

Zachary Bessinger¹, Chris Stauffer², and Nathan Jacobs¹

¹University of Kentucky
Lexington, KY, USA

²Visionary Systems and Research
Boston, MA, USA

<http://wgt.csr.uky.edu>



Overview



Goal: Construct maps that show what people look like at different places in the world.

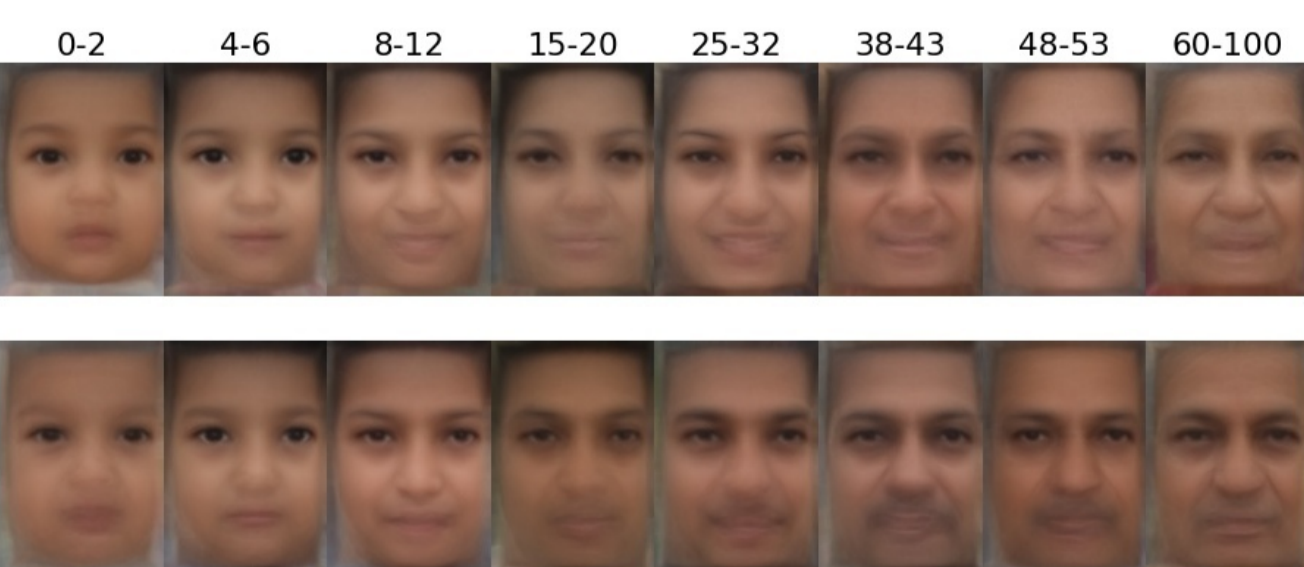
Our Approach: Use geotagged social media images by detecting faces and learning useful features to develop various mapping applications. We propose to model how age, gender, facial shape, and location affect the expected appearance.

Dataset

We constructed a face dataset using 48 million geotagged Flickr images as our primary source [1].

Dataset details:

- 2,106,468 faces
- 34 metadata attributes
- Location, pose, date uploaded, ...
- Deep CNN image features
 - Face recognition [2]
 - Age and Gender [3]



