Can We Detect Face Morphing to Prevent Identity Theft?

Sophie Nightingale, Shruti Agarwal, & Hany Farid

snightingale@berkeley.edu
shruti_agarwal@berkeley.edu
hfarid@berkeley.edu

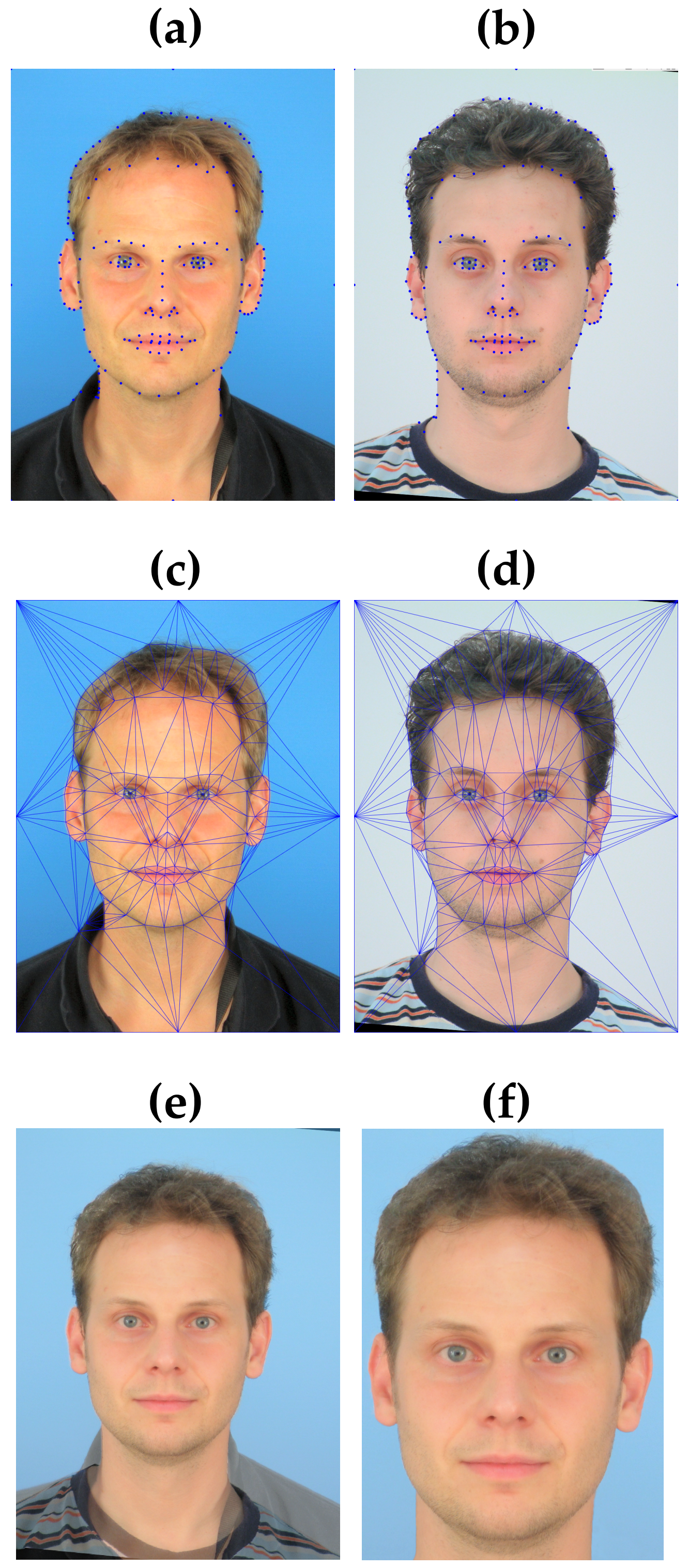
Introduction

Face morphing is a relatively new type of identity fraud that involves digitally blending images of two individuals to create a single facial image that resembles each of the original identities. People's ability to detect face morphing is limited[1,2,3], and the effect of training on performance is mixed[2,3]. We created a face dataset more representative of the type and variety of morphs fraudsters use in the real world. We then examine the efficacy of perceptual and computational detection of face morphing.

