Chapter 2

Cognitive and Computational Approaches to Face Recognition

Alice J. O’Toole

The use of computational models for understanding human face perception and recognition has a long and intriguing history that runs parallel to efforts in the engineering literature to develop algorithms for computer-based face recognition systems. Over the years, intermittent collaborative interactions between computational and psychological approaches have offered numerous insights into the kinds of face representations capable of supporting the many tasks humans accomplish with faces. At the core of the intersection between cognitive and computational approaches is the physical reality of the human face itself, which is a complex three-dimensional object with overlapping pigmentation. Most of the time, we experience human faces in action, moving and deforming in complex and meaningful ways. The face is a primary source of the information we use for recognizing people and for classifying them into visually derived semantic classes (e.g. sex, race, age) (Bruce and Young, 1986). It also provides information useful for assessing a person’s momentary emotional state (e.g. from smiles, frowns, etc.) and for determining their focus of attention (i.e. through gaze direction and head orientation).

Human expertise with faces is defined generally by the fact that we can remember hundreds, if not thousands, of faces as “individuals.” This suggests an ability to extract and encode the information that makes a face unique. The computational challenges involved in this problem are of three sorts. The first is the general inverse optics problem of vision (henceforth, Computational Challenge 1), which is the fact that the visual system has, as input for the task of face recognition, two two-dimensional projections of the three-dimensional world—one image on each retina. The neural code for faces, and indeed, the rest of the visual world, must either recreate the lost third dimension (Marr, 1982) or must do without it (Gibson, 1966). Both strategies, by definition, force the visual system to rely on an error prone, or at least limited, estimate of the real visual world. The dissociable information at the core of this problem can be illustrated using data from a laser scan of a face (Figure 2.1), which separates the three-dimensional shape and reflectance/pigmentation in the face. Any unique combination of shape and reflectance information, (e.g. from an individual face) can produce a nearly limitless variety of two-dimensional retinal images, depending on the illumination conditions and the relative position of the viewer with respect to the head. This myriad of images must somehow be mapped to a single set of physical realities about the structure and reflectance properties of a face.

The second computational challenge is that, regardless of which strategy the neural system has evolved to deal with its limited quality sample of the world, it must ultimately encode and quantify the complex information in faces that we need to survive—the uniqueness of individual faces, the properties that specify age, sex, and race, and the social and emotional communication information faces convey (Computational Challenge 2). Thus, the single set of physical realities about the structure and reflectance properties of a face must be sufficient for the face processing tasks we do. Assuming that the visual system is capable of computing a reasonably veridical set of
measurements of faces, there is still the problem of finding the task-relevant information. Quantifying the complex variations in the three-dimensional shape and reflectance of a human face is by no means trivial. Again let's look at Figure 2.1. On the left, we see a laser scanner sampling the surface of a face. The output of the scan produces $512 \times 512$ (262,144) shape measurements expressible in $x, y, z$ coordinates (Figure 2.1, upper right). The reflectance information is equally voluminous with the same number of samples expressed in $r, g, b$ coordinates (somewhat analogous to the visual systems' filtering through the three cone channels) (Figure 2.1, lower right). From this complex input, which the human visual system likely processes in a highly different form, we are able to identify a "hooked nose," wide-set eyes, and a "perky" face.

The third computational challenge is that the "unique" information that specifies face identity does not exist in absolute terms, but rather is dependent on a reference population of relevant faces (Computational Challenge 3). A nose of a certain type has diagnostic value for identifying a person, only to the extent that is "unusual" within the context of a population of faces. A neural code for faces will be most efficient if it references the amount and type of variability in a relevant population of human faces. Some element of this population referencing is needed to account for several phenomena in human face recognition, including the recognition of other-race faces. In the case of other-race face recognition, the appropriate prototype or reference face is different and our limited experience with other-race face populations may constrain the quality of the representation we can create.
These three computational challenges apply to understanding human face representations from a neural and psychological point of view and are at the core of the productive dialog that has taken place between psychologists and computer vision researchers over the past two decades. Neuroscientists have come into this dialog more recently as functional neuroimaging studies have uncovered some of the basic structure of the neural organization of face processing. A further opportunity for dialog between neuroscientists and psychologists has arisen with findings of high-level visual after-effects for faces elicited with perceptual adaptation methods (Leopold et al., 2001; Webster and MacLin, 1999). We will return to this approach at the end of the chapter.

