 (-0.17, 0.34)	0.00, -0.05 (-0.24, 0.15)
IR-Trans. C	0.21, 0.06 (-0.11, 0.23)	0.18, 0.21 (-0.02, 0.41)
IR-Trans. S	**0.20, 0.16 (-0.00, 0.32)**	0.21, 0.14 (-0.08, 0.34)
IR-Trans. A	0.22,-0.02 (-0.20, 0.17)	0.14, -0.00 (-0.12, 0.17)
EIG C	0.14,-0.02 (-0.21, 0.17)	0.14, -0.03 (-0.19, 0.27)
EIG S	**0.14, 0.16 (-0.00, 0.32)**	0.14, 0.00 (-0.21, 0.22)
EIG A	0.14 -0.10 (-0.28, 0.08)	0.14, 0.02 (-0.15, 0.20)
VGG-Face	0.15, 0.17 (-0.02, 0.38)	0.14, -0.05 (-0.22, 0.12)
Human acc.	0.17, - (-,-)	0.17, - (-,-)



Model accuracies (normalized and approximately matched to human performance levels via softmax) and per-trial correlations with human error rates (95% bootstrap Cis in parentheses). Across experiments, only Inverse Rendering models that target shape(S) correlate with human percepts reliably above chance.