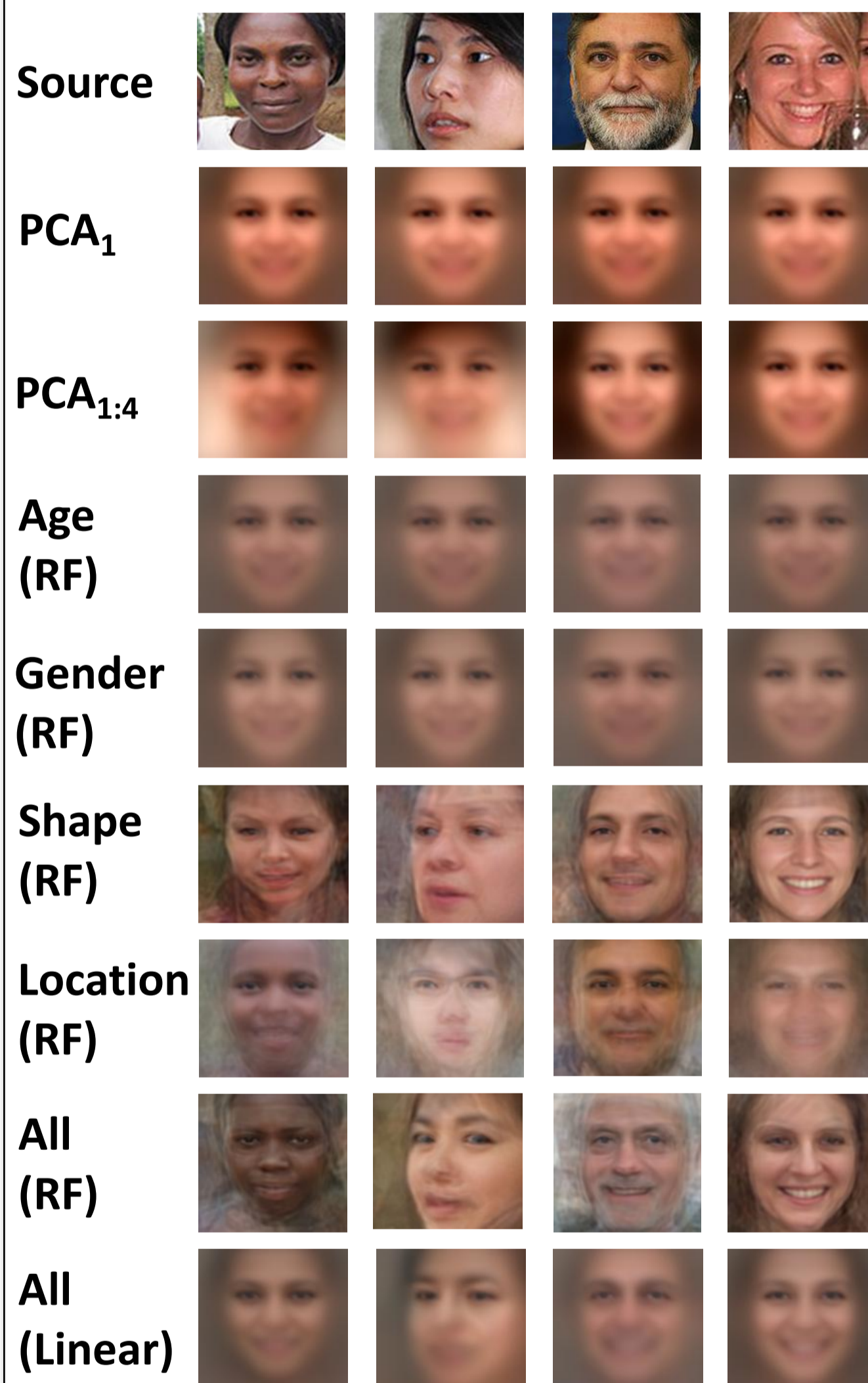
Indian females & males (top/bottom)

[1] B. Thomee, D.A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, L. Li, "YFCC100M: The New Data in Multimedia Research", Communications of ACM, 2016.
 [2] Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep face recognition." British Machine Vision Conference. Vol. 1. No. 3. 2015.
 [3] Levi, Gil, and Tal Hassner. "Age and gender classification using convolutional neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2015.

Learning Appearance

What aspects of appearance are captured by features?

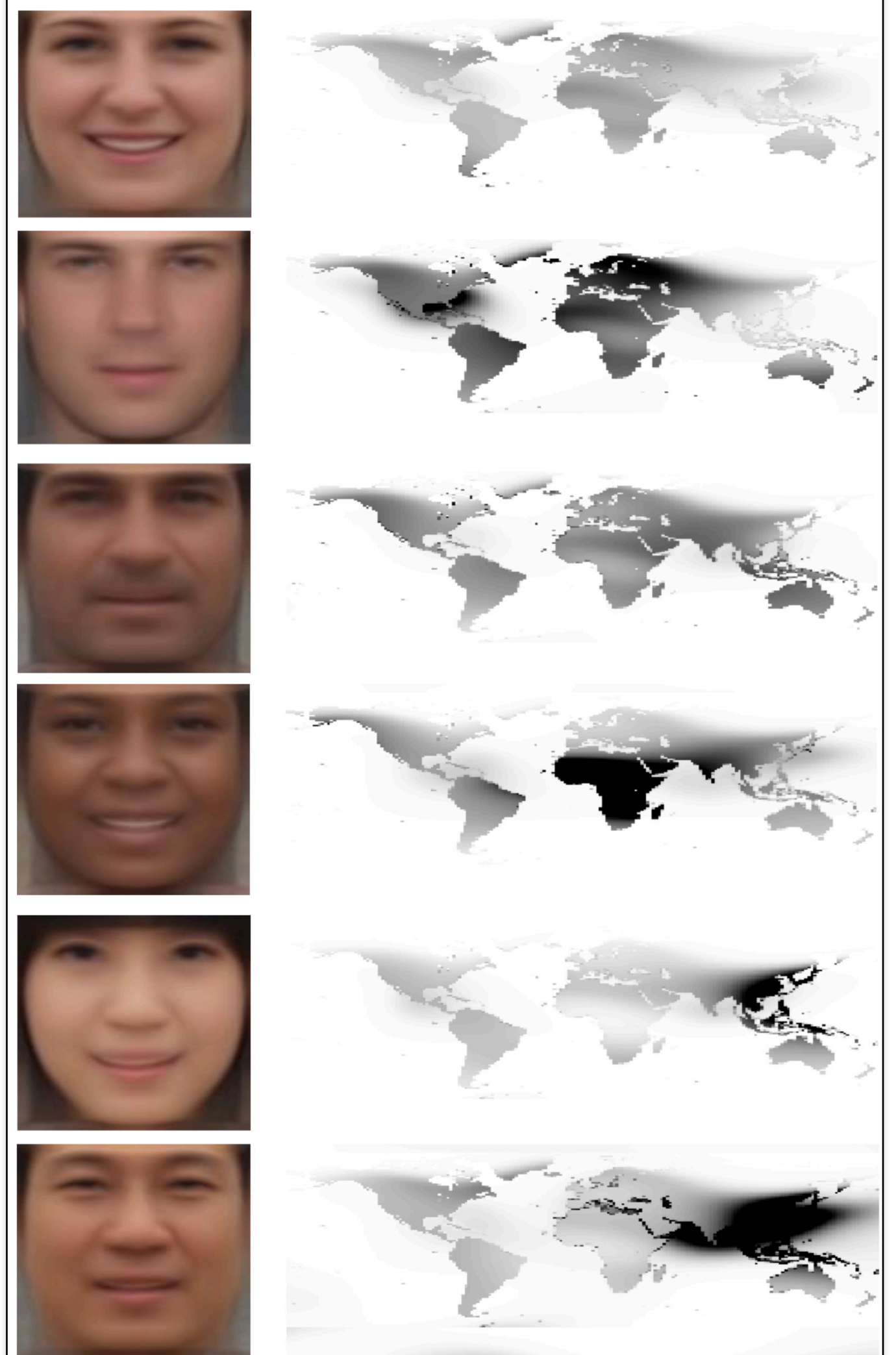
Use PCA to form a qualitatively "good" basis. Then, extract PCA components and apply linear and random forest regression, conditioned on various inputs.



