On Face Recognition

Jitendra Malik

Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About

Increased knowledge about the ways people recognize each other may help to guide efforts to develop practical automatic face-recognition systems.

By Pawan Sinha, Benjamin Balas, Yuri Ostrovsky, and Richard Russell

Recognition as a function of available spatial resolution

Result 1: Humans can recognize familiar faces in very low-resolution images.

Result 2: The ability to tolerate degradations increases with familiarity.

Result 3: High-frequency information by itself is insufficient for good face recognition performance.

The nature of processing: Piecemeal versus holistic

Result 4: Facial features are processed holistically.

Result 5: Of the different facial features, eyebrows are among the most important for recognition.

Result 6: The important configural relationships appear to be independent across the width and height dimensions.

The nature of cues used: Pigmentation, shape and motion

Result 7: Face-shape appears to be encoded in a slightly caricatured manner.

Result 8: Prolonged face viewing can lead to high-level aftereffects, which suggest prototype-based encoding.

Result 9: Pigmentation cues are at least as important as shape cues.

Result 10: Color cues play a significant role, especially when shape cues are degraded.

Result 11: Contrast polarity inversion dramatically impairs recognition performance, possibly due to compromised ability to use pigmentation cues.

Result 12: Illumination changes influence generalization.

Result 13: View-generalization appears to be mediated by temporal association.

Result 14: Motion of faces appears to facilitate subsequent recognition.

Developmental progression

Result 15: The visual system starts with a rudimentary preference for face-like patterns.

Result 16: The visual system progresses from a piecemeal to a holistic strategy over the first several years of life.

Neural underpinnings

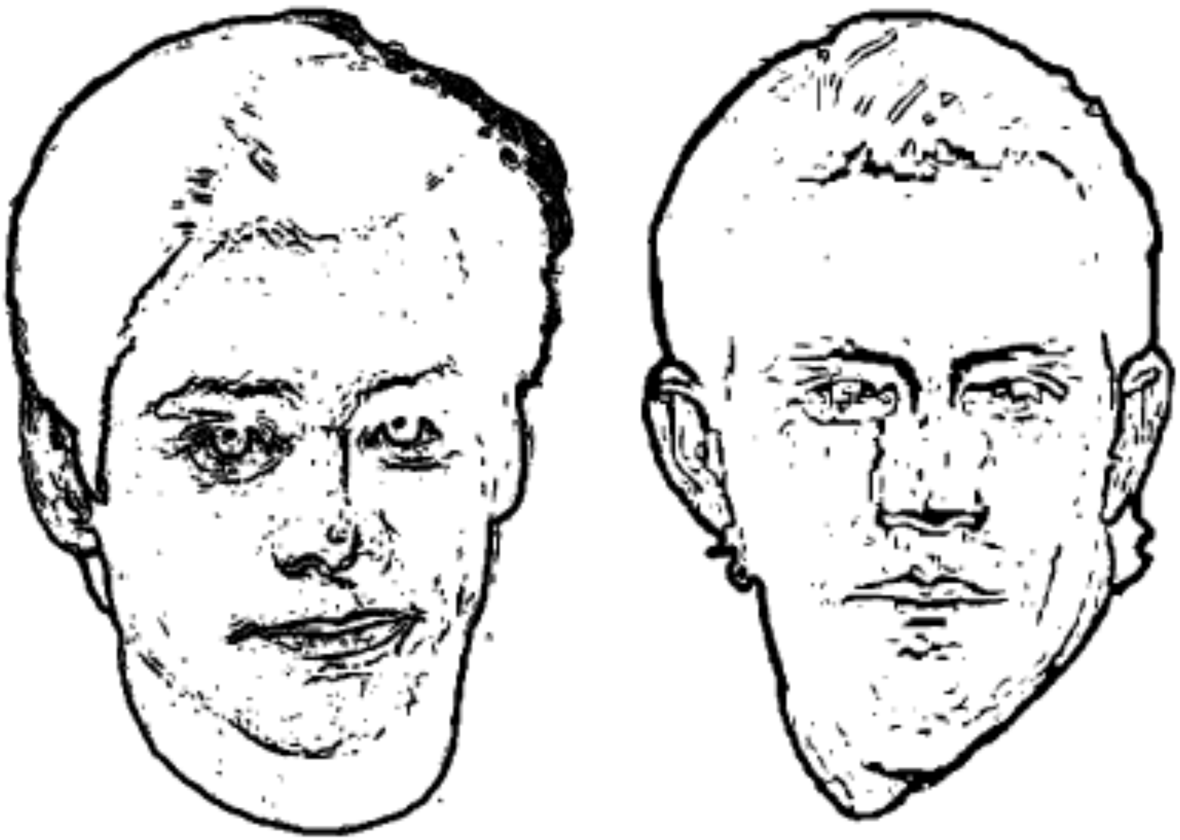
Result 17: The human visual system appears to devote specialized neural resources for face perception.

Result 18: Latency of responses to faces in inferotemporal (IT) cortex is about 120 ms, suggesting a largely feedforward computation.

Result 19: Facial identity and expression might be processed by separate systems.


