

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

ABSTRACT | A key goal of computer vision researchers is to create automated face recognition systems that can equal, and eventually surpass, human performance. To this end, it is imperative that computational researchers know of the key findings from experimental studies of face recognition by humans. These findings provide insights into the nature of cues that the human visual system relies upon for achieving its impressive performance and serve as the building blocks for efforts to artificially emulate these abilities. In this paper, we present what we believe are 19 basic results, with implications for the design of computational systems. Each result is described briefly and appropriate pointers are provided to permit an in-depth study of any particular result.

KEYWORDS | Benchmarks; configuration; face pigmentation; face recognition; human vision; neural correlates; resolution; visual development

I. INTRODUCTION

Notwithstanding the extensive research effort that has gone into computational face recognition algorithms, we have yet to see a system that can be deployed effectively in an unconstrained setting, with all of the attendant variability in imaging parameters such as sensor noise, viewing distance, and illumination. The only system that

does seem to work well in the face of these challenges is the human visual system. It makes eminent sense, therefore, to attempt to understand the strategies this biological system employs, as a first step towards eventually translating them into machine-based algorithms. With this objective in mind, we review here 19 important results regarding face recognition by humans. While these observations do not constitute a coherent theory of face recognition in human vision (we simply do not have all the pieces yet to construct such a theory), they do provide useful hints and constraints for one. We believe that for this reason, they are likely to be useful to computer vision researchers in guiding their ongoing efforts. Of course, the success of machine vision systems is not dependent on a slavish imitation of their biological counterparts. Insights into the functioning of the latter serve primarily as potentially fruitful starting points for computational investigations.

We have endeavored to bring together in one place several diverse results to be able to provide the reader a fairly comprehensive picture of our current understanding regarding how humans recognize faces. Each of the results is briefly described and, whenever possible, accompanied by its implications for computer vision. While the descriptions here are not extensive for reasons of space, we have provided relevant pointers to the literature for a more in-depth study. The results are organized along the following broad themes.

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.

Manuscript received July 12, 2005; revised March 15, 2006.

P. Sinha, B. Balas, and Y. Ostrovsky are with the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: psinha@mit.edu; bjbaldas@mit.edu; yostr@mit.edu).

R. Russell is with the Department of Psychology, Harvard University, Cambridge, MA 02138 USA (e-mail: russell@fas.harvard.edu).

Digital Object Identifier: 10.1109/JPROC.2006.884093

- 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.

A. Recognition as a Function of Available Spatial Resolution

1) *Result 1: Humans Can Recognize Familiar Faces in Very Low-Resolution Images:* Progressive improvements in cam-

era resolutions provide ever-greater temptation to use increasing amounts of detail in face representations in machine vision systems. Higher image resolutions allow recognition systems to discriminate between individuals on the basis of fine differences in their facial features. The advent of iris-based biometric systems is a case in point. However, the problem that such details-based schemes often have to contend with is that high-resolution images are not always available. This is particularly true in situations where individuals have to be recognized at a distance. In order to design systems more robust against image degradations, we can turn to the human visual system for inspiration. Everyday, we are confronted with the task of face identification at a distance and must extract the critical information from the resulting low-resolution images. Precisely how does face identification performance change as a function of image resolution? Pioneering work on face recognition with low-resolution imagery was done by Harmon and Julesz [30], [31]. Working with block averaged images of familiar faces, they found high recognition accuracies even with images containing just 16×16 blocks. Yip and Sinha [89] found that subjects could recognize more than half of an unprimed set of familiar faces that had been blurred to have equivalent image resolutions of merely 7×10 pixels (see Fig. 1), and recognition performance reached ceiling level at a resolution of 19×27 pixels. While the remarkable tolerance of the human visual system to resolution reduction is now indisputable, we do not have a clear idea of exactly how this is accomplished. At the very least, this result demonstrates that fine featural details are not necessary to obtain good face recognition performance. Furthermore, given the indistinctness of the individual features at low resolutions, it appears likely that diagnosticity resides in their overall configuration. However, precisely which aspects of this configuration are important, and how we can computationally encode them, are open questions.

2) *Result 2: The Ability to Tolerate Degradations Increases With Familiarity:* In trying to uncover the mechanisms underlying the human ability to recognize highly degraded face images, we might wonder whether this is the result of some general purpose compensatory processes, i.e., a biological instantiation of model-free “super resolution.” However, the story appears to be more complicated. The ability to handle degradations increases dramatically with amount of familiarity. Bruce et al. [9] demonstrated observers’ poor performance on the task of matching two different photographs of an unfamiliar person. Burton et al. [10] have shown that observers’ recognition performance with low-quality surveillance video is much better when the individuals pictured are familiar colleagues, rather than those with whom the observers have interacted infrequently. Additionally, body structure and gait information are much less useful for identification than facial information,



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.

even though the effective resolution in that region is very limited. Recognition performance changes only slightly after obscuring the gait or body, but is affected dramatically when the face is hidden, as illustrated in Fig. 2. This does not appear to be a skill that can be acquired through general experience; even police officers with extensive forensic experience perform poorly unless they are familiar with the target individuals. The fundamental question this finding, and others like it [49], [66], bring up is the following: How does the facial representation and matching strategy used by the visual system change with increasing familiarity, so as to yield greater tolerance to degradations? We do not yet know exactly what aspect of the increased experience with a given individual leads to an increase in the robustness of the encoding; is it the greater number of views seen or is the robustness an epiphenomenon related to some biological limitations such as slow memory consolidation rates? Notwithstanding our limited understanding, some implications for computer vision are already evident. In

considering which aspects of human performance to take as benchmarks, we ought to draw a distinction between familiar and unfamiliar face recognition. The latter may end up being a much more modest goal than the former and might constitute a false goal towards which to strive. The appropriate benchmark for evaluating machine-based face recognition systems is human performance with familiar faces.

3) *Result 3: High-Frequency Information by Itself Does Not Lead to Good Face Recognition Performance:* We have long been enamored of edge maps as a powerful initial representation for visual inputs. The belief is that edges capture the most important aspects of images (the discontinuities) while being largely invariant to shallow shading gradients that are often the result of illumination variations. In the context of human vision as well, line drawings appear to be sufficient for recognition purposes. Caricatures and quick pen portraits are often highly recognizable. Do these observations mean that high spatial frequencies are critical, or at least sufficient, for face recognition? Several researchers have examined the contribution of different spatial frequency bands to face recognition [14], [21]. Their findings suggest that high spatial frequencies might not be too important for face perception. In the particular domain of line drawings, Graham Davies and his colleagues have reported [16] that images which contain exclusively contour information are very difficult to recognize (specifically, they found that subjects could recognize only 47% of the line drawings compared to 90% of the original photographs; see Fig. 3). How can we reconcile such findings with the observed recognizability of line drawings in everyday experience? Bruce and colleagues [6], [7] have convincingly argued that such depictions do, in fact, contain significant photometric cues and that the contours included in such a depiction by an accomplished artist correspond not just to a low-level edge map, but in

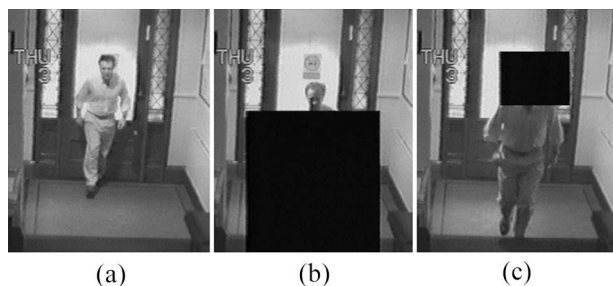


Fig. 2. Frames from video-sequences used in Burton et al. [10] study. (a) Original input. (b) Body obscured. (c) Face obscured. Based on results from such manipulations, researchers concluded that recognition of familiar individuals in low-resolution video is based largely on facial information.

