To appear in *Face Recognition: Models and Mechanisms*; Academic Press

Face Recognition by Humans

Pawan Sinha*, Benjamin J. Balas, Yuri Ostrovsky, Richard Russell
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139

* Corresponding author <psinha@mit.edu>

The human visual system is remarkably proficient at the task of identifying faces, even under severely degraded viewing conditions. A grand quest in computer vision is to design automated systems that can match, and eventually exceed, this level of performance. In order to accomplish this goal, a better understanding of the human brain's face processing strategies is likely to be helpful. With this motivation, here we review four key aspects of human face perception performance: 1. Tolerance for degradations, 2. Relative contribution of geometric and photometric cues, 3. The development of face perception skills, and 4. Biologically inspired approaches for representing facial-images. Taken together, the results provide strong constraints and guidelines for computational models of face recognition.

Introduction

The events of September 11, 2001, in the United States compellingly highlighted the need for systems that can identify individuals with known terrorist links. In rapid succession, three major international airports, Fresno, St. Petersburg and Logan, began testing face recognition systems. While such deployment raises complicated issues of privacy invasion, of even greater immediate concern is whether the technology is up to the task requirements.

Real-world tests of automated face-recognition systems have not yielded encouraging results. For instance, face recognition software at the Palm Beach International Airport, when tested on fifteen volunteers and a database of 250 pictures, had a success rate of less than fifty per cent and nearly fifty false alarms per five thousand passengers (translating to two to three false alarms/hour per checkpoint). Having to respond to a terror alarm every twenty minutes would, of course, be very disruptive for airport operations. Furthermore, variations such as eyeglasses, small facial rotations and lighting changes, proved problematic for the system. Many other such tests have yielded similar results.

The primary stumbling block in creating effective face recognition systems is that we do not know how to quantify similarity between two facial images in a perceptually meaningful manner.

Figure 1 illustrates this issue. Images 1 and 3 show the same individual from the front and oblique viewpoints, while image 2 shows a different person from the front. Conventional measures of image similarity (such as the Minkowski metrics (Duda and Hart, 1973)) would rate images 1 and 2 to be more similar than images 1 and 3. In other words, they fail to generalize across important and commonplace transformations. Other transforms that lead to similar difficulties include lighting variations, aging and expression changes. Clearly, similarity needs to be computed over attributes more complex than raw pixel values. To the extent that the human brain appears to have figured out which facial attributes are important for subserving robust recognition, it makes sense to turn to neuroscience for inspiration and clues.

Figure 1. *Most conventional measures of image similarity would declare images 1 and 2 to be more similar than images 1 and 3, even though both members of the latter pair, but not the former, are derived from the same individual. This example highlights the challenge inherent in the face recognition task.*

Work in the neuroscience of face perception can influence research on machine vision systems in two ways. First, studies of the limits of human face recognition abilities provide benchmarks against which to evaluate artificial systems. Second, studies characterizing the response properties of neurons in the early stages of the visual pathway can guide strategies for image pre-processing in the front-ends of machine vision systems. For instance, many systems use a wavelet representation of the image that corresponds to the multi-scale gabor-like receptive fields found in the primary visual cortex (DeAngelis et al, 1993, Lee, 1996). We describe these and related schemes in greater detail later. However, beyond these early stages, it has been difficult to discern any direct connections between biological and artificial face-recognition systems. This is perhaps due to the difficulty in translating psychological findings into concrete computational prescriptions.

A case in point is an idea that several psychologists have emphasized - that facial configuration plays an important role in human judgments of identity (Bruce and Young, 1998; Collishaw and Hole, 2000). However, the experiments so far have not yielded a precise specification of what is meant by 'configuration' beyond the general notion that it refers to the relative placement of the different facial features. This makes it difficult to adopt this idea in the computational arena, especially when the option of using individual facial features such as eyes, noses and mouths is so much easier to describe and implement. Thus, several current systems for face recognition and also for the related task of facial composite generation (creating a likeness from a witness description), are based on a piecemeal approach.