Dataset

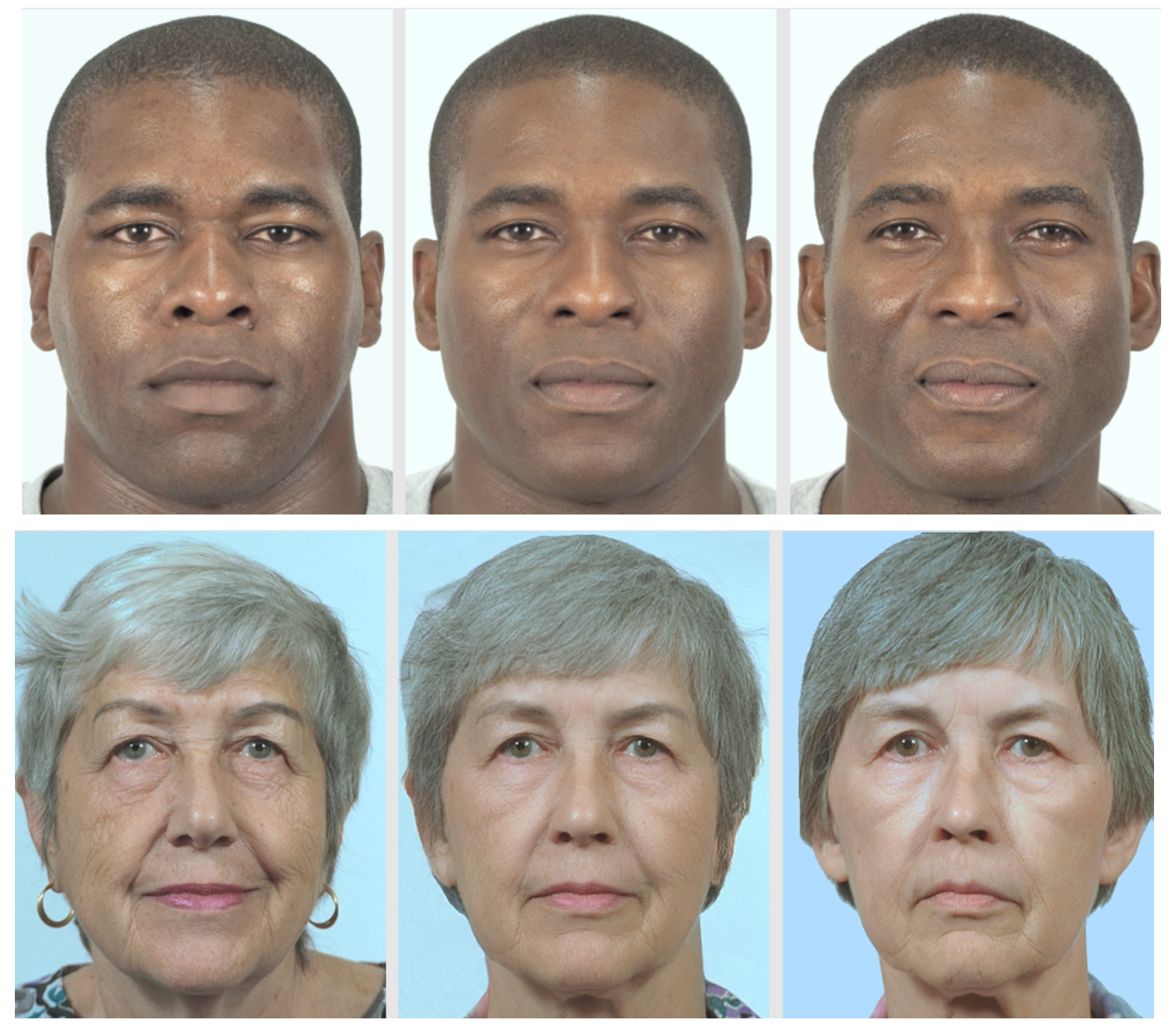
We created a high quality and diverse dataset:

- from 3,500 facial images selected 54 individuals ensuring diversity across race, gender, and age
- CNN descriptor (VGG[4]) to extract a low-dimensional, perceptually meaningful, representation of each face
- used these representations to match the 54 faces with their most similar looking counterpart
- corresponding facial landmark points were identified (a and b), aligned, and then the faces were morphed using a warping technique (c and d), to generate a mid-way morph (e), which was then cropped and manually touched-up (f)
- an analogous same individual dataset was created by selecting a new set of 54 facial images for which there were two or more distinct images of the same person