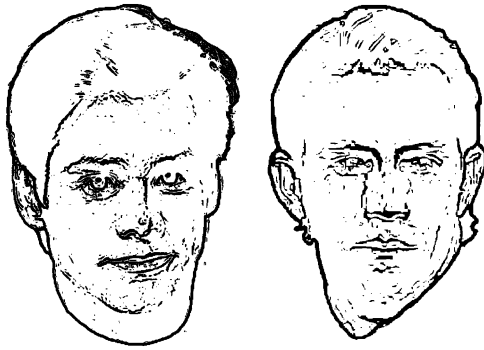


Fig. 3. Images which contain exclusively contour information are very difficult to recognize, suggesting that high-spatial frequency information, by itself, is not an adequate cue for human face recognition processes. Shown here are Jim Carrey (left) and Kevin Costner.

fact embody a face's photometric structure. It is the skillful inclusion of these photometric cues that is believed to make human generated line drawings more recognizable than computer generated ones [59]. The idea that "line drawings" contain important photometric cues leads to the prediction that recognition performance with line drawings would be susceptible to contrast negation, just as for gray-scale images. This prediction is indeed supported by experimental data [60].

B. Nature of Processing: Piecemeal Versus Holistic

1) *Result 4: Facial Features Are Processed Holistically:* Can facial features (eyes, nose, mouth, eyebrows, etc.) be processed independently from the rest of the face? Faces can often be identified from very little information. Sadr *et al.* [70] and others [15], [23] have shown that just one feature (such as the eyes or, notably, the eyebrows) can be enough for recognition of many famous faces. However, when features on the top half of one face are combined with the bottom half of another face, the two distinct identities are very difficult to recognize [91] (see Fig. 4). The holistic context seems to affect how individual features are processed. When the two halves of the face are misaligned, presumably disrupting normal holistic processing, the two identities are easily recognized. These results suggest that when taken alone, features are sometimes sufficient for facial recognition. In the context of a face, however, the geometric relationship between each feature and the rest of the face can override the diagnosticity of that feature. Although feature processing is important for facial recognition, this pattern of results suggests that configural processing is at least as important, and that facial recognition is dependent on "holistic" processes involving an interdependency between featural and configural information. Recent work has explored how

one might learn to use holistic information [67] and the contribution of holistic processing to the analysis of facial expressions [11].

2) *Result 5: Of the Different Facial Features, Eyebrows Are Among the Most Important for Recognition:* Not all facial features are created equal in terms of their role in helping identify a face [15], [19], [23], [90]. Experimental results typically indicate the importance of eyes followed by the mouth and then the nose. However, one facial feature has, surprisingly, received little attention from researchers in this domain—the eyebrows. Sadr *et al.* [70] have presented evidence suggesting that the eyebrows might not only be important features, but that they might well be among the most important, comparable to the eyes. These researchers digitally erased the eyebrows from a set of 50 celebrity face images (Fig. 5). Subjects were shown these images individually and asked to name them. Subsequently, they were asked to recognize the original set of (unaltered) images. Performance was recorded as the proportion of faces a subject was able to recognize. Performance with the images lacking eyebrows was significantly worse relative to that with the originals, and even with the images lacking eyes. These results suggest that the eyebrows may contribute in an important way to the representations underlying identity assessments.

How might one reasonably explain the perceptual significance of eyebrows in face recognition? There are several possibilities. First, eyebrows appear to be very important for conveying emotions and other nonverbal signals. Since the visual system may already be biased to attend to the eyebrows in order to detect and interpret such signals, it may be that this bias also extends to the task of facial identification. Second, for a number of reasons, eyebrows may serve as a very "stable" facial feature. Because they tend to be relatively high-contrast and large facial features, eyebrows can survive substantial image degradations. For instance, when faces are viewed at a



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.

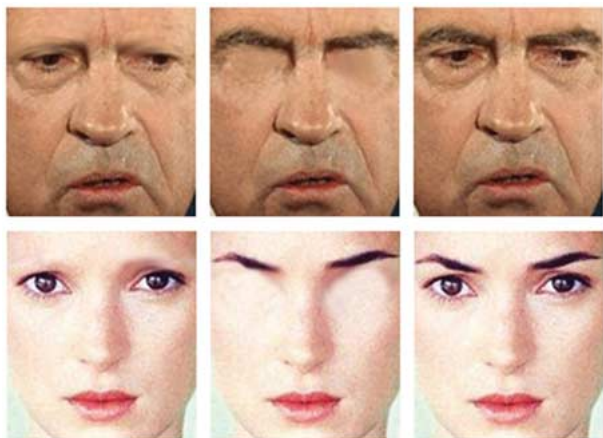


Fig. 5. Sample stimuli from Sadr et al.'s [70] experiment assessing the contribution of eyebrows to face recognition: original images of President Richard M. Nixon and actor Winona Ryder, along with modified versions lacking either eyebrows or eyes.

distance, the eyebrows continue to make an important contribution to the geometric and photometric structure of the observed image. Also, since eyebrows sit atop a convexity (the brow ridge separating the forehead and orbit), as compared to some other parts of the face, they may be less susceptible to shadow and illumination changes. Further, although the eyebrows can undergo a wide range of movements, the corresponding variations in the appearance of the eyebrows themselves do not rival those observed within the eyes and mouth, for example, as they run through the gamut of their own movements and deformations.

3) *Result 6: Important Configural Relationships Appear to be Independent Across the Width and Height Dimensions:* Taking up where the previous result left off, we can ask what aspects of the spatial structure of a head are important for judgments of identity? At least a few computer vision systems involve precise measurements of attributes such as inter-eye distance, width of mouth, and length of nose. However, it appears that the human visual system does not depend critically on these measurements. Evidence in favor of this claim comes from investigations of recognition with distorted face images [35]. A face can be compressed greatly, with no loss in its recognizability (see Fig. 6). Clearly, such compressions play havoc with absolute interfeature distance measurements, and also distance ratios across the x and y dimensions. Nevertheless, recognition performance stays invariant. One set of spatial attributes that stay unchanged with compressions are ratios of distances within the same dimension. It is possible then that human encoding of faces utilizes such ratios (we refer to them as iso-dimension ratios), and this might constitute a useful strategy for computer vision systems as well. Why might the human visual system have

adopted such a strategy, given that image compressions were not particularly commonplace until the recent advent of photography? To a limited extent, rotations in depth around the x and y axes approximate two-dimensional (2-D) compressions. Perhaps the human visual system has adopted an iso-dimension ratio encoding strategy to obtain a measure of tolerance to such transformations.

C. Nature of Cues Used: Pigmentation, Shape, and Motion

1) *Result 7: Face-Shape Appears to be Encoded in a Slightly Caricatured Manner:* Intuitively, successful face recognition requires that the human visual system should encode previously seen faces veridically. Errors in the stored representation of a face obviously weaken the potential to match new inputs to old.

However, it has been demonstrated that some departures from veridicality are actually beneficial for human face recognition. Specifically, "caricatured" versions of faces have been demonstrated to support recognition performance at least equal to or better than that achieved with veridical faces [63]. Caricatured faces can be created to exaggerate deviations in shape alone [3] or a combination of deviations in both shape and pigmentation cues [1]. This is illustrated in Fig. 7. In both cases, subjects display small, but consistent, preferences for caricatured faces as determined by several different measures [43], [44]. Shape caricaturing is evident for objects other than faces as well [27] suggesting that caricatured representations may be a widely applied strategy.

These results have been taken to suggest a norm-based representational space for faces, often referred to in the literature as "face space" [82]. This hypothesis may usefully constrain the kinds of encoding strategies employed by computational face recognition systems. It



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.

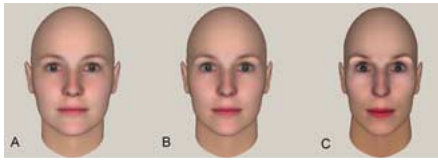


Fig. 7. Example of a face caricature. (A) Average female face for a particular face population is displayed, as well as a (B) “veridical” image of an exemplar face. (C) We create a caricatured version of the exemplar by moving away from the norm, thus exaggerating differences between the average face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

should also be noted that caricature effects tend to be strongest in images that are somehow degraded (line drawings, rapidly presented images). This may suggest that the exaggeration of individual variation plays a more important role in recognition when ordinary processing is compromised. At the very least, an interesting test for any recognition scheme is whether or not it displays “caricature effects” similar to those found in human recognition.