1
2
3
4
5
6
7
8
9
10
11
12

Fig. 1. *Unlike current machine-based systems, human observers are able to handle significant degradations in face images. For instance, subjects are able to recognize more than half of all familiar faces shown to them at the resolution depicted here. Individuals shown in order are: Michael Jordan, Woody Allen, Goldie Hawn, Bill Clinton, Tom Hanks, Saddam Hussein, Elvis Presley, Jay Leno, Dustin Hoffman, Prince Charles, Cher, and Richard Nixon.*



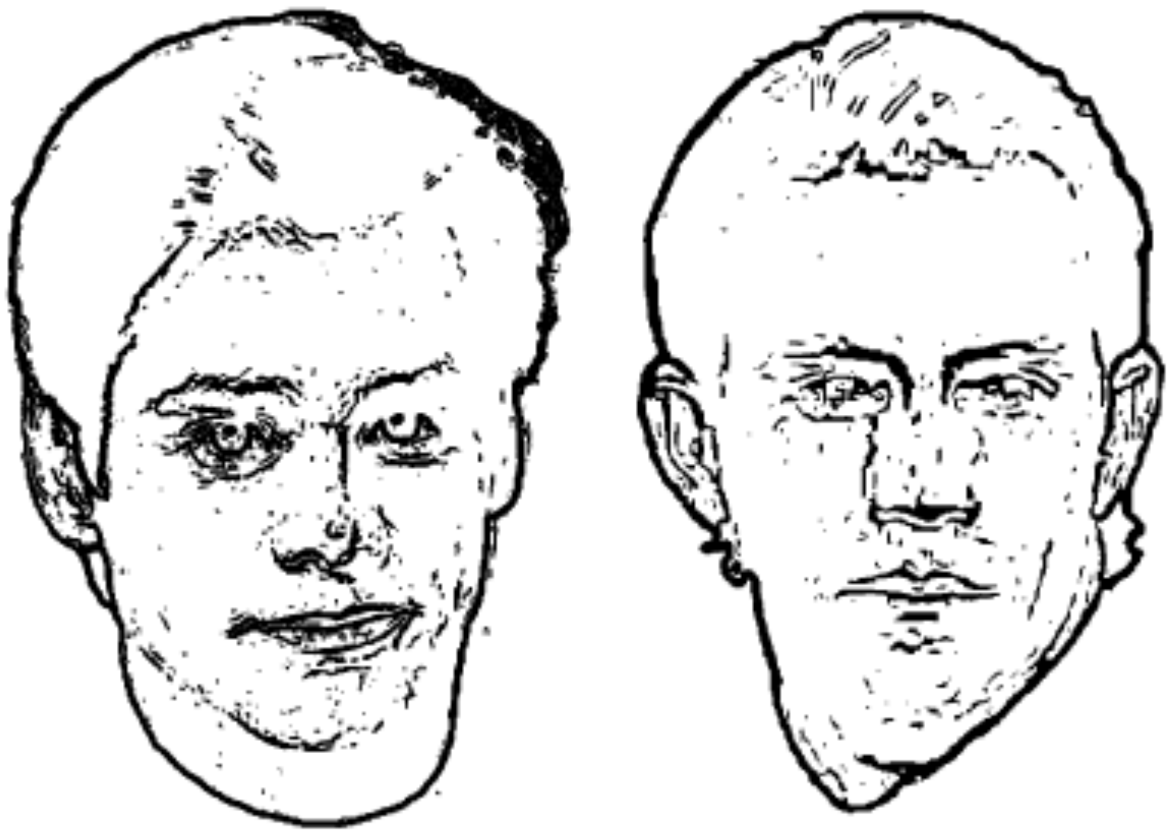


Fig. 3. ***Images which contain exclusively contour information are very difficult to recognize, suggesting that high-spatial frequency information, by itself, is not an adequate cue for human face recognition processes. Shown here are Jim Carrey (left) and Kevin Costner.***

Fig. 4. *Try to name the famous faces depicted in the two halves of the left image. Now try the right image. Subjects find it much more difficult to perform this task when the halves are aligned (left) compared to misaligned halves (right), presumably because holistic processing interacts (and in this case, interferes) with feature-based processing. The two individuals shown here are Woody Allen and Oprah Winfrey.*

Fig. 6. *Even drastic compressions of faces do not render them unrecognizable. Here, celebrity faces have been compressed to 25% of their original width. Yet, recognition performance with this set is the same as that obtained with the original faces.*

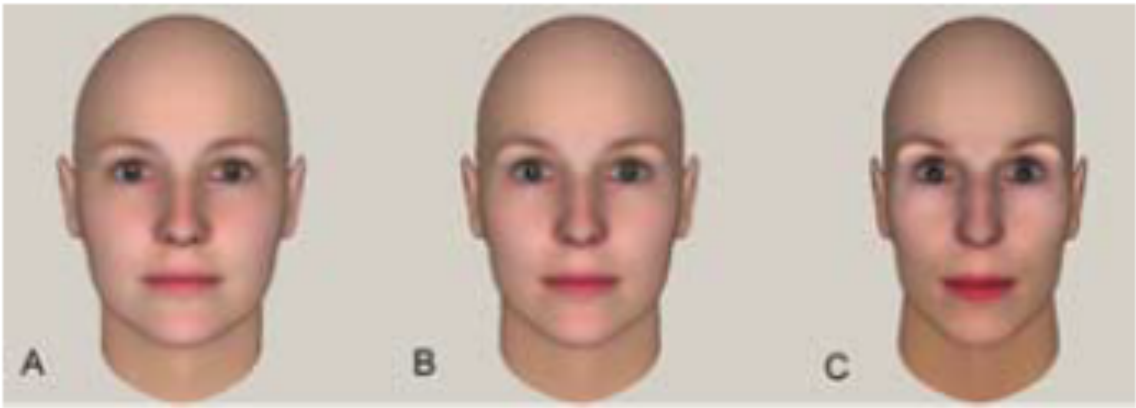


Fig. 7. *Example of a face caricature. (A) Average female face for a particular face population is displayed, as well as a (B) ''veridical'' image of an exemplar face. (C) We create a caricatured version of the exemplar by moving away from the norm, thus exaggerating differences between the average face and the exemplar. Result is a face with ''caricatured'' shape and pigmentation. Such caricatures are recognized as well or better than veridical images.*