As an illustration of the problems associated with the piecemeal approach, consider the facial composite generation task. The dominant paradigm for having a witness describe a suspect's face to a police officer involves having him/her pick out the best matching features from a large collection of images of disembodied features. Putting these together yields a putative likeness of the suspect. The mismatch between this piecemeal strategy and the more holistic facial encoding scheme that may actually be used by the brain can lead to problems in the quality of reconstructions as shown in figure 2. In order to create these faces, we enlisted the help of an individual who had several years of experience with the IdentiKit system, and had assisted the police department on multiple occasions for creating suspect likenesses. We gave him photographs of fourteen celebrities and requested him to put together IdentiKit reconstructions. There were no strict time constraints (the reconstructions were generated over two weeks) and the IdentiKit operator did not have to rely on verbal descriptions; he could directly consult the images we had provided him. In short, these reconstructions were generated under ideal operating conditions.

Figure 2. *Four facial composites generated by an IdentiKit operator at the authors' request. The individuals depicted are all famous celebrities. The operator was given photographs of the celebrities and was asked to create the best likenesses using the kit of features in the system. Most observers are unable to recognize the people depicted, highlighting the problems of using a piecemeal approach in constructing and recognizing faces. The celebrities shown are, from left to right: Bill Cosby, Tom Cruise, Ronald Reagan and Michael Jordan.*

The wide gulf between the face recognition performance of humans and machines suggests that there is much to be gained by improving the communication between human vision researchers on the one hand, and computer vision scientists on the other. This paper is a small step in that direction.

What aspects of human face recognition performance might be of greatest relevance to the work of computer vision scientists? The most obvious one is simply a characterization of performance that could serve as a benchmark. In particular, we would like to know how human recognition performance changes as a function of image quality degradations that are common in everyday viewing conditions, and that machine vision systems are required to be tolerant to. Beyond characterizing performance, it would be instructive to know about the kinds of facial cues that the human visual system uses in order to achieve its impressive abilities. This provides a tentative prescription for the design of machine-based face recognition systems. Coupled with this problem of system design is the issue of the balance between hard-wiring and on-line learning. Does the human visual system rely more on innately specified strategies for face processing, or is this a skill that emerges via learning? Finally, a computer vision scientist would be interested in knowing whether

there are any direct clues regarding face representations that could be derived from a study of the biological systems. With these considerations in mind, we shall explore the following four fundamental questions in the domain of human vision.

1. What are the limits of human face recognition abilities, in terms of the minimum image resolution needed for a specified level of recognition performance?
2. What are some important cues that the human visual system relies upon for judgments of identity?
3. What is the timeline of development of face recognition skills? Are these skills innate or learned?
4. What are some biologically plausible face representation strategies?

1. What are the limits of human face recognition skills?

The human visual system (HVS) often serves as an informal standard for evaluating machine vision approaches. However, this standard is rarely applied in any systematic way. In order to be able to use the human visual system as a useful standard to strive towards, we need to first have a comprehensive characterization of its capabilities.

In order to characterize the HVS's face recognition capabilities, we shall describe experiments that address two issues: 1. how does human performance change as a function of image resolution? and 2. what are the relative contributions of internal and external features at different resolutions? Let us briefly consider why these two questions are worthy subjects of study.

The decision to examine recognition performance in images with limited resolution is motivated by both ecological and pragmatic considerations. In the natural environment, the brain is typically required to recognize objects when they are at a distance or viewed under sub-optimal conditions. In fact, the very survival of an animal may depend on its ability to use its recognition machinery as an early-warning system that can operate reliably with limited stimulus information. Therefore, by better capturing real-world viewing conditions, degraded images are well suited to help us understand the brain's recognition strategies.