2) *Result 8: Prolonged Face Viewing Can Lead to High-Level Aftereffects, Which Suggest Prototype-Based Encoding:* Visual aftereffects that occur following prolonged exposure to an “adapting” stimulus have yielded many insights into the neural processing of basic visual attributes like motion, orientation, and color. In recent years, it has been shown that adaptation can lead to powerful aftereffects for more complex stimuli such as basic shapes [78] and faces [85].

The basic phenomenon of seeing any sort of aftereffect following prolonged viewing of a particular face stimulus provides strong evidence for norm-based contrastive coding of faces. The induced aftereffect can be as straightforward as a face distorted in the opposite manner as the adapting face [85], or as complex as an “anti-face” with a specific identity and no discernible distortions (see Fig. 8), suggesting multiple dimensions along which neural populations can be tuned. Furthermore, there is good reason to suspect that these aftereffects are the result of adaptation at relatively high levels of the visual system. Face aftereffects are robust to rotations of the face image [84] as well as changes in size [93], ruling out contributions from lower level mechanisms that process very small image regions.

Face adaptation and the associated aftereffects make “face space” a real neural possibility rather than a useful metaphor and also provide a means for examining its structure. For example, recent work shows that it is possible to simultaneously induce distinct distortion aftereffects for male and female faces, suggesting separate neural substrates for each gender [47].

In terms of computational models, face aftereffects provide both a clue to a useful encoding strategy

(prototype-based encoding with high-level “contrast”) and an interesting test for existing systems (determining whether identity-specific biases can result from exposing a model to a particular individual). These phenomena also indicate that human face perception is a highly plastic process, adjusting itself continually to the faces that surround us [64], [86].

3) *Result 9: Pigmentation Cues Are at Least as Important as Shape Cues:* There are two basic ways in which faces can differ—in terms of their shape, and in terms of how they reflect light, or their pigmentation. By “pigmentation,” we refer to all surface reflectance properties, including albedo, hue, specularity, translucency, and spatial variation in these properties. When referring to *all* surface reflectance properties of faces, we prefer the term “pigmentation” (or “surface appearance”) to the terms “texture” or “color,” which invite confusion because they are commonly used to refer to specific subsets of surface reflectance properties (spatial variation in albedo and greater reflectance of particular wavelengths, respectively).

Recent studies have investigated whether shape or pigmentation cues are more important for face recognition. The approach taken has been to create sets of faces that differ from one another in terms of only their shape or only their pigmentation, using either laser-scanned models of faces [57], artificial faces [68], or morphing photographs of faces (in which case shape is defined in terms of the 2-D

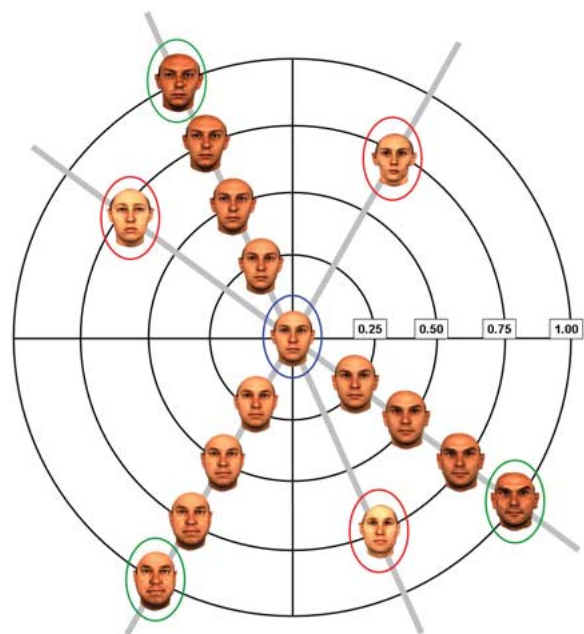


Fig. 8. Faces and their associated “anti-faces” in a schematic face space. Prolonged viewing of a face within a green circle will cause the central face to be misidentified as the individual within the red circle along the same “identity trajectory” (from [45]).



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.

outlines of the face and individual features, pictured in Fig. 9) [68]. With each of these classes of stimuli, subjects have performed about equally well using either shape or pigmentation cues. This provides evidence that the two kinds of cues are used about equally by humans to recognize faces. A study from our laboratory investigating the use of these cues for the recognition of familiar faces also found that shape and pigmentation are about equally important. An implication of this work is that artificial face recognition systems would benefit from representing pigmentation as well as shape cues.

4) *Result 10: Color Cues Play a Significant Role Especially When Shape Cues Are Degraded:* The luminance structure of face images is undoubtedly of great significance for recognition. Past research has suggested that the use of these cues may adequately account for face-identification performance with little remaining need to posit a role for color information. Furthermore, people tend to accurately identify faces that are artificially colored [40]. However, recent evidence [89] counters the notion that color is unimportant for human face recognition and suggests instead that when shape cues in images are compromised (say, by reductions in resolution), the brain relies on color cues to pinpoint identity. In such circumstances, recognition performance with color images is significantly better than with gray-scale images. Precisely how does color information facilitate face recognition? One possibility is that color provides diagnostic information. The expression “diagnostic information” refers to color cues that are specific to an individual, for instance the particular hue of their hair or skin that may allow us to identify them. On

the other hand, color might facilitate low-level image analysis, and thus indirectly aid face recognition. An example of such a low-level task is image segmentation—determining where one region ends and the other starts. As many years of work in computer vision has shown [20], [29], this task is notoriously difficult and becomes even more intractable as images are degraded. Color may facilitate this task by supplementing the luminance-based cues and thereby lead to a better parsing of a degraded face image in terms of its constituent regions. Experimental data favor the second possibility. Recognition performance with pseudo-colored face images (which do not contain diagnostic hue information) is just as high as with natural color images (and both are significantly better than grayscale images, when shape cues are degraded). Fig. 10 illustrates this idea. The images show the luminance and color components of sample face inputs. They suggest that color distributions can supplement luminance information to allow for a better estimation of the boundaries, shapes, and sizes of facial attributes such as eyes and hair lines.

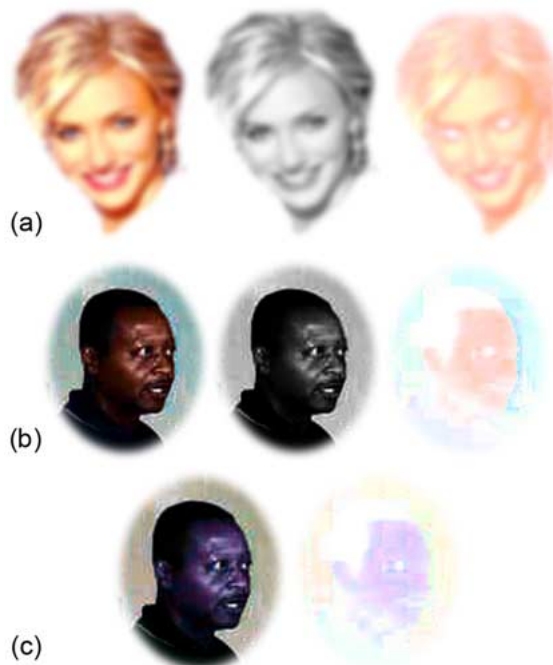


Fig. 10. Examples that illustrate how color information may facilitate some important low-level image analysis tasks such as segmentation. (a) Hue distribution (right panel) allows for a better estimation of the shape and size of the eyes than the luminance information alone (middle panel). Left panel shows the original image. Similarly, in (b), hue information (right panel) allows for a better segmentation and estimation of the location and shape of hair line than just luminance information (middle panel). This facilitation of low-level analysis happens with other choices of colors as well, such as in the pseudo-color image shown on the left in (c). Hue distribution here, as in (b), aids in estimating the position of facial attributes such as hair line.



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.)

5) *Result 11: Contrast Polarity Inversion Dramatically Impairs Recognition Performance, Possibly Due to Compromised Ability to Use Pigmentation Cues:* Skilled darkroom technicians working in the photo retouching industry several decades ago noticed that faces were particularly difficult to recognize when viewed in reversed contrast, as in photographic negatives (as illustrated in Fig. 11). Subsequently, the phenomenon has been studied extensively in the vision science community, with the belief that determining how recognition can be impaired helps us understand how it works under normal conditions. Contrast negation is a reversible manipulation that does not remove any information from the image. Though no information is lost, our ability to use the information in the image is severely compromised. This suggests that some normally useful information is rendered unusable by negation.