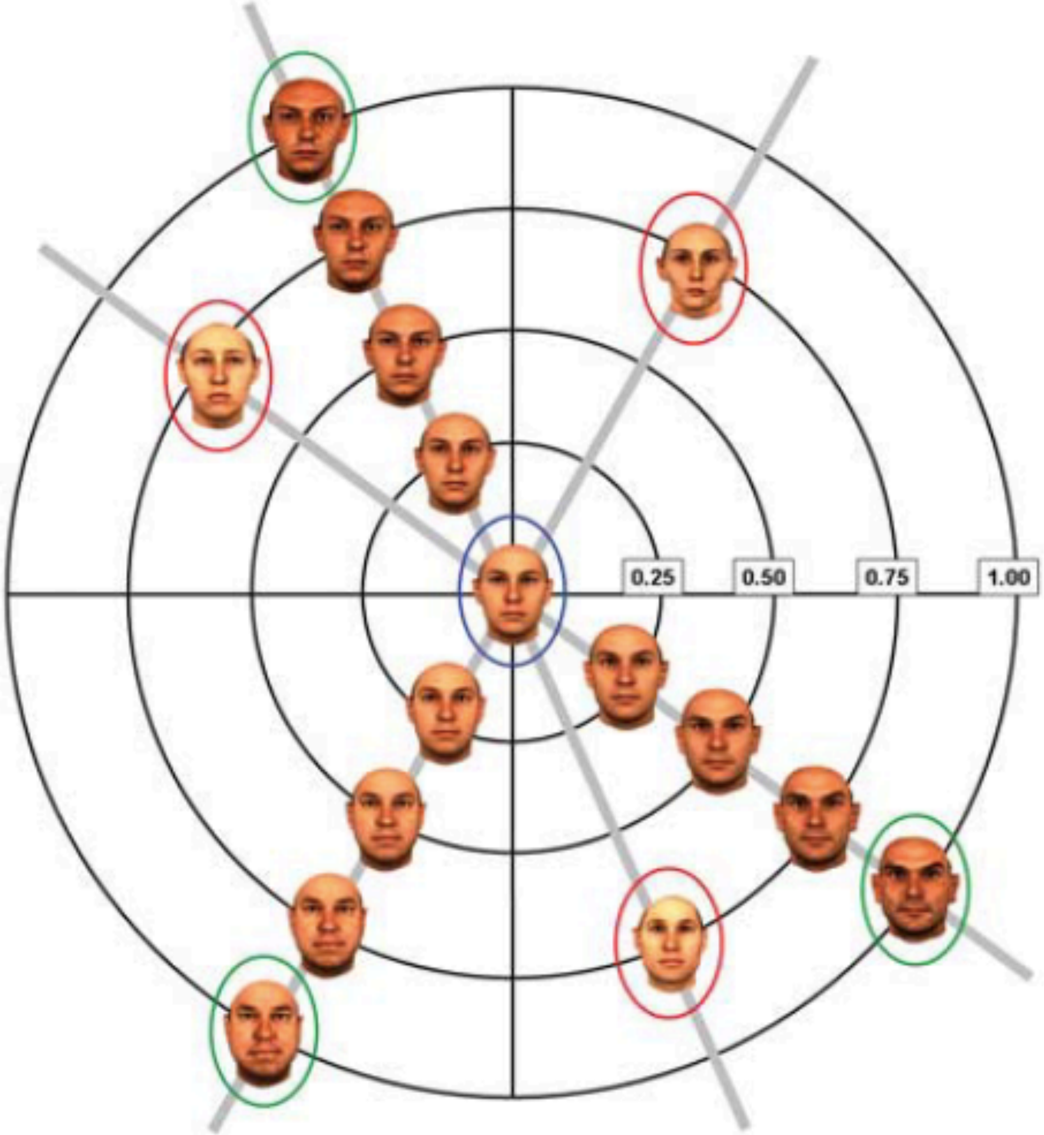


Fig. 8. *Faces and their associated "anti-faces" in a schematic face space. Prolonged viewing of a face within a green circle will cause the central face to be misidentified as the individual within the red circle along the same "identity trajectory" (from [45]).*

Fig. 9. *Faces in the bottom row are all images of laser-scanned faces. They differ from one another in terms of both shape and pigmentation. Faces in the middle row differ from one another in terms of their pigmentation but not their shape, while faces in the top row differ from one another in terms of their shape but not their pigmentation. From the fact that the faces in either the top or middle row do not look the same as each other, it is evident that both shape and pigmentation cues play a role in facial identity.*

Fig. 11. ***Image contains several well-known singers, whose likenesses would be easily recognizable to many readers of this publication. However, when presented in negative contrast, it is difficult, if not impossible, to recognize them. (Photographed during the recording of "We Are the World" song.)***