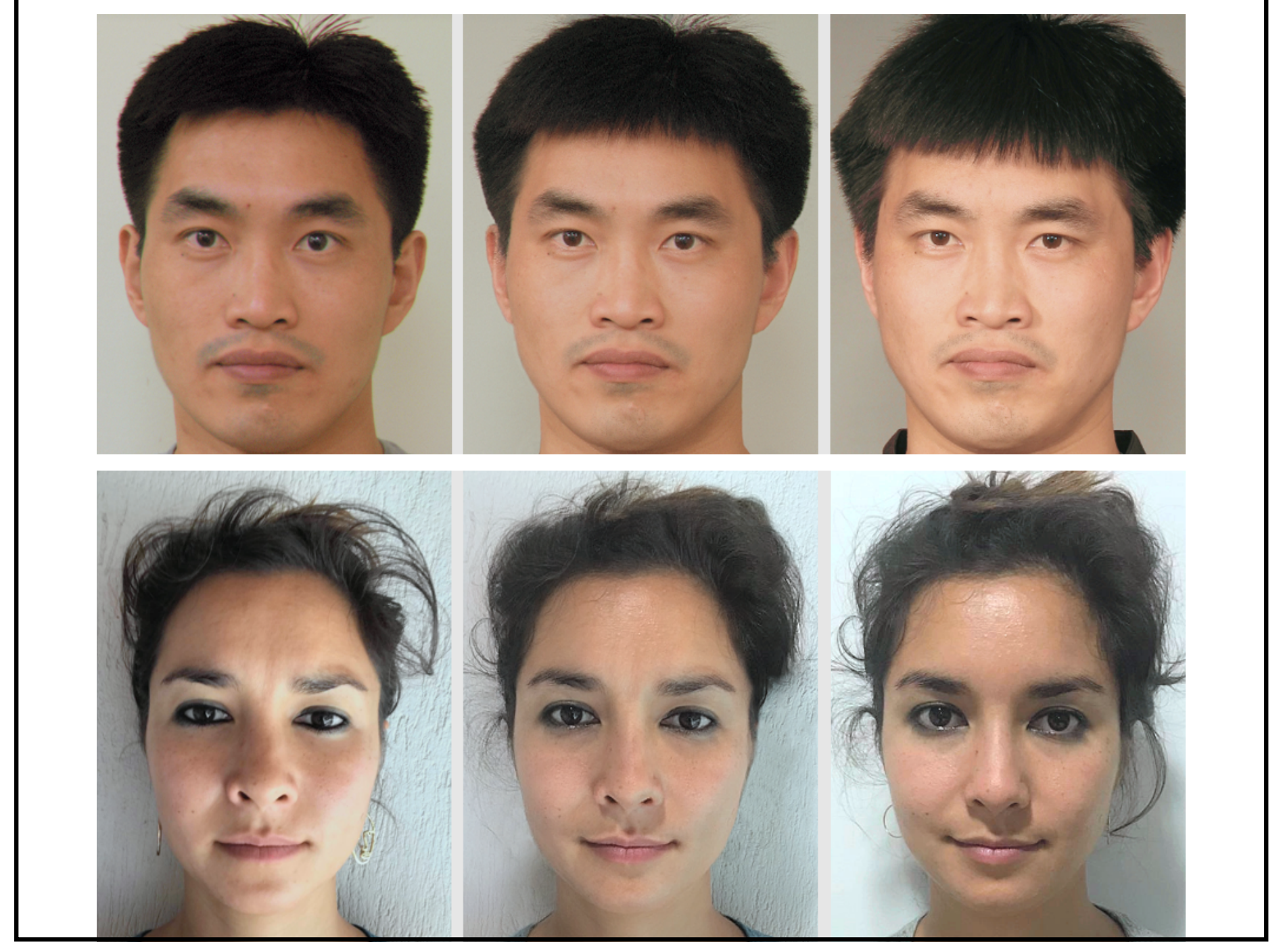


Example stimuli

Different individuals (left/right) + mid-way morph (center)



Same individuals (left/right) + mid-way morph (center)



General method

- Each of the five perceptual studies was completed online (N = 100), using a within-subject design with 108 trials
- In 1a, 1b, and 1c, on each trial, participants saw two images side-by-side and specified if they were the same person or not
- In 2a and 2b, on each trial, participants saw a single original or morphed face and specified if it was a morph or not

References

[1] Robertson, D. J., Kramer, R. S., & Burton, A. M. (2017). Fraudulent ID using face morphs: Experiments on human and automatic recognition. *PLoS One*, *12*, e0173319.