In this chapter, we will consider the insights gained from combining computational and cognitive approaches to the study of human face recognition. We will discuss the ways in which computational models have informed studies of human face processing and vice versa and how these interactions may be directed in the future to push progress in both fields forward. The chapter is organized as follows. We begin with the concept of a face space, in its abstract, psychological, and physical/computational forms. Next, we consider how physical face spaces have evolved via increasingly sophisticated kinds of inputs that alter the predictions we make about face processing as a function of the topography of these spaces. Different inputs change the nature of the space and its suitability as a model of human face representations. We will then look briefly at how adaptation as a method is beginning to reveal properties of neural representations in a way that connects with the cognitive/perceptual approach. Finally, we will discuss recent progress in state-of-the-art computational models of face recognition, which offer us a new perspective on the cognitive-computational dialog.

I will concentrate primarily on computer models of facial identity recognition, but readers can find information on computer recognition of other facial cues (e.g. facial expressions) in chapters by Cottrell and Hsiao (Chapter 21); Stewart-Bartlett and Whitehill (Chapter 25); and Calder (Chapter 22) elsewhere in this volume. Additional related studies of computational models of face processing for identification can be found in Chapter 20 by Vetter and Walker.

**Representing faces**

**Measuring the information in a face and representing it in a face space**

The *face space* model, introduced by Valentine (1991), serves as a metaphor for understanding human face representations. It is virtually impossible to explain the interaction between the computational and cognitive approaches to understanding face recognition without reference to this model. It serves as the glue that binds the theoretical and computational aspects of the problem together. The framework can be described simply with a few basic premises. First, faces are defined metaphorically as points in a multidimensional space. Second, the axes of the multidimensional face space define a “feature” set with which individual faces can be encoded. As such, each face has a value on each feature axis, and the combination of these feature values specifies its place in the space. Third, the similarity of two faces in the space can be measured as the Euclidean distance between them.

**Abstract face space**

As introduced by Valentine (1991), the face space was an *abstract* or theoretical construct for understanding some intrinsic factors underlying human face perception. In particular, the effect of face typicality on recognition (Light et al., 1979) and classification (Valentine and Bruce, 1986) are conceptually easy to assimilate in this framework. Typical faces are in the center of a crowded face space, and thus are recognized less accurately than unusual faces, because they are more
easily confused with other faces. Typical faces are also classified as faces faster than unusual faces (Valentine and Bruce, 1986), presumably, because this classification involves a comparison between the object and the face prototype (though see Burton and Vokey, 1998, for a more technical account of a face space with higher dimensionality).

The concept of an abstract face space offers a framework for understanding the importance of Computational Challenge 3—that human face processing skills must be understood in the context of a relevant population of faces. The face space even in this simple abstract form forced psychologists to consider the statistical structure of populations of faces and its implications for human performance. I will argue later in this chapter that this is a challenge that the developers of computer-based face recognition algorithms have yet to appreciate fully.

Psychological face space

A psychological face space represents human judgments of face similarity in a direct, map-like way and can be helpful for understanding the “feature” space that underlies these judgments. To create a psychological face space, a similarity or distance matrix is created from similarity judgments made to pairs of faces. These judgments are averaged over a large number of subjects and are submitted to a multidimensional scaling analysis, often implemented as a simple linear principal components analysis (PCA). The analysis produces a face space, of \( n \) dimensions, with the dimensions ordered according to the proportion of variance explained in the human similarity judgments. A researcher decides the number of dimensions needed to provide a good account of the human similarity data—often as few as three or four. These dimensions can sometimes be interpreted visually, by simply placing individual face images on the graph at their derived locations and observing the differences between the facial traits that land at opposite ends of individual dimensions. Often these “features” include face shape (round faces at one end, thin faces at the other end), facial complexion (dark versus light complexions) and other generally Gestalt-like variations in facial appearance.

This psychological similarity map can be useful in sifting through the inferred feature dimensions that make faces more or less perceptually similar. It can be used also to test the validity of various types of physical face spaces. There are several problems with this approach, however, that have limited its broader use by researchers. First, the axes must be interpreted in an ad hoc way. Second, different subjects vary in the similarity criteria they use for generating these measures, making the technique susceptible to overt strategies applied by subjects ("I will look at the eyes"). Third, and perhaps the most limiting, is that the technique does not provide enough dimensions to capture all or even most of the information we represent from faces. At its best, a psychological face space highlights the role of the statistical structure of the face population (Computational Challenge 3) and offers a technique for inferring the saliency of some of the information we encode about faces (Computational Challenge 2).