When pigmentation cues are unavailable, as in uniformly pigmented three-dimensional (3-D) face models (derived from laser scans) or in other stimuli for which pigmentation cues are unavailable (see Result 9 for examples), recognition is not significantly worse with negative contrast [8], [69]. This suggests that pigmentation cues might be disrupted by negation. Other work with uniformly pigmented face models has found evidence that shading cues are disrupted by contrast negation, but only for faces lit from above [48]. These findings suggest that human face recognition uses representations that are sensitive to contrast direction and that pigmentation and shading play important roles in recognition.

6) *Result 12: Illumination Changes Influence Generalization:* Some computational models of recognition assume that a face must be viewed under many different illumination conditions for robust representations. However, there is evidence that humans are capable of

generalizing representations of a face to radically novel illumination conditions. In one recent study [2], subjects shown a laser scanned image of an unfamiliar face with illumination coming from one side, were subsequently shown a face illuminated strongly from the other side, and were asked whether both images were of the same face (see Fig. 12). Subjects were well above chance at deciding whether the second face was the same as the first, indicating significant ability to generalize the representation of the face to novel illumination conditions. However, the subjects were significantly impaired at this task relative to when the two faces were presented under the same illumination, indicating that the generalization to novel illumination conditions is not perfect.

An implication of this result is that human recognition of faces is sensitive to illumination direction, but is capable of significant generalization to novel illumination conditions even after viewing only a single image.

7) *Result 13: View-Generalization Appears to be Mediated by Temporal Association:* Recognizing a familiar face across variations in viewing angle is a very challenging computational task that the human visual system can solve with remarkable ease. Despite the fact that image-level differences between two views of the same face are much larger than those between two different faces viewed at the same angle [56], human observers are somehow able to link the correct images together.

It has been suggested that temporal association serves as the “perceptual glue” that binds different images of the same object into a useful whole. Indeed, close temporal association of novel images viewed in sequence is sufficient to induce some IT neurons to respond similarly to arbitrary image pairs [53]. Behavioral evidence from human observers exposed to rotating “paperclip” objects supports rapid learning of image sequences as well [74], [75].

In terms of human face recognition, temporal association of two unique faces (one frontally viewed, the other viewed in profile) has been demonstrated to have intri-

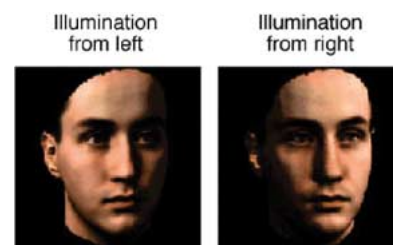


Fig. 12. Stimuli from Braje et al. [2]. These two images demonstrate the kind of lighting used in this experiment. After being shown an image like the one on the left, subjects were well above chance at determining whether a subsequently presented image such as the one on the right represented the same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

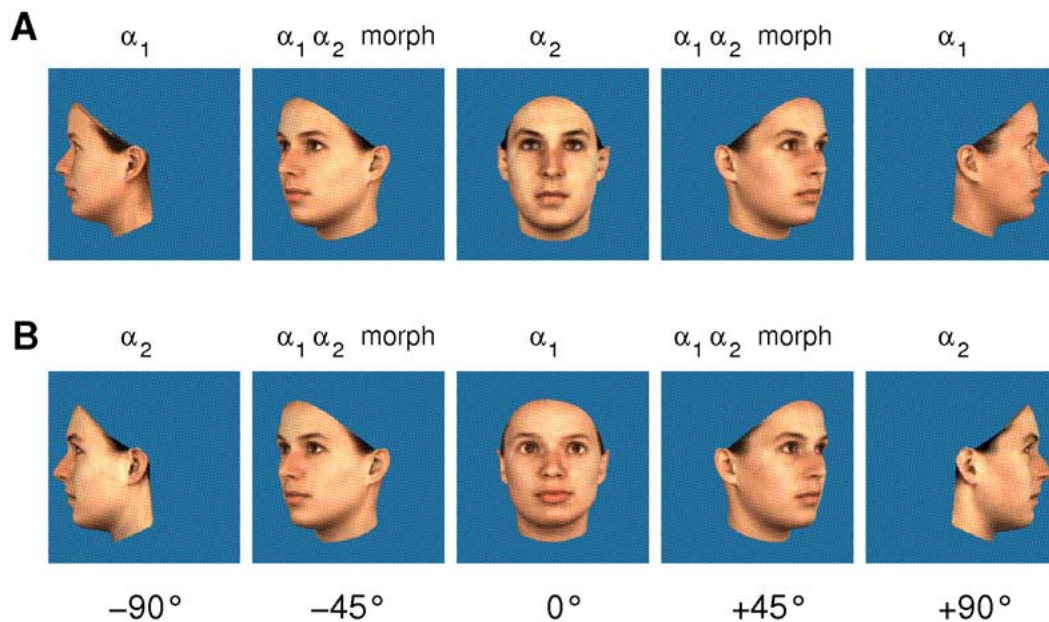


Fig. 13. Time course of sequences shown to observers in Wallis and Bulthoff [83]. Faces α_1 and α_2 are each used as the frontally viewed face in separate sequences, and combined with the other face profile in their respective movies. 3/4 morphs between α_1 and α_2 are used to interpolate between the frontally viewed faces and the profiles to create a smooth motion sequence. Same/Different performance for faces appearing in the same sequence is impaired relative to pairs of faces appearing in different sequences.

guing consequences for recognition. Brief exposure to movies containing a rotating head which morphs between one individual and another as it rotates from frontal to profile views can impair observers’ ability to distinguish between the two faces contained in the sequence [83] (see Fig. 13).

Taken together, these results suggest that the temporal proximity of images is a powerful tool for establishing object representations. Studying recognition performance using images that lack a temporal context may be a profound handicap to our understanding of how view invariance is achieved. Exploring image sequences using mechanisms that make explicit temporal associations [22] may be a powerful means for view generalization.

8) *Result 14: Motion of Faces Appears to Facilitate Subsequent Recognition:* Do dynamic cues aid face recognition? The answer is “yes” but only in some cases. Rigid motion (such as that obtained from a camera rotating around a motionless head) can facilitate recognition of previously viewed faces [58], [71] but there seems to be very little, if any, benefit of seeing these views during the learning phase. By contrast, nonrigid motion (where the individuals exhibit emotive facial expressions or speech movements) plays a greater role. Experiments in [41], using subtle morphs of form and facial motion in novel (i.e., unfamiliar) faces, showed that nonrigid facial motion from one face applied to the form of another face can bias an observer to misidentify the latter as the former (see

Fig. 14). Experiments with famous (i.e., highly familiar) faces [42] again showed a facilitation in recognition with dynamic cues from expressive or talking movements, but not from rigid motion. Facilitation was most pronounced for faces whose movement was judged as “distinctive.” Note also that facilitation comes from a natural sequence of moving images, not merely from having more views avail-

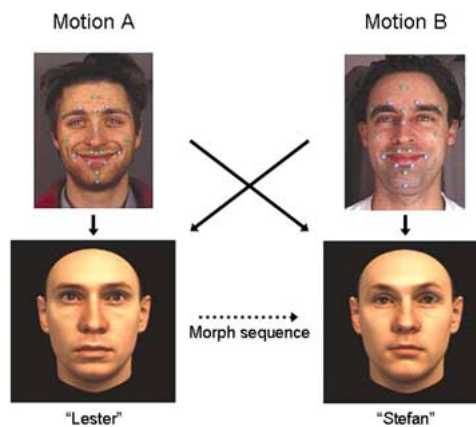


Fig. 14. Facial motion from expressions and talking were morphed onto forms of “Lester” and “Stefan.” Subjects could be biased to identify an anti-caricatured (morphed towards the average) form of Lester as Stefan when Stefan’s movements were imposed onto Lester’s form. (From [41].)