[2] Robertson, D. J., Mungall, A., Watson, D. G., Wade, K. A., Nightingale, S. J., & Butler, S. (2018). Detecting morphed passport photos: a training and individual differences approach. *Cognitive research: principles and implications*, *3*(1), 1-11.

[3] Kramer, R. S., Mireku, M. O., Flack, T. R., & Ritchie, K. L. (2019). Face morphing attacks: Investigating detection with humans and computers. *Cognitive research: principles and implications*, *4*(1), 28.

[4] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. In British Machine Vision Conference.

Results - perceptual

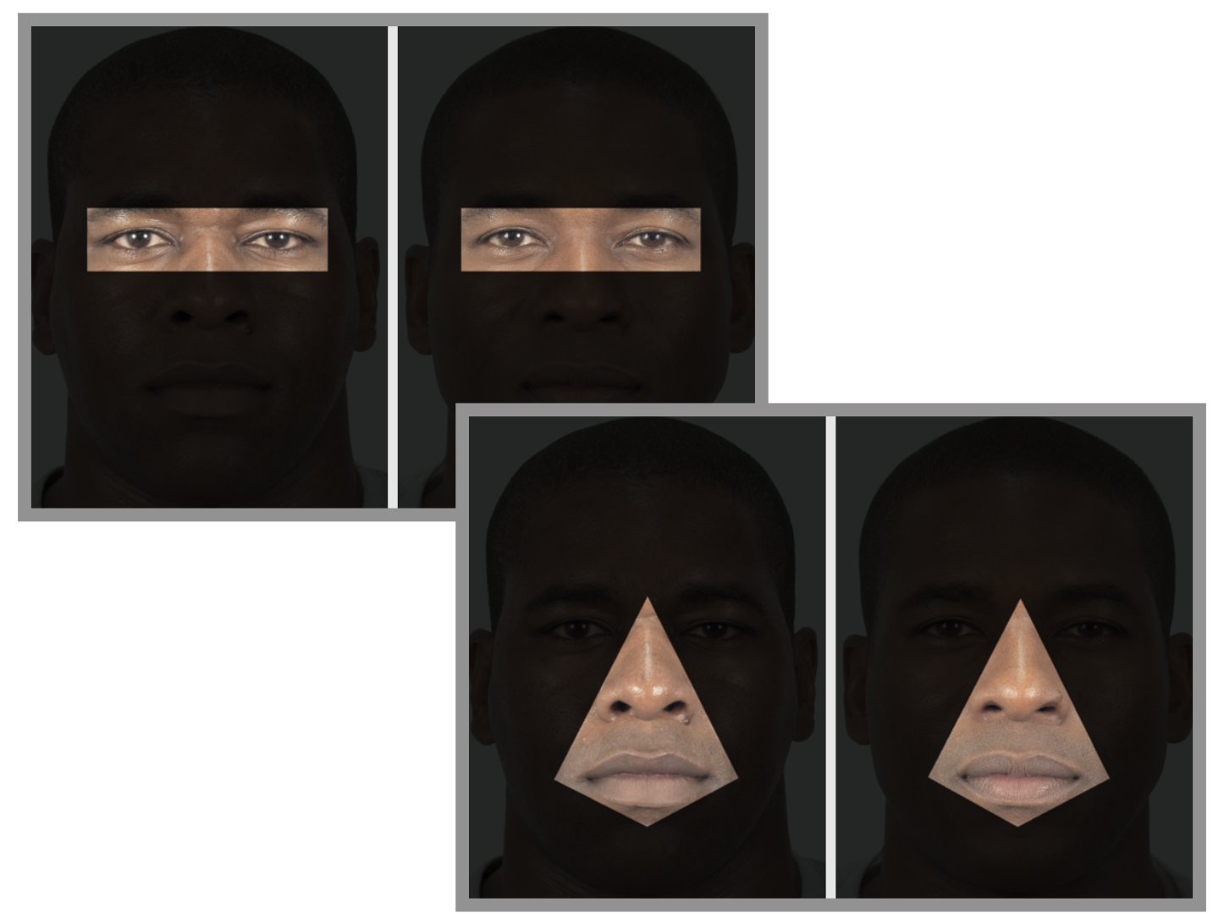
1a. One original, one morph per trial; half of the trials were different individuals + mid-way morph and half were same individuals + mid-way morph