Physical face space

A physical face space provides a method for quantifying the information in a set of faces and produces a statistically derived feature set for representing faces. This opens the door for asking whether a particular derived physical face space can account for various aspects of human face processing. A physical face space is created as follows. A set of physical measures is taken on a large set of faces, and encoded in individual vectors (one per face). The measures can include anything quantifiable about a face (e.g. nose length, pixel value, etc.). As with human similarity judgments, a covariance matrix of these physical measurements can be submitted to a PCA to produce a face space, of \( n \) dimensions, with the dimensions ordered according to the proportion of variance explained in the similarity matrix. These axes are now tangible, physical measures of the information in the set of faces from which they are derived.
This physical space represents the statistical structure of the similarity among faces in the set that depends directly on the types of measures made at the outset and on the variability of the population of faces analyzed. Although the interpretation of individual axes in the space must be made visually, a researcher has one additional interpretation tool for a physical face space that is not available with a psychological face space. Because the space is linear, and because the information is directly quantifiable from the face, it is usually possible to alter individual faces by tweaking the feature values within a face and re-synthesizing the face. By looking at how these manipulations change the face, often a guess about the nature of the information represented by the feature axes is possible.

The use of physical face space models to understand human face recognition has progressed over time with the creation of increasingly sophisticated spaces via the analysis of increasingly sophisticated face quantification schemes. In this way, computational face modeling progress has been front-loaded. By this, I mean that most of the insights we have gained have come from changes in what goes into the space rather than from more sophisticated ways of deriving the space. In fact, most researchers in both psychology and in some aspects of computer vision still use the simplest and least assumption-laden method available, PCA. In what follows, for simplicity, I will use PCA as a place-holder for any multidimensional space-creating analysis. Bear in mind that anything that works to create a graphic multidimensional space from the stimuli is an option.

Before proceeding, it is perhaps worth noting that the construct of the physical face space offers a basic framework for approaching Computational Challenges 2 and 3, but not Computational Challenge 1—the inverse optics problem. For the former two challenges, the face space lays theoretical groundwork for understanding the nature of the information in faces and how it relates to face populations as a whole. It does not directly address questions of how to overcome the inverse optics challenge to face recognition, which is tied to the specifics of the representation. In the next section, we consider the importance of what goes into the space, how it has changed over time, and the insights these changes have offered into understanding human face processing.

Feature quantification systems: input to the physical face space models

To make a long story short, in the era after Marr’s (1982) classic book *Vision*, computer vision researchers were intent on extracting (or computing) representations of the three-dimensional structure of objects. This goal filtered into the first attempts to formalize computational approaches to face recognition. At a minimum, this required measuring the information in faces in a way that stayed true to the veridical three-dimensional, object-centered, invariant, attributes of the face—a goal that turned out to be unrealistic for both objects and faces given the technology and algorithms of the time. The post-Marr era began roughly with Poggio and Edelman’s (1990) argument that extracting three-dimensional object-centered representations might not be necessary for object recognition. This approach was bolstered by psychological findings that indicated some degree of view-dependence in human performance for recognizing objects and faces (Bülthoff and Edelman, 1992; O’Toole et al., 1998a; Troje and Bülthoff, 1996; though see Biederman, 1987). The underlying assumption was that view-dependent object representations should yield view-dependent human performance. In hindsight, it is clear this assumption was overly simplistic. In particular, the assumption does not consider what happens to a face representation as a face becomes familiar (Jiang et al., 2009; Wallis and Bülthoff, 1999). Notwithstanding, computational trends and human object and face recognition data spurred the use of simpler image-based inputs, although at the beginning the use of these codes for psychological models was controversial.
In perspective, it is worth remembering that there are important differences between object and face recognition. Poggio and Edelman (1990), for example, explored computationally how much object recognition ("That is a chair") could be done using viewer-centered, image-based measurements. Object recognition researchers were mostly concerned with solving the problem of recognizing objects (classifying exemplars into object categories) over changes in viewpoint. Face recognition researchers, on the other hand, were more concerned with finding a recognition code that would capture enough of the subtle information in faces to quantify the uniqueness of individual faces ("That is Shimon"). Given that all faces have the same set of features, arranged in roughly the same configuration, it seemed evident that face codes would have to capture (implicitly or explicitly) more than a few simple discrete features. This provided an even stronger push in face recognition for assuming that an analysis of facial images might provide a good starting point for input to a face space.

Image-based representations

The first psychologically-motivated image-based face recognition model appeared under the name of auto-associative networks (O'Toole et al., 1988). Storing face images in an auto-associative memory allowed for content addressable retrieval of faces and used distributed storage mechanisms (i.e. individual face memories shared the same coding space rather than being assigned to different memory locations). In particular, this face code allowed generalization of recognition between images that varied in spatial frequency content (O'Toole et al., 1988). The paper was published in one of the first issues of the journal Neural Networks and was of interest to neural network researchers because it made use of Kohonen's (1984) auto-associative memory model. It further drew on Kohonen's use of face images to demonstrate that associative memory models could fill in missing information from occluded, blurred, or partial images previously stored in the memory. Of note, Kohonen (1984) made clear the link between auto-associative memory models and PCA (eigenvector analysis), which had not been explored in any detail within the context of an image memory application.