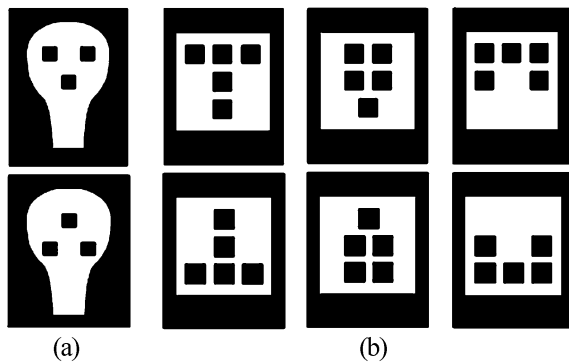


Fig. 15. (a) Newborns preferentially orient their gaze to face-like pattern on top, rather than one shown on bottom, suggesting some innately specified representation for faces (from [36]). (b) As a counterpoint to idea of innate preferences for faces, Simion et al. [73] have shown that newborns consistently prefer top-heavy patterns (left column) over bottom-heavy ones (right column). It is unclear whether this is the same preference exhibited in earlier work, and if it is, whether it is face-specific or some other general-purpose or artifactual preference.

able: The facilitation is greatly lessened when the same frames are presented in random order or in a static array.

These results suggest that face motion is more than just a sequence of viewpoints to the face recognition system. The dynamic cues from expressive and talking movements provide information about aspects of facial structure that transcend the gains of simply having multiple viewpoints.

D. Developmental Progression

1) *Result 15: Visual System Starts With a Rudimentary Preference for Face-Like Patterns:* What, if any, are the face-specific biases that the human visual system starts out with? The answer to this question will help a computer vision researcher decide between two alternatives: 1) program explicit face-specific templates into a face recognition system or 2) allow implicit templates to form through learning processes, be they face-specific or object-general. Newborns selectively gaze at “face-like” patterns only hours after birth. A pattern that is face-like can be something as simple as that shown in Fig. 15(a): three dots within an oval that represent the two eyes and a mouth. An impossible face (created by vertically inverting the triad of dots) does not attract the newborn’s attention as much as the more normal face. However, the specificity of the response to the three-dot arrangement has been called into question. More recent work [73] suggests that newborns simply prefer “top-heaviness” [Fig. 15(b)]. Thus, it remains unclear whether this is a general preference (perhaps with no practical significance) or a face-specific orienting response to prime the infant in bootstrapping its nascent face recognition system. Even if this preference really is an innate face-orienting mechanism, it may be more for the

benefit of the mother (e.g., to form the mother–child bond) than the infant’s face processing capabilities.

A simple arrangement of three dots within an oval may serve as an appropriate template for detecting faces in the bootstrapping stages of a face-learning system. Similar templates have been used with reasonable success in some applications (for example, [76]) of face detection.

2) *Result 16: Visual System Progresses From a Piecemeal to a Holistic Strategy Over the First Several Years of Life:* Normal adults show a remarkable deficit in recognition of inverted faces versus upright faces, whereas the deficit is quite small for inverted images of nonface objects such as houses [88]. A number of studies have shown, however, that this pattern of results takes many years to develop ([13], [34], [50], [54], [55], [61], [72]). Six-year-old children are not affected by inversion when it comes to recognizing seen faces in a seen–unseen pair [13]; eight-year-olds show some inversion effect and ten-year-olds exhibit near adult-like performance (see Fig. 16). In [54], the authors selectively manipulated spacing (moving the location of features on a face) versus features (taking eyes or mouth from different faces) and found what may be the source of the developmental progression of the inversion effect: six- and eight-year-olds show a relative deficit in the processing of spacing in both upright and inverted faces, but ten-year-olds resemble adults in that they show the deficit for inverted but not upright faces. Thus, it looks as though the processing of spacing matures later than featural processing. Interestingly, although six-year-old children are not sensitive to inversion in the tests mentioned previously, they are susceptible to the Thatcher Illusion (Thompson, 1980 [46]), suggesting that the limited holistic processing that is available to the six-year-old is sufficient for orientation-sensitive local feature parsing.

This pattern of behavior suggests that over the course of several years, a shift in strategy occurs. Initially, infants and toddlers adopt a largely piecemeal, feature-based strategy for recognizing faces. Gradually, a more sophisticated holistic strategy involving configural information evolves. This is indirect evidence for the role of configural information in achieving the robust face recognition performance that adults exhibit ([24], [65]).

E. Neural Underpinnings

1) *Result 17: Human Visual System Appears to Devote Specialized Neural Resources for Face Perception:* Whether or not faces constitute a “special” class of visual stimuli has been the subject of much debate for many years. Since the first demonstrations of the “inversion effect” described previously [88], it has been suspected that unique cognitive and neural mechanisms may exist for face processing in the human visual system.

Indeed, there is a great deal of evidence that the primary locus for human face processing may be found on

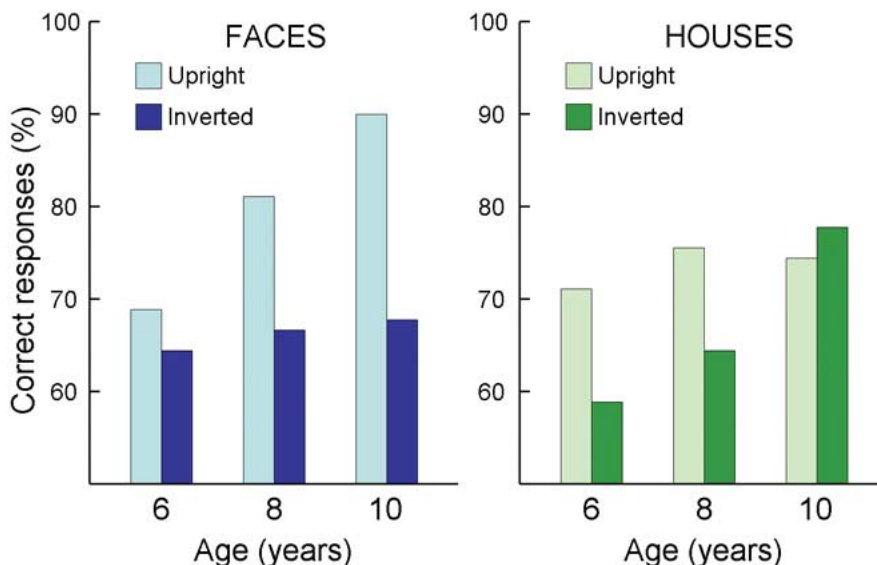


Fig. 16. Generally, six-year-olds are rather poor at upright and inverted faces. As their age approaches ten years, their performance improves dramatically on upright faces, but hardly any improvement is exhibited on inverted faces. (Data from Carey and Diamond, 1971.)

the fusiform gyrus of the extra-striate visual cortex [38], [51]. This region shows an intriguing pattern of selectivity (schematic faces do not give rise to much activity) and generality (animal faces do elicit a good response) [80], suggesting a strong domain-specific response for faces (see Fig. 17). In keeping with behavioral results, the “fusiform face area” (FFA) also appears to exhibit an “inversion effect” [39]. Overall, the characterization of the FFA as a dedicated face processing module appears very strong.

However, it must be noted that the debate over faces being “special” is far from over. It has been suggested that rather than being a true “face module,” the FFA may be responsible for performing either subordinate or “expert-level” categorization of generic objects. There are results from both behavioral studies [18], [25] and

neuroimaging studies [26] that lend some support to this “perceptual expertise” account. Recent findings appear to favor the original “face module” account of the FFA’s function, however [28].

The full breadth and depth of the arguments supporting both positions are beyond the scope of this review (see [52] for a more thorough treatment), but it is important to recognize that specialized face processing mechanisms in the human visual system are a very real possibility. Whatever its ultimate status, the response profile of the FFA provides a potentially valuable set of constraints for computational systems, indicating the extent of selectivity and generality we should expect from face recognition systems.