Many automated vision systems too need to have the ability to interpret degraded images. For instance, images derived from present-day security equipment are often of poor resolution due both to hardware limitations and large viewing distances. Figure 3 is a case in point. It shows a frame from a video sequence of Mohammad Atta, a perpetrator of the World Trade Center bombing, at a Maine airport on the morning of September 11, 2001. As the inset shows, the resolution in the face region is quite poor. For the security systems to be effective, they need to be able to recognize suspected terrorists from such surveillance videos. This provides strong pragmatic motivation for our work. In order to understand how the human visual system interprets such images and how a machine-based system could do the same, it is imperative that we study face recognition with such degraded images.

Fig. 3. *A frame from a surveillance video showing Mohammad Atta at an airport in Maine on the morning of the 11th of September, 2001. As the inset shows, the resolution available in the face region is very limited. Understanding the recognition of faces under such conditions remains an open challenge and motivates the work reported here.*

Furthermore, impoverished images serve as 'minimalist' stimuli, which, by dispensing with unnecessary detail, can potentially simplify our quest to identify aspects of object information that the brain preferentially encodes.

The decision to focus on the two types of facial feature sets – internal and external, is motivated by the marked disparity that exists in their use by current machine-based face analysis systems. It is typically assumed that internal features (eyes, nose and mouth) are the critical constituents of a face, and the external features (hair and jaw-line) are too variable to be practically useful. It is interesting to ask whether the human visual system also employs a similar criterion in its use of the two types of features. Many interesting questions remain unanswered. Precisely how does face identification performance change as a function of image resolution? Does the relative importance of facial features change across different resolutions? Does featural saliency become proportional to featural size, favoring more global, external features like hair and jaw-line? Or, are we still better at identifying familiar faces from internal features like the eyes, nose, and mouth? Even if we prefer internal features, does additional information from external features facilitate recognition? Our experiments were designed to address these open questions by assessing face recognition performance across various resolutions and by investigating the contribution of internal and external features.

Considering the importance of these issues, it is not surprising that a rich body of research has accumulated over the past few decades. Pioneering work on face recognition with low-resolution imagery was done by Harmon and Julesz (1973a, 1973b). Working with block averaged images of familiar faces of the kind shown in figure 4, they found high recognition accuracies even with images containing just 16x16 blocks. However, this high level of performance could have been due at least in part to the fact that subjects were told which of a small set of people they were going to be shown in the experiment. More recent studies too have suffered from this problem. For instance, Bachmann (1991) and Costen et al. (1996) used six high-resolution photographs during the 'training' session and low-resolution versions of the same during the test sessions. The prior subject priming about stimulus set and the use of the same base photographs across the training and test sessions renders these experiments somewhat non-representative of real-world recognition situations. Also, the studies so far have not performed some important comparisons.

Specifically, it is not known how performance differs across various image resolutions when subjects are presented full faces versus when they are shown the internal features alone.

Figure 4. *Images such as the one shown here have been used by several researchers to assess the limits of human face identification processes.(After Harmon and Julesz, 1973)*

Our experiments on face recognition were designed to build upon and correct some of the weaknesses of the work reviewed above. Here, we describe an experimental study with two goals: 1. assessing performance as a function of image resolution and 2. determining performance with internal features alone versus full faces.

The experimental paradigm we used required subjects to recognize celebrity facial images blurred by varying amounts (a sample set is shown in figure 5). We used 36 color face images and subjected each to a series of blurs. Subjects were shown the blurred sets, beginning with the highest level of blur and proceeding on to the zero blur condition. We also created two other stimulus sets. The first of these contained the individual facial features (eyes, nose and mouth), placed side by side while the second had the internal features in their original spatial configuration. Three mutually exclusive groups of subjects were tested on the three conditions. In all these experiments, subjects were not given any information about which celebrities they would be shown during the tests. Chance level performance was, therefore, close to zero.

Figure 5. *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 famous faces shown to them at the resolution depicted here. The individuals shown from left to right, are: Prince Charles, Woody Allen, Bill Clinton, Saddam Hussein, Richard Nixon and Princess Diana.*

Results

Figure 6 shows results from the different conditions. It is interesting to note that in the full-face condition, subjects can recognize more than half of the faces with image resolutions of merely 7x10 pixels. Recognition reaches almost ceiling level at a resolution of 19x27 pixels.