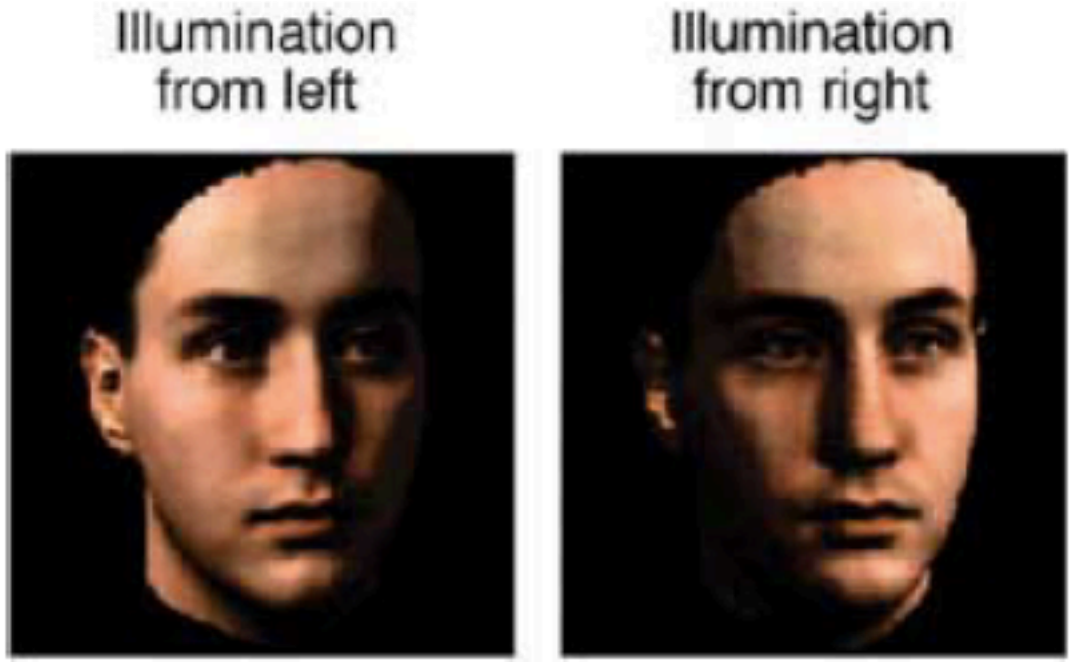
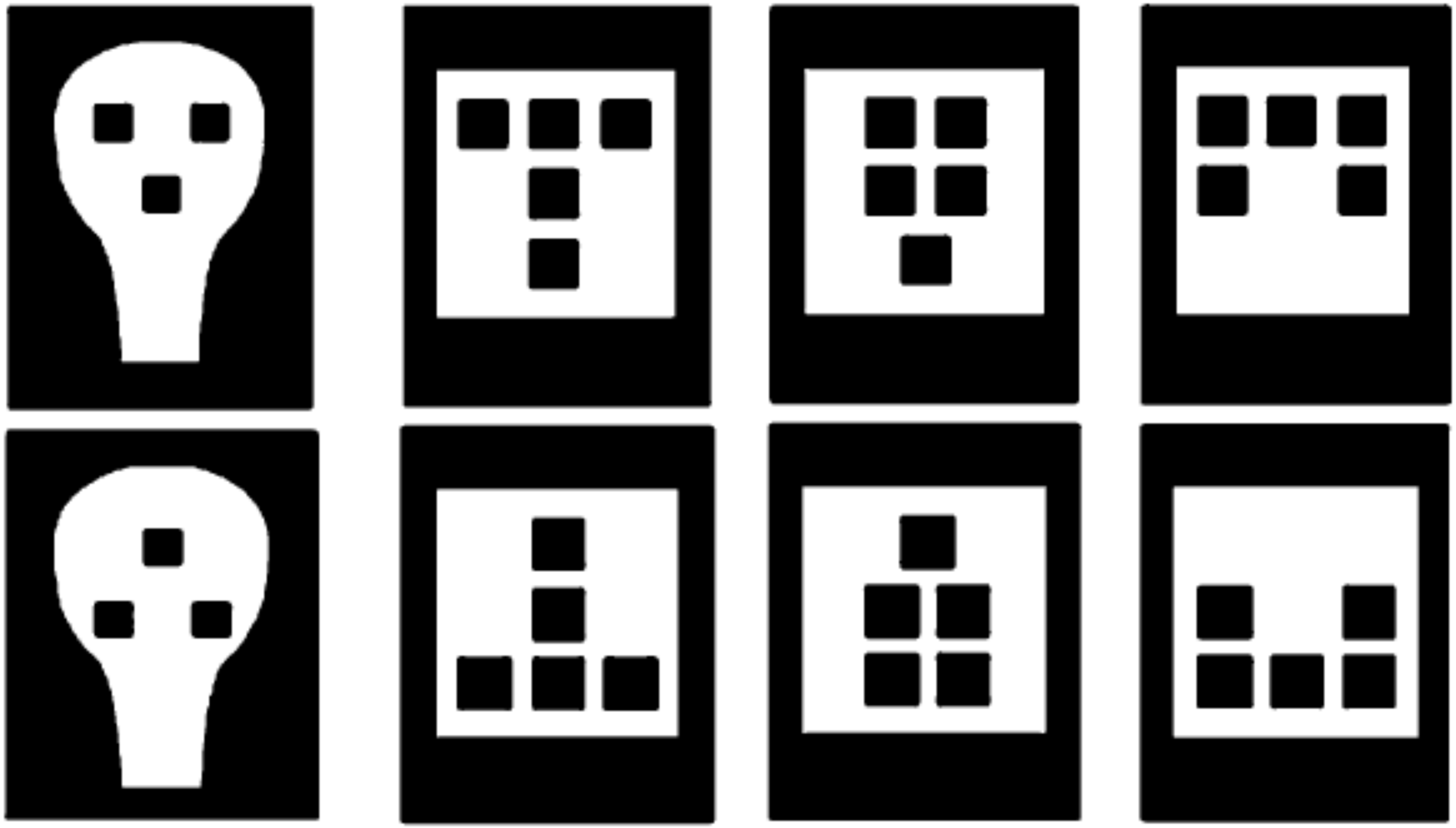