The second reason our paper ended up in Neural Networks was because it kept getting rejected from psychology journals, with reviewers expressing serious reservations about the use of image-based codes in a model of human face processing. I was sympathetic to these reviews, but I still thought that the model had some intriguing psychological properties that were worth exploring. First of all, it worked really well as a content addressable memory, having lossless encoding and relatively high-capacity storage. In other words, you could store many faces in the memory with no quality loss, using a single area of memory that shared resources. Thus, it did not force localized or separate storage of each face in a different place. Second, you could retrieve the stored images from "cue images" that were occluded, blurred, or otherwise frequency-filtered. Third, you did not have to do feature selection—a psychological perk that seemed to resonate with configural face codes. Fourth, retrieval of a face image did not require a serial search through a database of stored face images and the time required did not vary as a function of the number of images in the memory.

Although psychologists had reservations, computer vision researchers embraced the practical advantages of a simple, image-based PCA for face recognition applications. In a highly influential paper, Turk and Pentland (1991) proposed an image-based eigenanalysis (PCA) model for face recognition. Zhao et al. (2003), in their review of computational face recognition algorithms, describe Turk and Pentland's (1991) work as "the first really successful demonstration of machine recognition of faces" (p. 412). In addition to the fact that the image-based PCA could do face recognition better than it had been done before, other work suggested that the representation it used was amenable to being dissected and analyzed in interesting ways. Sirovich and Kirby (1987)
had explored low-dimensional representations of faces in PCA, using only PCs/eigenvectors that explained large proportions of variance in the image set. O'Toole et al. (1993) found that PCs explaining large amounts of variance were better for face categorization tasks (e.g. by sex), whereas PCs explaining smaller amounts of variance were better for recognizing faces. This made intuitive sense, because the information that makes a face male or female is relevant for all faces in a population, whereas the information that makes individual faces unique is likely to explain little variance in a population of faces. It also gave us insight into understanding how the complex information in faces that could be used for different tasks (e.g. recognition, classification) could be dissociated.

In short, it was possible with this model to examine in some detail, the importance of computationally extracted features for different face processing tasks. With relatively little effort, this image-based version of a physical face space could be linked to some basic phenomena in the psychology of face perception, including the perception of face gender (Abdi et al., 1995; O'Toole et al., 1998b), and recognition of own- versus other-race faces (O'Toole et al., 1994).

The most critical problem with the pure image-based eigenface analysis was the fact that it required averaging face images without regard to the correspondence of feature locations. The correspondence problem for faces involves finding the locations of landmark features (e.g. tip of the nose, etc.) in all face images and coding these locations explicitly. This is easy to do by eye but challenging to do automatically. The problem with image-based eigenanalysis was quickly apparent to both psychologists and computer vision researchers. Because of its practical value, however, and the ease of roughly aligning and scaling face images (i.e. relative to difficulty of solving the correspondence problem), pure image-based models have persisted longer in machine-based face recognition algorithms (Zhao et al., 2003) than in psychological models of face recognition.

Two-dimensional morphable models

Researchers from both psychology and computer vision worked on going to the next step. From a computational perspective, Craw and Cameron (1991) were able to eliminate deviations from the average face shape from individual faces before submitting the faces to PCA. These “shape-free” faces were made by: (1) locating (by hand) a set of landmark feature points on each of a set of face; (2) averaging the locations of these points over a set of faces; and (3) morphing individual face images into the average shape. A PCA of these shape-free faces captures the image-based information in faces, free of the artifacts inherent in combining misaligned images. Unfortunately, a PCA on shape-free faces is also free of information about face configuration. More precisely, each face is morphed into the configuration of the average face, so that the features (nose, mouth, etc.) align with the features of the average face. Because facial configuration in this model does not vary across the set of faces, it loses its diagnostic value for identification. This is surely overkill for a psychological model of face perception. Hancock and colleagues (1996) added back the configural information, but in a form that allowed for separate shape and shape-free face image codes. The shape representation came from information about the two-dimensional deformation of individual faces (i.e. their landmark point locations) from the average face. The shape-free “texture” information came from the corresponded shape-free face. Next, they analyzed the shape and shape-free information in separate PCAs. The PCA predictions based on shape and shape-free face analyses dissociated hits and false positives in a face recognition task performed by humans. Hancock et al. (1996) found that the shape-free information predicted false positives and the shape information predicted hits. There was also evidence indicating that PCs explaining larger amounts of variance were generally more predictive of false positives. This physical framework made strong progress in modeling the effects of typicality on face recognition and in isolating the kinds of information in faces that contribute to different errors.
It is worth noting that two-dimensional morphable representations are also at the core of facial caricaturing. Caricaturing had long been possible using two-dimensional shape spaces that dragged the image points (pixel values) along with the shape deformation (cf. Rhodes, 1996 for a complete review), but did not alter pixel values. Benson and Perrett (1991), for example, were able to make precise photorealistic caricatures by two-dimensional shape deformations applied to photographs. They found that magnitude of the caricature advantage for human subjects recognizing faces correlated with the familiarity of the faces and with the quality of the caricaturing process as judged by caricature experts. They also showed an advantage for the speed of naming faces that were caricatured. Lee et al. (2000) found that this type of multidimensional face model accounted well for identification accuracy and distinctiveness ratings of veridical, caricatured, and anti-caricatured faces. The work of Lee and Perrett (1997) took this caricaturing process one step further to include caricaturing the color values in face images—a method that bridges to a more complete face representation.