59.2% correct,
95% CI [57.6, 60.7]
d′ = 0.68
β = 1.81

1b. Two original images per trial; half different individuals and half same individuals

80.8% correct
95% CI [78.8, 82.8]
d′ = 1.74
β = 1.03

1c. As 1a but masking to highlight eyes, nose and mouth, plus accuracy feedback after each trial

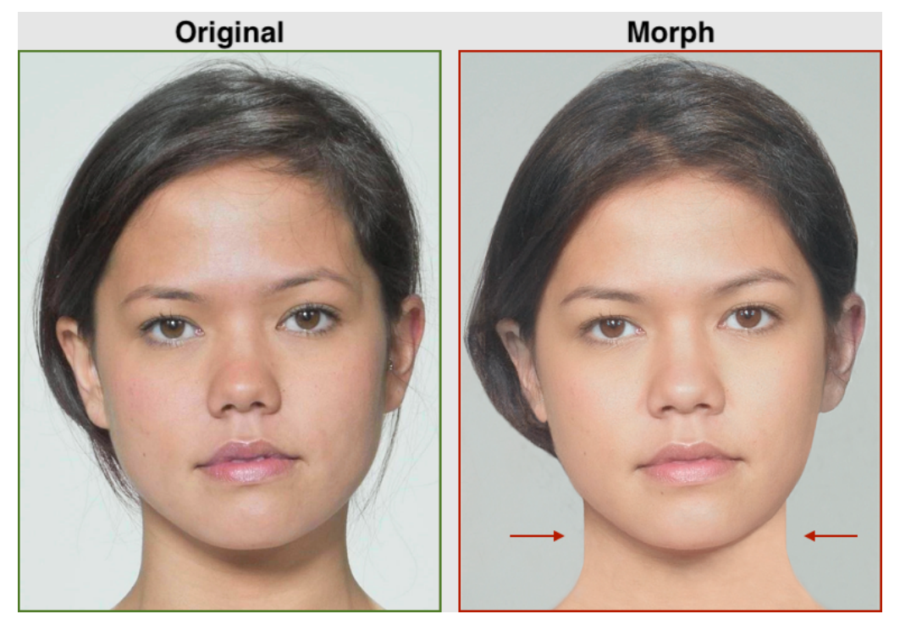


61.3% correct
95% CI [60.0, 62.6]
d′ = 0.58
β = 1.09

2a. One image per trial; half morphed, half original

54.0% correct
95% CI [52.5, 55.5]
d′ = 0.21
β = 0.98

2b. As 2a but with training to highlight morphing artifacts and accuracy feedback after each trial



60.4% correct
95% CI [58.9, 61.9]
d′ = 0.53
β = 0.92

Summary - perceptual

Participants distinguished two unfamiliar faces with reasonable accuracy (1b) but were error-prone when determining identity in morphed faces (1a), even after guidance and feedback (1c). Participants performed close to chance in detecting whether a face had been morphed (2a), and training and feedback had little effect on performance (2b).

Results - computational

Identification

Using VGG-based[4] face recognition to perform the same task as in 1a. Distances were computed between the VGG representation of the two original faces and between the VGG representation of each of the 54 pairs of different individuals and their mid-way morph. Receiver Operator Characteristic analysis revealed an area under the curve (AUC) of 0.38 (chance classifier = 0.50).

Classification

Compared to two distinct original photos of an individual, a pair of images, one of which is a mid-way morph, are more photometrically similar (higher luminance mutual information) and more geometrically similar (smaller warp-field gradient). This observation may be useful in identifying morphed faces.

Conclusion

Morph detection is a difficult task. Human participants show high error rates and the effect of training is limited. Even a state-of-the-art, machine-learning, face recognition algorithm struggles to distinguish between an individual and their morphed version.

We identify a limitation of the morphing process that can be leveraged as a way to flag suspicious images at the issuance stage using a computational classification technique.