Fig. 12. ***Stimuli from Braje et al. [2]. These two images demonstrate the kind of lighting used in this experiment. After being shown an image like the one on the left, subjects were well above chance at determining whether a subsequently presented image such as the one on the right represented the same or a different individual (in this case the same).***



(a) (b)

Fig. 15. ***(a) Newborns preferentially orient their gaze to face-like pattern on top, rather than one shown on bottom, suggesting some innately specified representation for faces (from [36]). (b) As a counterpoint to idea of innate preferences for faces, Simion et al. [73] have shown that newborns consistently prefer top-heavy patterns***

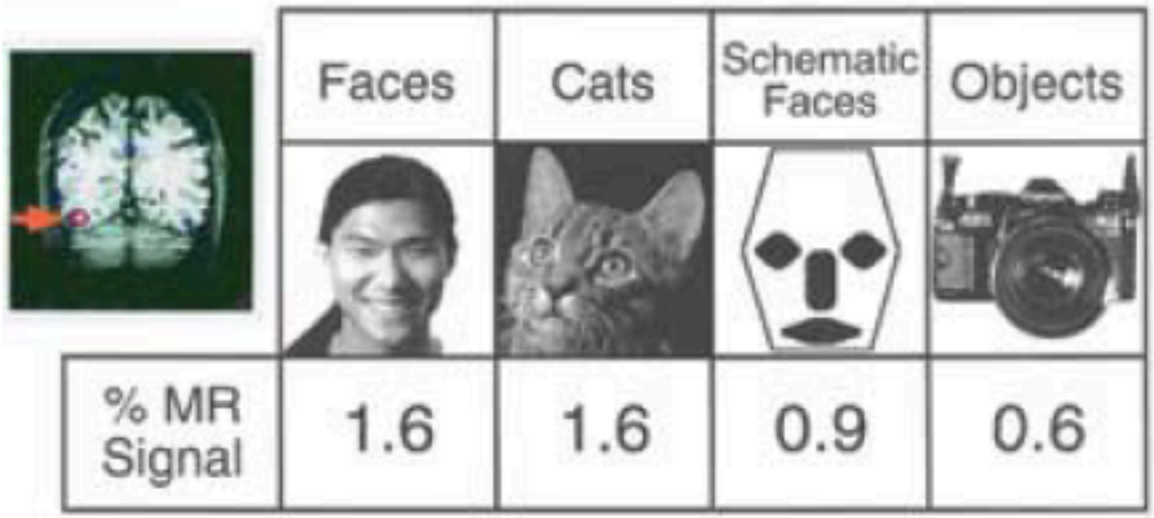


Fig. 17. *Upper left, an example of FFA in one subject, showing right-hemisphere lateralization. Also included here are example stimuli from Tong et al. [80], together with amount of percent signal change observed in FFA for each type of image. Photographs of human and animal faces elicit strong responses, while schematic faces and objects do not. This response profile may place important constraints on the selectivity and generality of artificial recognition systems.*