It is worth pausing to consider the differences between the two-dimensional “semi-corresponded” face-space model and the pure image analysis. Semi-corresponded models code the shape, but not reflectance, in terms that relate individual faces to the average face. One difference is that it is possible to morph selectively through the part of the space that encodes the two-dimensional configural shape (i.e. via the facial landmark position codes). In the pure image-based face space, morphing can produce images that “leave the face space.” This means that when there are correspondence artifacts or errors, a face with “two noses” or “two chins” can be synthesized in the morphing process. In the pure shape space, only valid faces lie on the trajectories between any two faces. A second difference is that the average of the shape space is necessarily a meaningful reference for all other faces, because the faces are defined as deformations from the average. The representation of a face’s shape in this space is a “point in the multidimensional space” (as in the image space), but one that is directly connected to the average face in the space. Concomitantly, the inclusion of the shape-free space also retains a measure of the image-based reflectance (although one that is not inherently dissociable from viewing parameters, e.g. illumination). The representation of a face’s reflectance in this space is a point, similar to the point representation used in the image space (i.e. corresponded, but not linked formally to the average). Although it may look like a subtle difference, the Lee and Perrett (1997) model adds a “connected-to-the-average morphability” feature for color information. This is an important, but under-utilized, method in two-dimensional morphable spaces.

The idea of a dual shape- and shape-free space was an enormous improvement over a pure image-based space. One drawback in its use for computational modeling, however, was the work-intensive pre-requisite involved in hand-locating enough landmark points (more than 30) in enough faces (over 100) to make a meaningful and clean morph in a face space that represented a sufficiently diverse, “psychologically interesting,” population of faces. Although computer graphics researchers were making progress on the problem of automating the correspondence problem for faces (cf. Beymer et al., 1993; Lanitis et al., 1995; Vetter and Poggio, 1997), these techniques were sufficiently complex and technically demanding that they did not filter easily into psychology labs.

Three-dimensional morphable models

A completely automated solution to the correspondence problem for laser scan data on a large number of faces was put forth in 1999 by Blanz and Vetter (cf. Chapter 20, this volume, by Vetter and Walker). The laser scans they analyzed contained a direct and dense measure of the three-dimensional surface and a similarly dense measure of the point-for-point overlay of reflectance information. This latter measure did not confound viewing parameters, because of the use of an
ambient light source by the scanner. Blanz and Vetter (1999) took several impressive steps forward from what had been accomplished previously. First, correspondence was established on all of the roughly quarter of a million head sample points, rather than on a subset of facial landmark points. Second, this corresponded representation was the first truly object-centered representation of faces analyzed in a face space (no light required). Third, the space was completely morphable, either in a way that cohesively combined the three-dimensional shape and reflectance information, or in a way that separated the space and reflectance subspaces. Fourth, because the face is represented in object-centered coordinates, face variations produced by morphing between faces, or by morphing in arbitrary directions in the space, can be viewed (with computer graphics) from any viewpoint and under any illumination condition (by putting back the light).

This space offered psychologists a myriad of options for manipulating the information in faces in precise and interesting ways. The representation of a face’s shape and reflectance in this space is linked directly to the average face (with shape and reflectance included). An individual face’s location in the space, therefore, is not just a point, but rather a trajectory that is defined by the line originating at the average face and continuing to the location of the individual face in the space. It is, therefore, the direction of this vector in the space that defines a face’s identity trajectory.