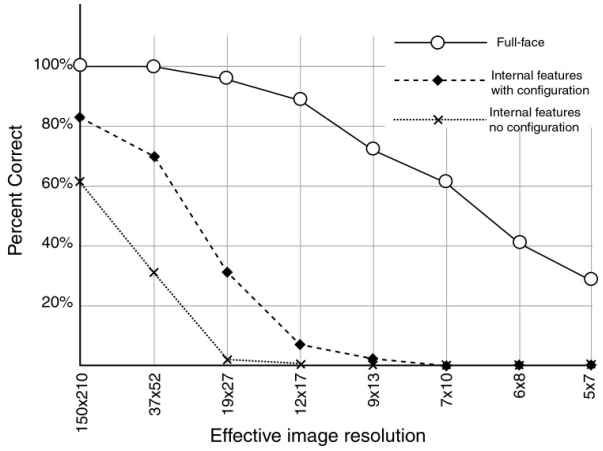

Figure 6. *Recognition performance with internal features (with and without configural cues). Performance obtained with whole head images is also included for comparison.*

Performance of subjects with the other two stimulus sets is quite poor even with relatively small amounts of blur. This clearly demonstrates the perceptual importance of the overall head configuration for face recognition. The internal features on their own and even their mutual configuration is insufficient to account for the impressive recognition performance of subjects with full face images at high blur levels. This result suggests that feature-based approaches to recognition are likely to be less robust than those based on the overall head configuration. Figure 7 shows an image that underscores the importance of overall head shape in determining identity.

Figure 7. *Although this image appears to be a fairly run-of-the-mill picture of Bill Clinton and Al Gore, a closer inspection reveals that both men have been digitally given identical inner face features and their mutual configuration. Only the external features are different. It appears,*

therefore, that the human visual system makes strong use of the overall head shape in order to determine facial identity. (From Sinha and Poggio, 1996)

In summary, these experimental results lead us to some very interesting inferences about face recognition:
1. A high-level of face recognition performance can be obtained even with resolutions as low as 12x14 pixels. The cues to identity must necessarily include those that can survive across massive image degradations. However, it is worth bearing in mind that the data we have reported here come from famous face recognition tasks. The results may be somewhat different for unfamiliar faces.
2. Details of the internal features, on their own are insufficient for subserving a high-level of recognition performance.
3. Even the mutual spatial configuration of the internal features is inadequate to explain the observed recognition performance, especially when the inputs are degraded. It appears that, unlike the belief in computer vision, where internal features are of paramount importance, the human visual system relies on the relationship between external head geometry and internal features. Additional experiments (Jarudi and Sinha, 2005) reveal a highly non-linear cue combination strategy for merging information from internal and external features. In effect, even when recognition performance with internal and external cues by themselves is statistically indistinguishable from chance, performance with the two together is very robust.

These findings are not merely interesting facts about the human visual system; rather, they help in our quest to determine the nature of facial attributes that can subserve robust recognition. In a similar vein of understanding the recognition of 'impoverished' faces, it might also be useful to analyze of the work of minimalist portrait artists, especially caricaturists, who are able to capture vivid likenesses using very few strokes. Investigating which facial cues are preserved or enhanced in such simplified depictions can yield valuable insights about the significance of different facial attributes.

We next turn to an exploration of cues for recognition in more conventional settings – high-quality images. Even here, it turns out, the human visual system yields some very informative data.

2. What cues do humans use for face recognition?

Beyond the difficulties presented by sub-optimal viewing conditions, as reviewed above, a key challenge for face recognition systems comes from the overall similarity of faces. Unlike many other classes of objects, faces share the same overall configuration of parts and the same scale. The convolutions in depth of the face surface vary little from one face to the next, and for the most part, the reflectance distributions across different faces are also very similar. The result of this similarity is that faces seemingly contain few distinctive features that allow for easy differentiation. Despite this fundamental difficulty, human face recognition is fast and accurate. This means that we must be able to make use of some reliable cues to face identity, and in this section we will consider what these cues could be.