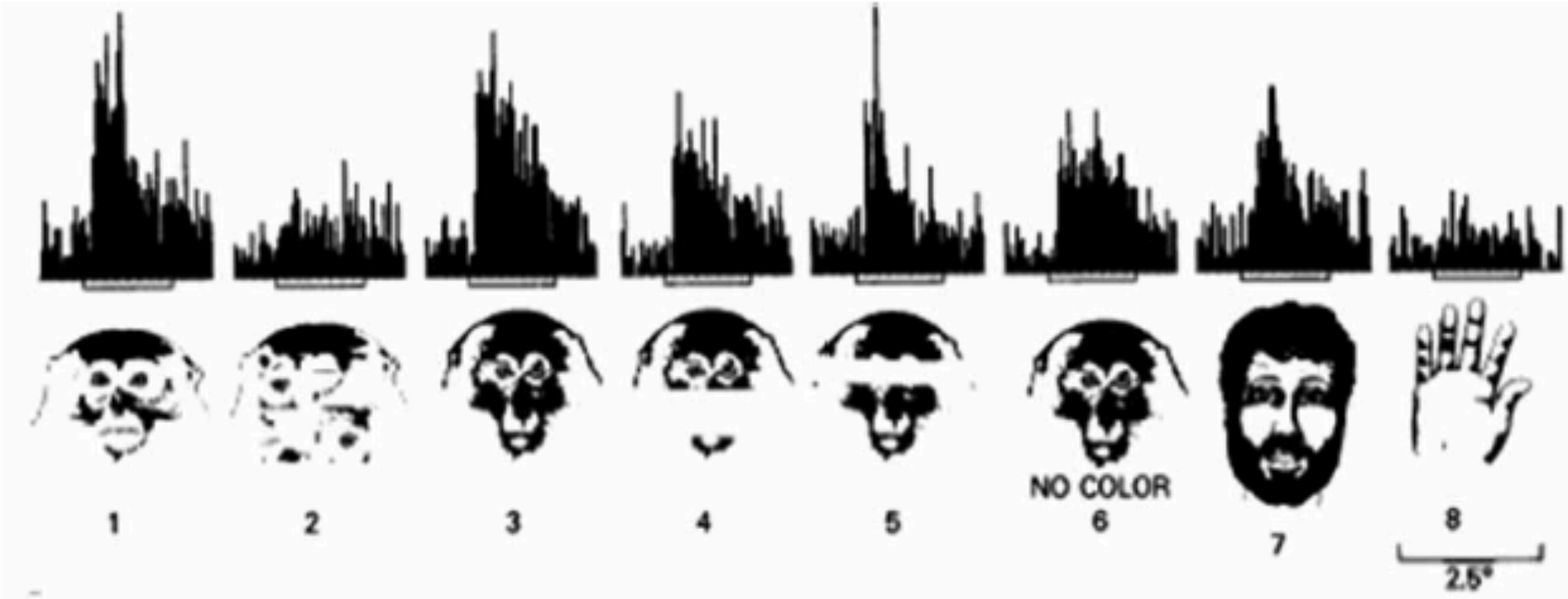


Fig. 18. *Example of a monkey IT cell's responses to variations on a face stimulus (from Desimone et al. [17]). Response is robust to many degradations of the primate face (save for scrambling) and also responds very well to a human face. Lack of a response to the hand indicates that this cell is not just interested in body parts, but is specific to faces. Cells in IT cortex can produce responses such as these with a latency of about 120 ms.*

The Thatcher Illusion (Thompson 1980)

DeepFace: Closing the Gap to Human-Level Performance in Face Verification

Yaniv Taigman Ming Yang Marc'Aurelio Ranzato Lior Wolf

Facebook AI Research
Menlo Park, CA, USA
{yaniv, mingyang, ranzato}@fb.com

Tel Aviv University
Tel Aviv, Israel
wolf@cs.tau.ac.il

Key Ideas

- “Frontalize” the face, so that the features can be computed in standard position
- Train a neural network to learn the features that are “identity specific” by treating examples of the same individual’s face as instances of the same category
- Use “local” receptive fields in higher layers, no longer maintaining the shift-invariant architecture of lower layers