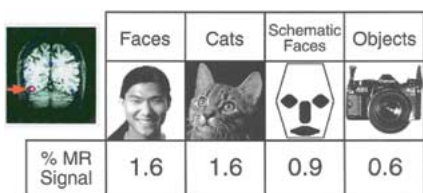


Fig. 17. Upper left, an example of FFA in one subject, showing right-hemisphere lateralization. Also included here are example stimuli from Tong et al. [80], together with amount of percent signal change observed in FFA for each type of image. Photographs of human and animal faces elicit strong responses,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

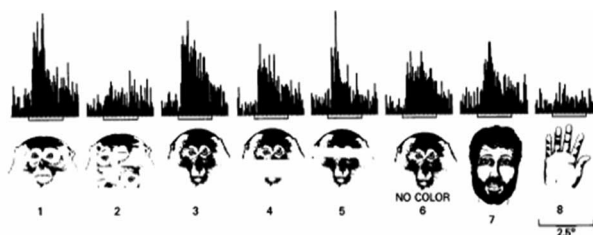


Fig. 18. Example of a monkey IT cell’s responses to variations on a face stimulus (from Desimone et al. [17]). Response is robust to many degradations of the primate face (save for scrambling) and also responds very well to a human face. Lack of a response to the hand indicates that this cell is not 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.

2) *Result 18: Latency of Responses to Faces in IT Cortex is About 120 ms, Suggesting a Largely Feedforward Computation:* Human observers can carry out visual recognition tasks very rapidly. Behavioral RTs are already quite fast and represent a potentially large overestimate of the time required for recognition due to the motor component of signaling a response. Indeed, when a neural marker of recognition is used, accurate performance on such seemingly complex tasks as determining the presence/absence of an animal in a natural scene appears to require as little as 50 ms [79].

Recently, it has been shown that though this particular task (animal/no animal) seems quite complicated, it may be solvable using very low-level visual representations [37]. That said, there is neurophysiological evidence that truly complex tasks, such as face recognition, may be carried out over a surprisingly short period of time.

Neurons in the primate inferotemporal (IT) cortex can exhibit selectivity to stimuli that are more complicated than the simple gratings and bars that elicit responses from cells in early visual areas. In particular, it has been noted that there are some cells in IT cortex that are selective for faces [17] (see Fig. 18). Moreover, the latency of response in these cells is in the neighborhood of 80–160 ms [62]. More recent results have demonstrated that fine-grained discrimination of face identity or expression is possible at approximately 50 ms after exposure [77].

The computational relevance of these results is that recognition as it is performed up to the level of IT cortex probably requires only one feedforward pass through the visual system. Feedback and iterative processing are likely not major factors in the responses recorded in these studies, especially if the stimuli are clear, undegraded images. While impoverished images will likely require some amount of iterative processing (and thus more time), relatively clean images can be dealt with very rapidly. This is a very important constraint on recognition algorithms, as it indicates that sufficient information must be extracted immediately from the image and cannot necessarily be “cleaned up” later.

3) *Result 19: Facial Identity and Expression Might be Processed by Separate Systems:* To what extent is the processing of facial identity bound with the processing of facial expression? That is, is it possible to extract facial expression independently of the identity and *vice versa*, or are the two inextricably linked? Beyond being a mere academic point, the computational implications of this question would determine whether a biologically based implementation would be able to identify a person without

taking into account the person’s expression or to judge the facial emotions in a human–computer interaction application without going through the process of extracting a representation of identity.

The most popular theoretical model [5] and a recent neural systems model [33] both propose a separation of identity and expression processes early in the facial perception pathway, leaving each of these processes to act in parallel using distinct representations. This account has been supported by a large body of evidence. Behavioral studies [4] show that familiarity does not aid expression reportability; functional brain imaging [87] has identified distinct brain areas for identity versus expression; brain-injured patients [81], [92] have provided examples of selective impairments in identity or expression processing; and electrophysiology studies in primates [32] find that single neurons can be identified which are selective for either identity or expression. See a recent review of such results in [12].

On the other hand, Calder and Young [12] point out that although there seems to be a significant amount of dissociation between identity and expression, most studies do leave some room for overlap, perhaps at the representational stage. For example, although some neurons responded only to identity and some only to expression in the Hasselmo *et al.* study [32], a smaller subset of neurons responded to both factors. Such ambiguity leads Calder and Young to propose a statistical account which predicts a representation of identity, expression, and identity expression (i.e., the combination of the two) stemming from a uniform perceptual process. They still agree, however, that these representations are then processed largely independently.

II. CONCLUSION

The twin enterprises of visual neuroscience and computer vision have deeply synergistic objectives. An understanding of human visual processes involved in face recognition can facilitate and, in turn be facilitated by, better computational models. Our presentation of results in this paper is driven by the goal of furthering crosstalk between the two disciplines. The observations included here constitute 19 brief vignettes into what is surely a most impressive and rather complex biological system. We hope that these vignettes will help in the ongoing computer vision initiatives to create face recognition systems that can match, and eventually exceed, the capabilities of their human counterparts. ■

REFERENCES

- [1] P. J. Benson and D. I. Perrett, “Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images,” *Eur. J. Cognitive Psychol.*, vol. 3, no. 1, pp. 105–135, 1991.
- [2] W. L. Braje, D. Kersten, M. J. Tarr, and N. F. Troje, “Illumination effects in face recognition,” *Psychobiology*, vol. 26, pp. 371–380, 1998.
- [3] S. E. Brennan, “The caricature generator,” *Leonardo*, vol. 18, pp. 170–178, 1985.