Mapping Appearance

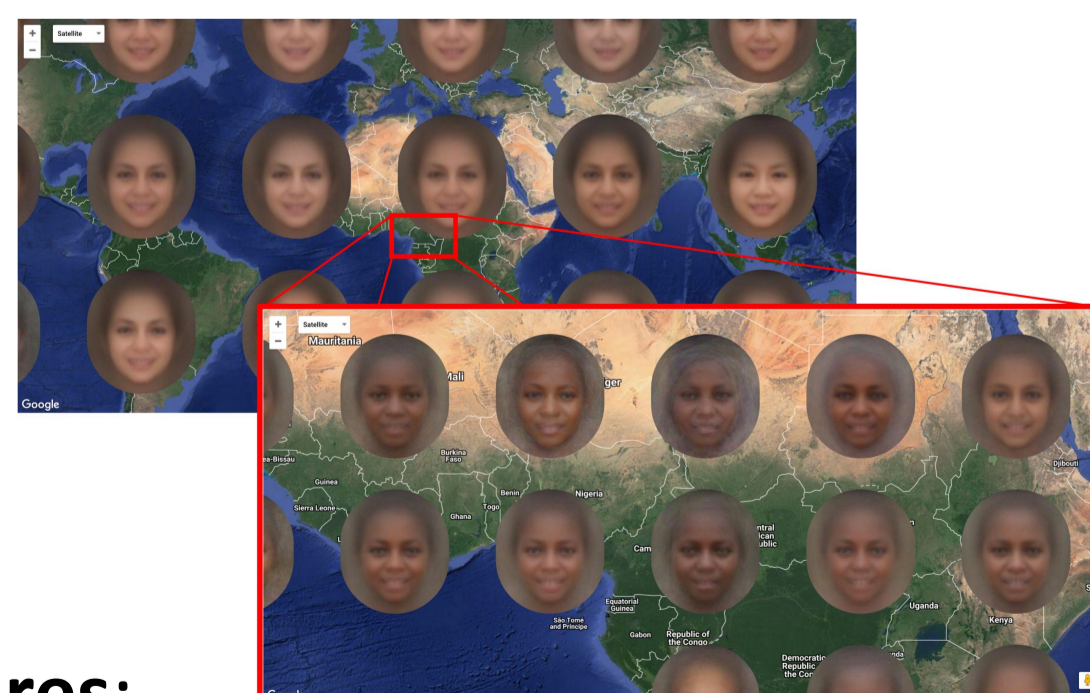
What are the most likely locations for a given appearance type?

Cluster the face recognition image features and map the distribution, $P(c = c_i | location)$, to show the spatial distribution of appearance types.



Mapping Appearance

Multiscale Web Interface



Features:

- Worldwide coverage
- Multiple zoom levels
- Interactive control of attributes

Quantifying Diversity

We use the fraction of variability, $\lambda_n = \frac{\sum_{i=1}^n \lambda_i}{\sum_{i=1}^N \lambda_i}$, where λ_i is an eigenvalue of the covariance matrix of face recognition features.