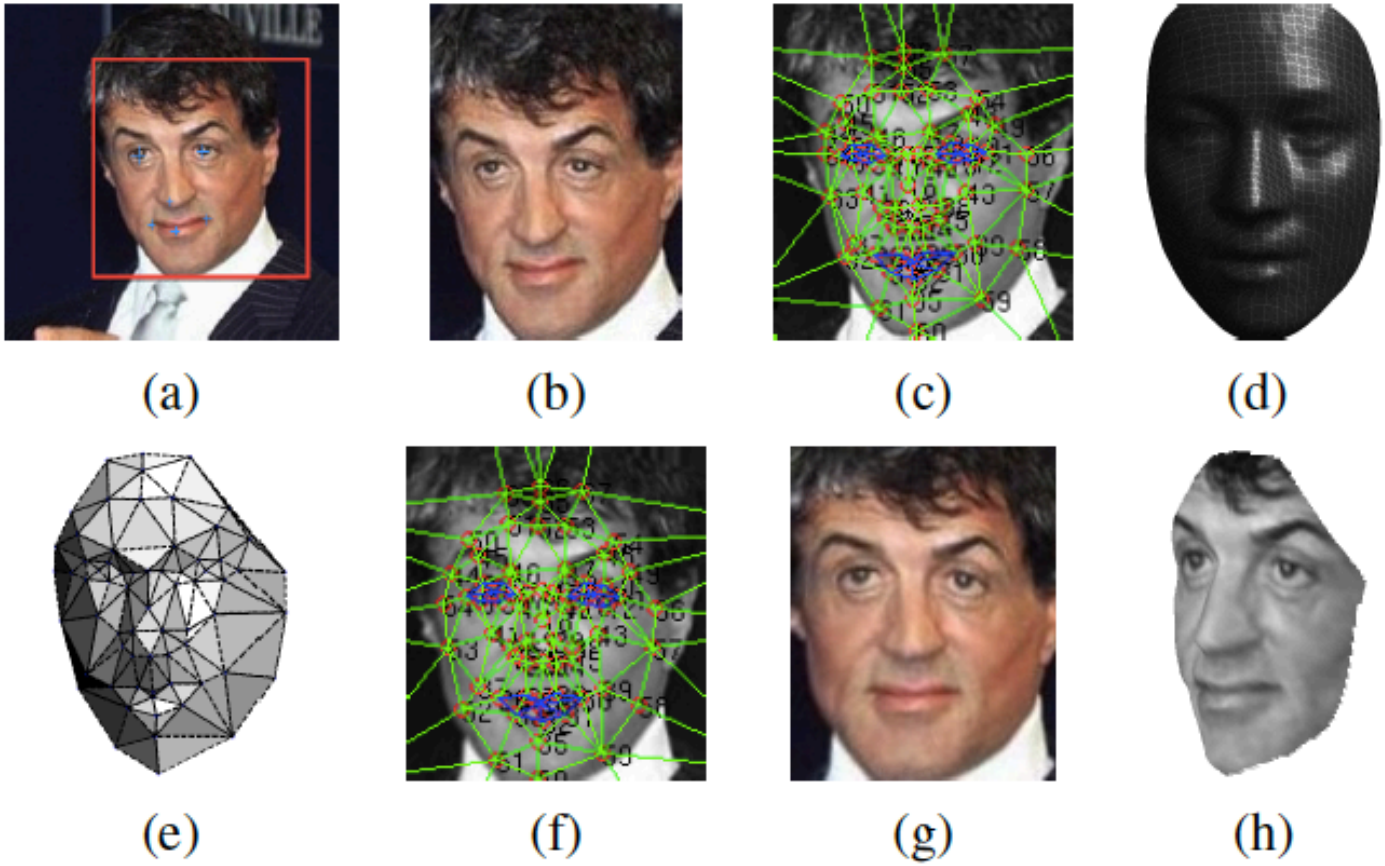
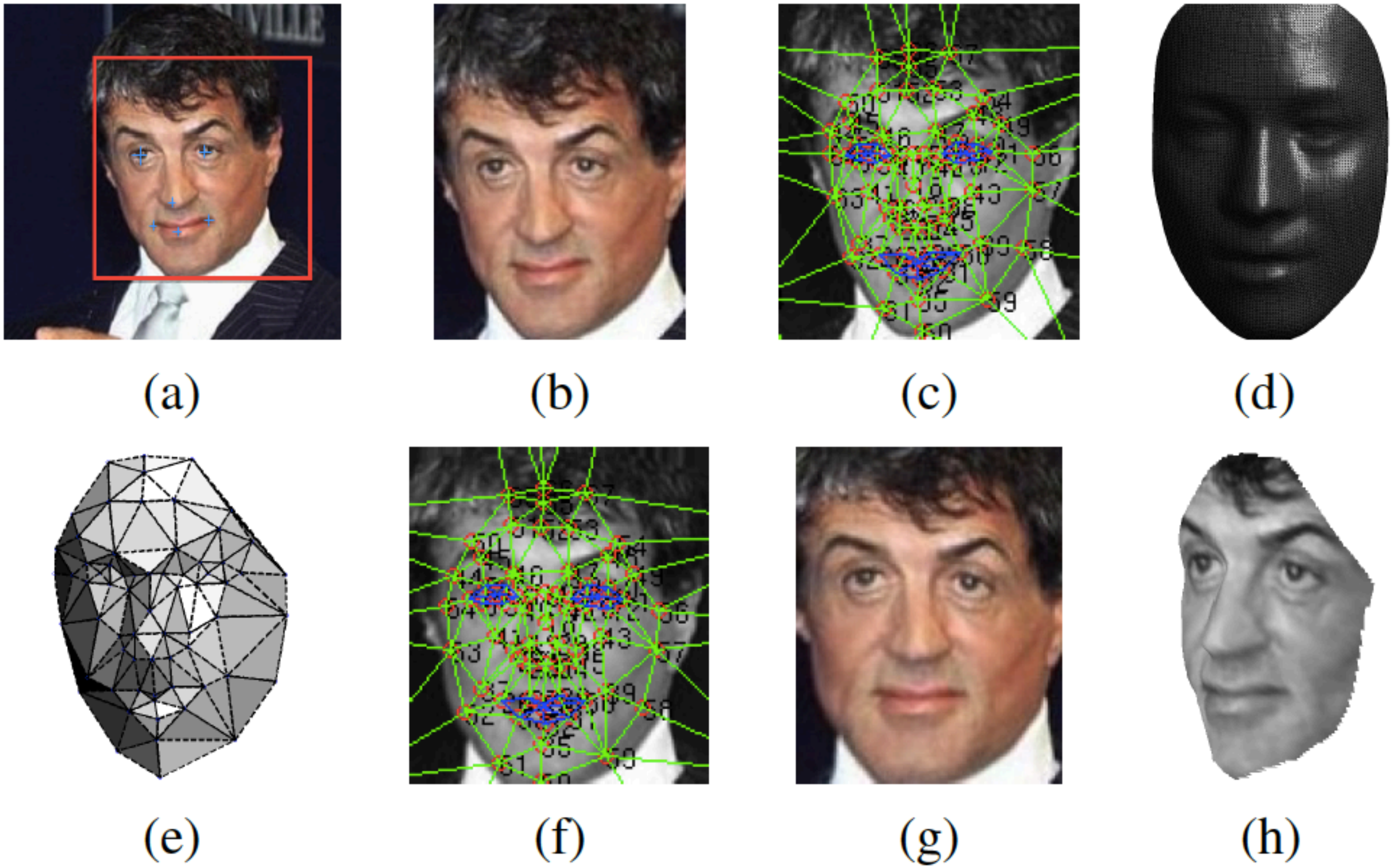


Figure 1. **Alignment pipeline.** (a) The detected face, with 6 initial fiducial points. (b) The induced 2D-aligned crop. (c) 67 fiducial points on the 2D-aligned crop with their corresponding Delaunay triangulation, we added triangles on the contour to avoid discontinuities. (d) The reference 3D shape transformed to the 2D-aligned crop image-plane. (e) Triangle visibility w.r.t. to the fitted 3D-2D camera; darker triangles are less visible. (f) The 67 fiducial points induced by the 3D model that are used to direct the piece-wise affine warpping. (g) The final frontalized crop. (h) A new view generated by the 3D model (not used in this paper).