- [4] V. Bruce, "Influences of familiarity on the processing of faces," *Perception*, vol. 15, pp. 387–397, 1986.
- [5] V. Bruce and A. W. Young, "Understanding face recognition," *Br. J. Psychol.*, vol. 77, pp. 305–327, 1986.
- [6] —, *In the Eye of the Beholder*. Oxford, U.K.: Oxford Univ. Press, 1998.
- [7] V. Bruce, E. Hanna, N. Dench, P. Healey, and M. Burton, "The importance of 'mass' in line drawings of faces," *Appl. Cognitive Psychol.*, vol. 6, pp. 619–628, 1992.
- [8] V. Bruce and S. Langton, "The use of pigmentation and shading information in recognizing the sex and identities of faces," *Perception*, vol. 23, pp. 803–822, 1994.
- [9] V. Bruce, Z. Henderson, K. Greenwood, P. J. B. Hancock, A. M. Burton, and P. I. Miller, "Verification of face identities from images captured on video," *J. Experimental Psychol.: Applied*, vol. 5–4, pp. 339–360, 1999.
- [10] A. M. Burton, S. Wilson, M. Cowan, and V. Bruce, "Face recognition in poor-quality video," *Psychol. Sci.*, vol. 10, pp. 243–248, 1999.
- [11] A. J. Calder, A. W. Young, J. Keane, and M. Dean, "Configural information in facial expression perception," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 26, pp. 527–551, 2000.
- [12] A. J. Calder and A. W. Young, "Understanding the recognition of facial identity and facial expression," *Nature Rev. Neurosci.*, vol. 6, no. 8, pp. 641–651, 2005.
- [13] S. Carey and R. Diamond, "From piecemeal to configurational representation of faces," *Science*, vol. 195, pp. 312–314, 1977.
- [14] N. P. Costen, D. M. Parker, and I. Craw, "Effects of high-pass and low-pass spatial filtering on face identification," *Perception Psychophys.*, vol. 58, pp. 602–612, 1996.
- [15] G. Davies, H. Ellis, and J. Shepherd, "Cue saliency in faces as assessed by the 'Photofit' technique," *Perception*, vol. 6, pp. 263–269, 1977.
- [16] —, "Face recognition accuracy as a function of mode of representation," *J. Appl. Psychol.*, vol. 63, pp. 180–187, 1978.
- [17] R. Desimone, T. D. Albright, C. G. Gross, and C. Bruce, "Stimulus-selective properties of inferior temporal neurons in the macaque," *J. Neurosci.*, vol. 4, no. 8, pp. 2051–2062, 1984.
- [18] R. Diamond and S. Carey, "Why faces are and are not special: An effect of expertise," *J. Experimental Psychol.: General*, vol. 115, no. 2, pp. 107–117, 1986.
- [19] H. D. Ellis, J. W. Shepherd, and G. M. Davies, "Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition," *Perception*, vol. 8, no. 4, pp. 431–439, 1979.
- [20] P. Felzenszwalb and D. Huttenlocher, "Image segmentation using local variation," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 1998, pp. 98–104.
- [21] A. Fiorentini, L. Maffei, and G. Sandini, "The role of high spatial frequencies in face perception," *Perception*, vol. 12, pp. 195–201, 1983.
- [22] P. Foldiak, "Learning invariance from transformation sequences," *Neural Comp.*, vol. 3, pp. 194–200, 1991.
- [23] I. H. Fraser, G. L. Craig, and D. M. Parker, "Reaction time measures of feature saliency in schematic faces," *Perception*, vol. 19, no. 5, pp. 661–673, 1990.
- [24] A. Freire, K. Lee, and L. A. Symons, "The face-inversion effect as a deficit in the encoding of configural information: Direct evidence," *Perception*, vol. 29, no. 2, pp. 159–170, 2000.
- [25] I. Gauthier and M. J. Tarr, "Becoming a 'Greeble' expert: Exploring the face recognition mechanism," *Vision Res.*, vol. 37, no. 12, pp. 1673–1682, 1997.
- [26] I. Gauthier, A. W. Anderson, M. J. Tarr, P. Skudlarski, and J. C. Gore, "Levels of categorization in visual objects studied with functional MRI," *Current Biol.*, vol. 7, pp. 645–651, 1997.
- [27] J. J. Gibson, *Motion Picture Testing Res.* Washington, DC: AAF Aviation Psychology Program, vol. 7, 1947.
- [28] K. Grill-Spector, N. Knouf, and N. Kanwisher, "The fusiform face area subserves face perception, not generic within-category identification," *Nature Neurosci.*, vol. 7, no. 5, pp. 555–562, 2004.
- [29] R. Haralick, "Survey, image segmentation techniques," *Computer Vision, Graphics, and Image Processing*, vol. 29, pp. 100–135, 1985.
- [30] L. D. Harmon and B. Julesz, "Masking in visual recognition: Effects of two-dimensional noise," *Science*, vol. 180, pp. 1194–1197, 1973a.
- [31] L. D. Harmon, "The recognition of faces," *Scientific American*, vol. 229, no. 5, pp. 70–83, 1973b.
- [32] M. E. Hasselmo, E. T. Rolls, and G. C. Baylis, "The role of expression and identity in face-selective responses of neurons in the temporal visual cortex of the monkey," *Behav. Brain Res.*, vol. 32, pp. 203–218, 1989.
- [33] J. V. Haxby, E. A. Hoffman, and M. I. Gobbini, "The distributed human neural system for face perception," *Trends Cogn. Sci.*, vol. 4, pp. 223–233, 2000.
- [34] D. C. Hay and R. Cox, "Developmental changes in the recognition of faces and facial features," *Infant Child Devel.*, vol. 9, pp. 199–212, 2000.
- [35] G. J. Hole, P. A. George, K. Eaves, and A. Razek, "Effects of geometric distortions on face recognition performance," *Perception*, vol. 31, no. 10, pp. 1221–1240, 2002.
- [36] M. H. Johnson, S. Dziurawiec, H. Ellis, and J. Morton, "Newborns preferential tracking of face-like stimuli and its subsequent decline," *Cognition*, vol. 40, pp. 1–19, 1991.
- [37] J. S. Johnson and B. A. Olshausen, "Timecourse of neural signatures of object recognition," *J. Vision*, vol. 3, pp. 499–512, 2003.
- [38] N. Kanwisher, J. McDermott, and M. Chun, "The fusiform face area: A module in human extrastriate cortex specialized for the perception of faces," *J. Neurosci.*, vol. 17, pp. 4302–4311, 1997.
- [39] N. Kanwisher, F. Tong, and K. Nakayama, "The effect of face inversion on the human fusiform face area," *Cognition*, vol. 68, pp. B1–B11, 1998.
- [40] R. Kemp, G. Pike, P. White, and A. Musselman, "Perception and recognition of normal and negative faces: The role of shape from shading and pigmentation cues," *Perception*, vol. 25, pp. 37–52, 1996.
- [41] B. Knappmeyer, I. M. Thornton, and H. H. Bülthoff, "The use of facial motion and facial form during the processing of identity," *Vision Res.*, vol. 43, pp. 1921–1936, 2003.
- [42] K. Lander and L. Chuang, "Why are moving faces easier to recognize?" *Visual Cognition*, vol. 12, pp. 429–442, 2005.
- [43] K. J. Lee and D. Perrett, "Presentation-time measures of the effects of manipulations in colour space on discrimination of famous faces," *Perception*, vol. 26, pp. 733–752, 1997.
- [44] —, "Manipulation of colour and shape information and its consequence upon recognition and best-likeness judgments," *Perception*, vol. 29, pp. 1291–1312, 2000.
- [45] D. A. Leopold, A. J. O'Toole, T. Vetter, and V. Blanz, "Prototype-referenced shape encoding revealed by high-level aftereffects," *Nature Neurosci.*, vol. 4, pp. 89–93, 2001.
- [46] M. B. Lewis, "Thatcher's children: Development and the Thatcher illusion," *Perception*, vol. 32, pp. 1415–21, 2003.
- [47] A. C. Little, L. M. DeBruine, and B. C. Jones, "Sex-contingent face after-effects suggest distinct neural populations code male and female faces," *Proc. Roy. Soc. London, Series B*, vol. 272, pp. 2283–2287, 2005.
- [48] C. H. Liu, C. A. Collin, A. M. Burton, and A. Chaudhuri, "Lighting direction affects recognition of untextured faces in photographic positive and negative," *Vision Res.*, vol. 39, pp. 4003–4009, 1999.
- [49] C. H. Liu, H. Seetzen, A. M. Burton, and A. Chaudhuri, "Face recognition is robust with incongruent image resolution: Relationship to security video images," *J. Experimental Psychol.: Applied*, vol. 9, pp. 33–41, 2003.
- [50] D. Maurer, R. Le Grand, and C. J. Mondloch, "The many faces of configural processing," *Trends in Cognitive Sciences*, vol. 6, pp. 255–260, 2002.
- [51] G. McCarthy, A. Puce, J. C. Gore, and T. Allison, "Face specific processing in the human fusiform gyrus," *J. Cognitive Neurosci.*, vol. 9, pp. 605–610, 1997.
- [52] E. McKone and N. Kanwisher, "Does the human brain process objects of expertise like faces? A review of the evidence," in *From Monkey Brain to Human Brain*, S. Dehaene, J. R. Duhamel, M. Hauser, and G. Rizzolatti, Eds. Cambridge, MA: MIT Press, 2005.
- [53] Y. Miyashita, "Inferior temporal cortex: Where visual perception meets memory," *Ann. Rev. Neurosci.*, vol. 16, pp. 245–263, 1993.
- [54] C. J. Mondloch, R. Le Grand, and D. Maurer, "Configural face processing develops more slowly than featural face processing," *Perception*, vol. 31, pp. 553–566, 2002.
- [55] C. J. Mondloch, S. Geldart, D. Maurer, and R. Le Grand, "Developmental changes in face processing skills," *J. Experimental Child Psychol.*, vol. 86, pp. 67–84, 2003.
- [56] Y. Moses, Y. Adini, and S. Ullman, "Face recognition: The problem of compensating for illumination changes," in *Proc. Eur. Conf. Computer Vision*, 1994, pp. 286–296.
- [57] A. J. O'Toole, T. Vetter, and V. Blanz, "Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: An application of three-dimensional morphing," *Vision Res.*, vol. 39, pp. 3145–3155, 1999.
- [58] A. J. O'Toole, D. A. Roark, and H. Abdi, "Recognizing moving faces: A psychological and neural synthesis," *Trends Cogn. Sci.*, vol. 6, pp. 261–266, 2002.
- [59] D. E. Pearson and J. A. Robinson, "Visual communication at very low data rates," *Proc. IEEE*, vol. 74, no. 4, pp. 795–812, Apr. 1985.
- [60] D. Pearson, E. Hanna, and K. Martinez, "Computer generated cartoons," in *Images and Understanding*, H. Barlow, C. Blakemore, and M. Weston-Smith, Eds. Cambridge,