In principle, any morphable space is useful for graphically altering human faces. The Blanz and Vetter (1999) space is an object-centered version of the more common viewer-centered two-dimensional morph spaces (Benson and Perrett, 1991; Hancock et al., 1996). In what follows, I will focus primarily on psychological findings that have emerged from work using the three-dimensional morphable model, incorporating data from two-dimensional morphable manipulations where relevant. Bear in mind, however, that the two-dimensional morphable model operates in an analogous viewer-centered way for shape.1

In the three-dimensional morphable space, the distinctiveness of faces can be manipulated easily by creating caricatures and anti-caricatures that incorporate both shape and reflectance information simultaneously. The caricature is made by stretching a face vector away from the average, while retaining the direction of its identity trajectory. Notably, caricatures can be made either with respect to the three-dimensional shape and/or reflectance. Caricaturing only the three-dimensional information in a face can age it decades, by enhancing slight creases in young faces into deep wrinkles (O’Toole et al., 1997; see Figure 2.2). This finding suggests that a part of information for face age is person-specific, rather than based on a general aging routine. Further, the finding complements Burt and Perrett’s (1995) demonstration of aging a face in a two-dimensional morph space. They morphed younger faces in the direction of the average of the older faces and found a role for an aging algorithm that generalizes across individuals.

The importance of shape versus reflectance information has been tested also for face recognition using manipulations of faces in the three-dimensional morphing space. O’Toole et al. (1999) measured human recognition performance for a set of faces that varied only in shape and a set of faces that varied only in reflectance. These stimuli were created by morphing individual reflectance maps onto the shape average (shape-normalized faces) and by morphing the average reflectance map onto the shape of the individual faces (reflectance-normalized faces). Human recognition performance benefited from both shape and reflectance information. Recognition performance for the original faces, which varied both in shape and reflectance, was approximately equal to the sum of performance for the faces that varied only in shape and the faces that varied only in reflectance.

---

1 One exception is Lee and Perrett’s (1997) model, which implements reflectance codes also in terms of the face average.
Fig. 2.2 Caricaturing only the three-dimensional shape information in a face ages the face. The original scan is of a man in his early thirties (second from the left). An anti-caricature appears to the left of the original and three increasingly exaggerated caricatures appear to the right.

By far the weirdest invention of the Blanz and Vetter (1999) space has been the “anti-face” or face opposite (Figure 2.3; see also Rhodes and Leopold, Chapter 14, this volume, for a face space illustration). A face’s anti-face is made in this space by reversing the direction of its identity trajectory, to pass back through the average face and continue along the line, roughly an equal distance in the opposite direction. This arrives at a point in space where the face’s values on each of the many multidimensional axes in the space are inverted (opposite) to their original values (Blanz et al., 2000). I half-remember the first time Volker Blanz showed me an anti-face, almost incidentally. He said something like “Just set the caricature value on the morpher model to –1.” What could be simpler? The anti-face was not designed into the morpher, it was just numerically there—on the other side of the mean. It can exist meaningfully only in a space where faces are in correspondence with the average. Although cognitively we understand that the face-anti-face pairs are a kind of opposite (round faces become elongated, dark complexions become light), there is a perceptual disconnect when the face crosses the mean and goes over to the other side. Faces retain their identity all along the identity trajectory from the average to the original face, but encounter a strong perceptual break with the original identity on the other side of the mean (Blanz et al., 2000). We found out soon after, however, that this perceptual break did not mean that these faces are disconnected from each other in the representational space. This connection became apparent in the form of high-level visual adaptation for faces.

Adaptation and the organization of the face space around the prototype

In recent years, one of the most fruitful methods for studying human face representation has been based on findings that face perception is susceptible to aftereffects elicited by adaptation.

Fig. 2.3 A face (left) and its anti-face opposite appear strongly dissimilar. The original face is thin and elongated with a dark complexion. The anti-face has a round shape with a light complexion.
The effects are much the same as other lower-order visual aftereffects (e.g. color; Hering, 1878). Because the face adaptation literature is covered in detail elsewhere in this book (Rhodes and Leopold, Chapter 14), I will mention just a few effects and will focus on how these effects relate to the computational constructs presented in this chapter. The classic finding of Webster and MacLin (1999) demonstrated that adaptation to a grotesquely expanded or contracted face biases the perception of a subsequently presented normal face to appear distorted in the opposite direction. In other work, Leopold et al. (2001) found that adaptation to an anti-face biases the perception of an identity-neutral “average” face to appear as the original face. Further findings by Webster et al. (2004) showed face aftereffects for gender, race, and expression. All of these results suggest that the representations of faces that are quite distant in the perceptual and neural face space are strongly interconnected.

The opponent direction of the perceptual bias suggests the importance of a prototype or average face in the representation of individual faces. Thus, it seems that the feature values of an individual face in the space are defined with respect to average values across a relevant population. These findings fit comfortably into the framework of a morphable space, where “deformation from the average” is the basic unit of analysis, and where it is possible to reset or move the average in the short term, based on experience or exposure to a particular type of face. Again, the connectedness of faces in the space, and their codependence in coding is revealed in a compelling way by the fact that perceiving a particular face can influence one’s subsequent perception of a highly dissimilar face, far away in the space.