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)

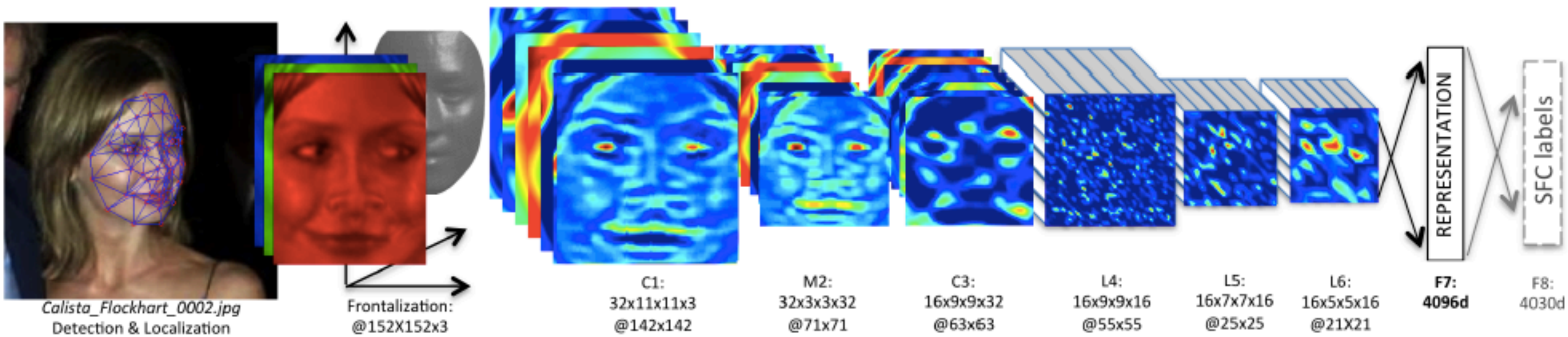


Figure 2. **Outline of the *DeepFace* architecture**. A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate feature maps produced at each layer. The net includes more than 120 million parameters, where more than 95% come from the local and fully connected layers.

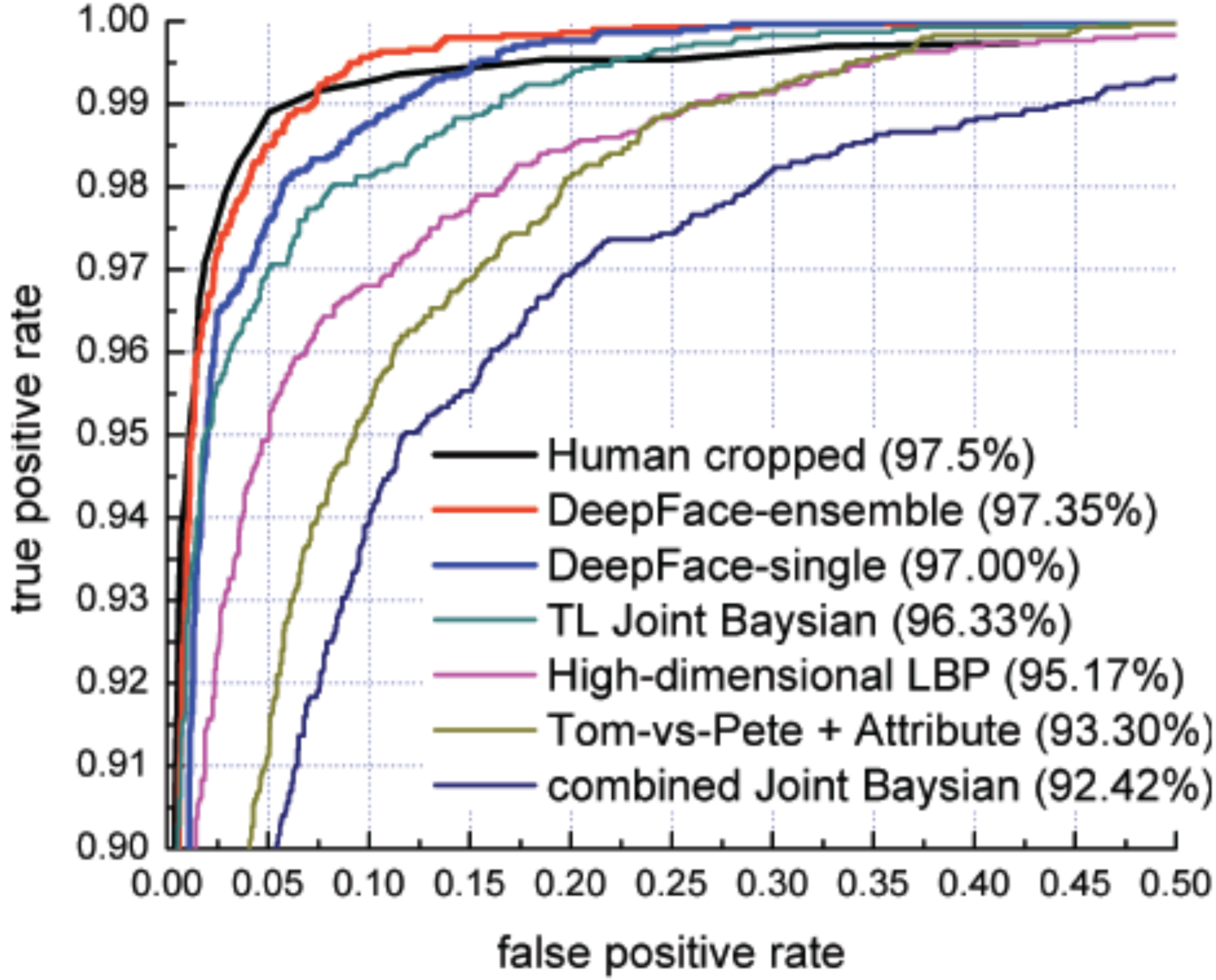


Figure 3. The ROC curves on the *LFW* dataset. Best viewed in color.