For a cue to be useful for recognition, it must differ between faces. Though this point is obvious, it leads us to the question, what kinds of visual differences *could* there be between faces? The objects of visual perception are surfaces. At the most basic level, there are three variables that go into determining the visual appearance of a surface 1) the light that is incident on the surface 2)

the shape of the surface, and 3) the reflectance properties of the surface. This means that any cue that could be useful for the visual recognition of faces can be classified as a lighting cue, a shape cue, or a surface reflectance cue. We will consider each of these three classes of cues, evaluating their relative utility for recognition.

Illumination has a large effect on the image level appearance of a face, a fact well known to artists and machine vision researchers. Indeed, when humans are asked to match different images of the same face, performance is worse when the two images of a face to be matched are illuminated differently (Braje et al 1998; Hill and Bruce 1996), although the decline in performance is not as sharp as for machine algorithms. However, humans are highly accurate at naming familiar faces under different illumination (Moses et al 1994). This finding fits our informal experience, in which we are able to recognize people under a wide variety of lighting conditions, and do not experience the identity of person as changing when we walk with them, say, from indoor to outdoor lighting. There are certain special conditions when illumination can have a dramatic impact on assessments of identity. Lighting from below is a case in point. However, this is a rare occurrence; under natural conditions, including artificial lighting, faces are almost never lit from below. Consistent with the notion that the representation of facial identity includes this statistical regularity, faces look odd when lit from below, and several researchers have found that face recognition is impaired by bottom lighting (Enns and Shore 1997; Hill and Bruce 1996; Johnston et al 1992; McMullen et al 2000). These findings overall are consistent with the notion that the human visual system does make some use of lighting regularities for recognizing faces.

The two other cues that could potentially be useful for face recognition are the shape and reflectance properties of the face surface. We will use the term 'pigmentation' here to refer to the surface reflectance properties of the face. Shape cues include boundaries, junctions, and intersections, as well any other cue that gives information about the location of a surface in space, such as stereo disparity, shape from shading, and motion parallax. Pigmentation cues include albedo, hue, texture, translucence, specularity, and any other property of surface reflectance. 'Second order relations', or the distances between facial features such as the eyes and mouth, are a subclass of shape. However, the relative reflectance of those features or parts of features, such as the relative darkness of parts of the eye or of the entire eye region and the cheek, are a subclass of pigmentation. This division of cues into two classes is admittedly not perfect. For example, should a border defined by a luminance gradient, such as the junction of the iris and sclera be classified as a shape or pigmentation cue? Because faces share a common configuration of parts, we classify such borders as shape cues when they are common to all faces (e.g. the iris-sclera border), but as pigmentation cues when they are unique to an individual (e.g. moles and freckles). It should also be noted that this is not an image-level description. For example, a particular luminance contour could not be classed as caused by shape or pigmentation from local contrast alone. Although the classification cannot completely separate shape and pigmentation cues, it does separate the vast majority of cues. We believe that dividing cues for recognition into shape and pigmentation is a useful distinction to draw.

Indeed, this division has been used to investigate human recognition of non-face objects, although in that literature, pigmentation has typically been referred to as 'color' or 'surface'. Much of this work has compared human ability to recognize objects from line drawings or images with pigmentation cues, such as photographs. The assumption here is that line drawings contain shape cues, but not pigmentation cues, and hence the ability to recognize an object from a line drawing indicates reliance on shape cues. In particular, these studies have found recognition of line drawings to be as good (Biederman and Ju 1988; Davidoff and Ostergaard 1988; Ostergaard and Davidoff 1985) or almost as good (Humphrey 1994; Price and Humphreys 1989; Wurm 1993) as

recognition of photographs. On the basis of these and similar studies, there is a consensus that, in most cases, pigmentation is less important than shape for object recognition (Palmer 1999; Tanaka et al 2001; Ullman 1996).