Adaptation, and the psychological questions that have been addressed with this technique, rest heavily on predictions formulated in the computational construct of a face space. These findings have opened the door to addressing questions about the way faces are coded neurally. Numerous studies have demonstrated that face adaptation shows tolerance to two-dimensional affine transformations in the size, orientation, and retinal position of a face (Anderson and Wilson, 2005; Jeffrey et al., 2006; Leopold et al., 2001; Rhodes et al., 2003; Watson and Clifford, 2003; Zhao and Chubb, 2001). This eliminates low-level retinotopic visual processing areas as the neural locus of the effects. Moreover, recent studies using functional neuroimaging (Loffler et al., 2005) and neurophysiological methods (Leopold et al., 2006) support a prototype-centered organization of face representations. Both studies show a stronger neural response for faces distant from the average face.

Returning to questions about the nature of the information captured by the human face representations, using stimuli created with the three-dimensional morpher, Jiang et al. (2006) showed that identity adaptation effects can be elicited by faces that are opposites either in shape only or in reflectance only. This suggests that both shape and reflectance are encoded in this prototype-centered space. Further, Jiang et al. (2006) and Jeffrey et al. (2006) found partial transfer of face aftereffects across three-dimensional viewpoint change. Incorporating in the question of how face representations change with familiarity, Jiang et al. (2007) showed that the magnitude and viewpoint transferability of identity adaptation increases with the familiarity of a face. This suggests the exciting possibility that adaptation can track the development of individual face representations as they evolve with experience.

On the question of Computational Challenge 1, one can ask “To what extent do humans build a representation that includes inherently three-dimensional information about faces?” This stands as an alternative or complement to building a representation that incorporates multiple viewer-centered templates. Jiang et al. (2009) showed evidence for the inclusion of some inherently three-dimensional information in face representations when the faces were learned from multiple viewpoints. The strategy Jiang et al. applied for getting at inherently three-dimensional information involved testing the generality of information learned from one three-dimensional-refencing
transformation (viewpoint change) for facilitating recognition across another three-dimensional-refencing transformation (illumination change). Identity adaptation over illumination change was facilitated for faces learned from multiple views, suggesting that learning a face from one type of three-dimensional transformation (e.g. viewpoint change) can benefit perceptual robustness to a different type of three-dimensional representation (e.g. illumination change). This provides a complementary role for three-dimensional codes to the well-established viewer-centered codes.

Although these results have been informative about the face representations that govern perception, more work needs to be done on the computational philosophy that connects adaptation findings to neural processing mechanisms. Non-opponent based explanations have been offered for other adaptation effects involving social attention cues, such as eye gaze (Calder et al., 2008) and body orientation (Lawson et al., 2009). This approach should be explored for other kinds of adaptation effects as well.

**Humans versus machines**

In the final part of this chapter, we will run a quick "reality check" on the progress of computational face recognition models and offer some observations about the future of the dialog between psychology and computer vision on face processing. It used to be possible for a psychologist (like me) to sit down and implement a simple version of a computational face recognition model and then run off and test some of the predictions the model made about human face processing. I could look at the nature of the representations used in a computational formulation of the problem and make some predictions about performance pitfalls or advantages that might go along with these representations. If human and machine performance were similar, I could infer that some properties of the computational model apply also to human representations, (and so on). Those days are gone. State-of-the-art computational models have come into their own as commercial products. They are now built from complex plug-in components, and solve a sequence of problems that span from getting a face out of the image, to running the information through preprocessing routines, to delivering a final response. They are also quite a bit more accurate than many of us would have predicted, even a few years ago.

United States government-sponsored evaluations, spanning more than a decade, have gauged the progress of these algorithms. In recent tests (the Face Recognition Grand Challenge, FRGC, and Face Recognition Vendor Test-2006, FRVT-06; Phillips et al., 2010), algorithms in one part of the competition were required to match face identity in over 128 million pairs of face images. The images comprising each pair varied substantially in illumination and the problem was quite challenging for algorithms, relative to tests where the illumination was controlled and matched.

We compared humans to algorithms on this FRGC test using a subset of computationally-defined "easy" and "difficult" face pairs (O'Toole et al., 2007a). In a head-to-head comparison on identical image pairs, three of the seven algorithms were more accurate than humans matching identity on the difficult face pairs. For the easy face pairs, six of seven algorithms were more accurate than humans. We replicated this result on two additional datasets used in the FRVT-06 tests. On the problem of matching identity in pairs of frontal face images, the best algorithms are now in range of human performance—though I will add some caveats to the generality of this claim shortly.