- U.K.: Cambridge Univ. Press, 1990, ch. 3, pp. 46–60.
- [61] E. Pellicano and G. Rhodes, “Holistic processing of faces in preschool children and adults,” *Psycholog. Sci.*, vol. 14, pp. 618–622, 2003.
- [62] D. I. Perrett, E. T. Rolls, and W. Caan, “Visual neurones responsive to faces in the monkey temporal cortex,” *Experimental Brain Res.*, vol. 47, no. 3, pp. 329–342, 1982.
- [63] G. Rhodes, *Superportraits: Caricatures and Recognition*. East Sussex, U.K.: Psychology Press, 1996.
- [64] G. Rhodes, L. Jeffery, T. L. Watson, C. W. G. Clifford, and K. Nakayama, “Fitting the mind to the world: Face adaptation and attractiveness aftereffects,” *Psycholog. Sci.*, vol. 14, pp. 558–566, 2003.
- [65] M. Riesenhuber, I. Jarudi, S. Gilad, and P. Sinha, “Face processing in humans is compatible with a simple shape-based model of vision,” *Proc. Roy. Soc. London B (Suppl.)*, vol. 271, pp. S448–S450, 2004.
- [66] D. A. Roark, A. J. O’Toole, and H. Abdi, “Human recognition of familiar and unfamiliar people in naturalistic video,” in *Proc. IEEE Int. Workshop Analysis and Modeling of Faces*, Nice, France, 2003, pp. 36–43.
- [67] R. Robbins and E. McKone, “Can holistic processing be learned for inverted faces?” *Cognition*, vol. 88, pp. 79–107, 2003.
- [68] R. Russell, P. Sinha, I. Biederman, and M. Nederhouser, “The utility of surface reflectance for the recognition of upright and inverted faces,” *Vis. Res.*, in press.
- [69] —, “Is pigmentation important for face recognition? Evidence from contrast negation,” *Perception*, vol. 35, pp. 749–759, 2006.
- [70] J. Sadr, I. Jarudi, and P. Sinha, “The role of eyebrows in face recognition,” *Perception*, vol. 32, pp. 285–293, 2003.
- [71] W. Schiff, L. Banka, and G. de Bordes, “Recognizing people seen in events via dynamic ‘mug shots’,” *Amer. J. Psychol.*, vol. 99, pp. 219–231, 1986.
- [72] G. Schwarzer, “Development of face processing: The effect of face inversion,” *Child Devel.*, vol. 71, pp. 391–401, 2000.
- [73] F. Simion, V. M. Cassia, C. Turati, and E. Valenza, “The origins of face perception: Specific versus non-specific mechanisms,” *Infant Child Devel.*, vol. 10, pp. 59–65, 2001.
- [74] P. Sinha and T. Poggio, “I think I know that face . . .,” *Nature*, vol. 384, p. 404, 1996.
- [75] —, “The role of learning in 3-D form perception,” *Nature*, vol. 384, pp. 460–463, 1996.
- [76] P. Sinha, “Qualitative representations for recognition,” *Biologically Motivated Computer Vision, Proc.*, vol. 2525, pp. 249–262, 2002.
- [77] Y. Sugase, S. Yamane, S. Ueno, and K. Kawano, “Global and fine information coded by single neurons in the temporal visual cortex,” *Nature*, vol. 400, pp. 869–873, 1999.
- [78] S. Suzuki and P. Cavanagh, “A shape-contrast effect for briefly presented stimuli,” *J. Experimental Psychology: Human Perception and Performance*, vol. 24, pp. 1315–1341, 1998.
- [79] S. Thorpe, D. Fize, and C. Marlot, “Speed of processing in the human visual system,” *Nature*, vol. 381, pp. 520–522, 1996.
- [80] F. Tong, K. Nakayama, M. Moscovitch, O. Weinrib, and N. Kanwisher, “Response properties of human fusiform face area,” *Cognitive Neuropsychol.*, vol. 17, no. 1, pp. 257–279, 2000.
- [81] D. Tranel, A. R. Damasio, and H. Damasio, “Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity,” *Neurology*, vol. 38, pp. 690–696, 1988.
- [82] T. Valentine, Ed., *Face-Space Models of Face Recognition*. Hillsdale, NJ: Lawrence Erlbaum, 1999.
- [83] G. Wallis and H. H. Bulthoff, “Effects of temporal association on recognition memory,” *Proc. Nat. Acad. Sci.*, vol. 98, no. 8, pp. 4800–4804, 2001.
- [84] T. L. Watson and C. W. G. Clifford, “Pulling faces: An investigation of the face-distortion aftereffect,” *Perception*, vol. 32, pp. 1109–1116, 2003.
- [85] M. A. Webster and O. H. MacLin, “Figural after-effects in the perception of faces,” *Psychonomic Bull. Rev.*, vol. 6, pp. 647–653, 1999.
- [86] M. A. Webster, D. Kaping, Y. Mizokami, and P. Duhamel, “Adaptation to natural facial categories,” *Nature*, vol. 428, pp. 557–560, 2004.
- [87] J. S. Winston, R. N. A. Henson, M. R. Fine-Goulden, and R. J. Dolan, “fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception,” *J. Neurophysiol.*, vol. 92, pp. 1830–1839, 2004.
- [88] R. K. Yin, “Looking at upside-down faces,” *J. Experimental Psychol.*, vol. 81, pp. 141–145, 1969.
- [89] A. Yip and P. Sinha, “Role of color in face recognition,” *Perception*, vol. 31, pp. 995–1003, 2002.
- [90] A. W. Young, D. C. Hay, K. H. McWeeny, B. M. Flude, and A. W. Ellis, “Matching familiar and unfamiliar faces on internal and external features,” *Perception*, vol. 14, pp. 737–746, 1985.
- [91] A. W. Young, D. Hellawell, and D. C. Hay, “Configurational information in face perception,” *Perception*, vol. 16, pp. 747–759, 1987.
- [92] A. W. Young, F. Newcombe, E. H. F. de Haan, M. Small, and D. C. Hay, “Face perception after brain injury: Selective impairments affecting identity and expression,” *Brain*, vol. 116, pp. 941–959, 1993.
- [93] L. Zhao and C. Chubb, “The size-tuning of the face-distortion aftereffect,” *Vision Res.*, vol. 41, pp. 2979–2994, 2001.

ABOUT THE AUTHORS

Pawan Sinha received the B.S. degree in computer science from the Indian Institute of Technology, New Delhi, and the M.S. and Ph.D. degrees from the Department of Computer Science, Massachusetts Institute of Technology (MIT), Cambridge.

He is an Associate Professor of neuroscience in the Department of Brain and Cognitive Sciences, MIT. Using a combination of experimental and computational modeling techniques, research in his laboratory focuses on understanding how the human brain recognizes objects and how this skill is learned through visual experience. He studies individuals with normal developmental histories and also those with neurological disorders such as autism. He has recently launched Project Prakash, a humanitarian and scientific initiative to help treat congenitally blind children in India and also to study how they develop visual skills after sight onset.

Dr. Sinha is a recipient of the Alfred P. Sloan Foundation Fellowship in Neuroscience, the John Merck Scholars Award for research on develop-



mental disorders, and the Jephtha and Emily Wade Award for creative research. He serves on the editorial board of ACM’s *Journal of Applied Perception*. He was named a Global Indus Technovator in 2003. Further information about Dr. Sinha’s lab, along with a forum for further discussion of the ideas presented in this paper, is available at <http://web.mit.edu/bcs/sinha/home.html>.

Benjamin Balas received the B.S. degree in the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge. He is working toward the Ph.D. degree from the same department.

His research interests include the domain of visual concept learning and the representation of moving objects.



Yuri Ostrovsky received the A.B. diploma in computer science from Harvard University, Cambridge, MA, in 1998. He is currently working toward the Ph.D. degree from the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.

His research focuses on the bootstrapping of visual processes through object motion information, as informed through computational modeling, as well as experimental behavioral work with visually impaired patients. He is interested in visual development, object segmentation, and object and face recognition from both the applied and basic research perspectives.



Richard Russell received the B.A. degree in neuroscience from Pomona College, in 1998, and the Ph.D. degree in cognitive science from the Massachusetts Institute of Technology, Cambridge, in 2005.

He is a Postdoctoral Fellow at Harvard University, Cambridge, MA. His research interests include face recognition and prosopagnosia.