The next question we asked was whether the algorithms and humans were performing the task in qualitatively similar ways (i.e. were they making the same errors). In days gone by, we could have looked at the algorithm strategy to see what predictions it made for human accuracy. Now, however, the proprietary nature of many of the algorithms made this impossible. So, we decided
to computationally fuse the algorithm and human identity match responses together to see if the combination would improve performance relative to the humans or algorithms operating alone (O'Toole et al., 2007b). Our rationale was that fusing together similar strategies results in only minimal performance improvements, whereas the fusion of highly different strategies will improve performance substantially. In fact, fusing the seven algorithms cut the error rate to about half of the error rate for best-performing algorithm alone—this suggested that the algorithms were not all qualitatively similar. Fusing humans with the seven algorithms pushed performance to nearly perfect, indicating that human strategies are at least partially different from the algorithms' strategies.

Finally, we looked to see if state-of-the-art algorithms from the FRVT-06 showed an “other-race effect” (O'Toole et al., 2008). First, we divided the algorithms into East Asian Algorithms (n = 5, originating in China, Korea, and Japan) and those originating in Western countries (n = 8, France, Germany, and the United States). Next, we created an East Asian and Western fusion algorithm by fusing together the identity match estimates from the five Western algorithms and the eight East Asian algorithms, separately. In two separate tests, we found evidence for an other-race effect in the performance of the East Asian and Western fusion algorithms matching identity in the Caucasian and East Asian face pairs. We surmise (but, cannot know for sure) that this other-race effect is due to differences in the “experience” of the algorithms. Experience in this context refers to the ethnic composition of the set of training faces used by the algorithms to extract a feature set for encoding the faces. This would make the root cause of the problem for algorithms similar to that of the human other-race effect. So, things now seem to come full circle, as we find ourselves looking to human performance to understand what the algorithms are doing. Maybe that’s progress.

Before closing this section, I offer some caveats on the reality test. First, in these human–machine comparison tests, humans matched identity in pairs of unfamiliar faces. This is fair as a test of what security guards do, but does not represent human performance at its best. Humans are at their best with familiar faces (see Burton and Jenkins, Chapter 15, this volume). Indeed, we would expect human performance to exceed algorithm performance if the faces were familiar. Second, although there are stark changes in illumination between the images in the face pairs, the conditions under which algorithm performance surpassed human performance involved identity comparisons between images that did not vary in viewpoint, (specifically, frontal images). In fact, for algorithms, the inverse optics problem is still far from solved (Computational Challenge 1). Further, although it is clear that the algorithm developers have spent a great deal of effort working on Challenge 1, the other-race effect findings suggest that they have under-appreciated the complexities added when the task must be carried out in the context of a relevant face population, which may vary across locations (airports) and even by time of day (the Frankfurt and Tokyo flights arrive at 8am, the New York and Mumbai flights at 9 am, and so on).

**Summary and conclusions**

In summary, the use of computational models for understanding human face perception and recognition runs parallel to efforts in the engineering literature to develop algorithms for computer-based face recognition systems. Collaborative interactions between computational and psychological approaches to face recognition have offered numerous insights into the kinds of face representations capable of supporting the many tasks humans accomplish with faces. Combined with findings in functional neuroimaging and high-level face adaptation studies, computational formulations have increased our understanding of the complex tasks humans solve when they recognize faces. Beyond the expectations of many psychologists, state-of-the-art
face recognition algorithms are now more accurate than humans on some challenging face recognition tasks. Moreover, recent studies show that optimal combinations of human and machine recognition judgments can improve recognition performance over that possible by either machines or humans alone. As machine recognition improves, however, there is clear evidence that they begin to show some of the weak points of human perception—such as the “other-race effect.”

The computational challenges in face recognition have focused researchers in psychology and neuroscience on the problem of formulating theories of representation that meet these computational challenges. The face space framework, in all of its implementations, has productively bootstrapped this research. Although the dialog between cognitive and computational approaches has been active now for nearly two decades, important challenges remain. One part of face space framework that remains under-formulated is the gap between viewer- and object-centered representations. Even if human representations are primarily viewer-centered, there is still a need to link representations in ways that make them robust to changes in the viewing conditions. This leads to the second open issue—that of how representations change as faces become familiar. The face space formulation, as it stands, does not support representations that are more or less enriched by experience. These questions may begin to play a prominent role in the computational-cognitive dialog of the next decade of research.

Acknowledgments

Thanks are due to Andy Calder, Vicki Bruce, Vaidehi Natu, David Raboy, and Dana Roark for helpful comments on a previous version of the chapter. Thanks are also due to the Technical Support Working Group, U.S. DOD, for funding the human-machine comparisons.

References