In the face recognition literature too, there is a broadly held implicit belief that shape cues are more important than pigmentation cues. Driven by this assumption, line drawings and other non-pigmented stimuli are commonly used as stimuli for experimental investigations of face recognition. Similarly, many models of human face recognition use only shape cues, such as distances, as the featural inputs. This only makes sense if it is assumed that the pigmentation cues being omitted are unimportant. Also, there are many more experimental studies investigating specific aspects of shape than of pigmentation, suggesting that the research community is less aware of pigmentation as a relevant component of face representations. However, this assumption turns out to be false. In the rest of this section, we will review evidence that both shape and pigmentation cues are important for face recognition.

First, we briefly review the evidence that shape cues alone can support recognition. Specifically, we can recognize representations of faces that have no variation in pigmentation, hence no useful pigmentation cues. Many statues have no useful pigmentation cues to identity because they are composed of a single material, such as marble or bronze, yet are recognizable as representations of a specific individual. Systematic studies with 3D laser-scanned faces that similarly lack variation in pigmentation have found that recognition can proceed with shape cues only (Bruce et al 1991; Hill and Bruce 1996; Liu et al 1999). Clearly, shape cues are important, and sometimes sufficient, for face recognition.

However, the ability to recognize a face in the absence of pigmentation cues does not mean that such cues are not used under normal conditions. To consider a rough analogy, the fact that we can recognize a face from a view of only the left side does not imply that the right side is not also relevant to recognition. There is reason to believe that pigmentation may also be important for face recognition. Unlike other objects, faces are much more difficult to recognize from a line drawing than from a photograph (Bruce et al 1992; Davies et al 1978; Leder 1999; Rhodes et al 1987), suggesting that the pigmentation cues thrown away by such a representation may well be important.

Recently, some researchers have attempted to directly compare the relative importance of shape and pigmentation cues for face recognition. The experimental challenge here is to find a way to create a stimulus face that appears naturalistic, yet does not contain either useful shape or pigmentation cues. The basic approach that was first taken by O'Toole and colleagues (O'Toole et al 1999) is to create a set of face representations with the shape of a particular face and the average pigmentation of a large group of faces, and a separate set of face representations with the pigmentation of an actual face and the average shape of many faces. The rationale here is that to distinguish among the faces from the set that all have the same pigmentation, subjects must use shape cues, and vice versa for the set of faces with the same shape. In O'Toole's study, the faces were created by averaging separately the shape and reflectance data from laser scans. The shape data from individual faces was combined with the averaged pigmentation to create the set of faces that differ only in terms of their shape, and vice versa for the set of faces with the same shape. Subjects were trained with one or the other set of face images, and were subsequently tested for memory. Recognition performance was about equal with both the shape and pigmentation sets, suggesting that both cues are important for recognition.

A question that arises when comparing the utility of different cues is whether the relative performance is due to differences in image similarity or differences in processing. One way to

address this is to equate the similarity of a set of images that differ by only one or the other cue. If the sets of images are equated for similarity, differences in recognition performance with the sets can be attributed to differences in processing. We investigated this question in our lab (Russell et al 2004) with sets of face images that differed only by shape or pigmentation cues and were equated for similarity using the 'gabor jet' model of facial similarity developed by von der Malsburg and colleagues (Lades et al 1993). We used photographic images that were manipulated with image morphing software. Before considering the results, we can see from the stimuli (some of which are shown in Figure 8) that both shape and pigmentation can be used to establish identity. With the similarity of the cues equated in grayscale images, we found slightly better performance with the shape cues. With the same images in full color (but not equated for similarity), we found slightly better performance with the pigmentation cues. Overall, the performance was approximately equal using shape or pigmentation cues. This provides evidence that both shape and pigmentation cues are important for recognition. The implication for computer vision systems is obvious – facial representations ought to encode both of these kinds of information to optimize performance.

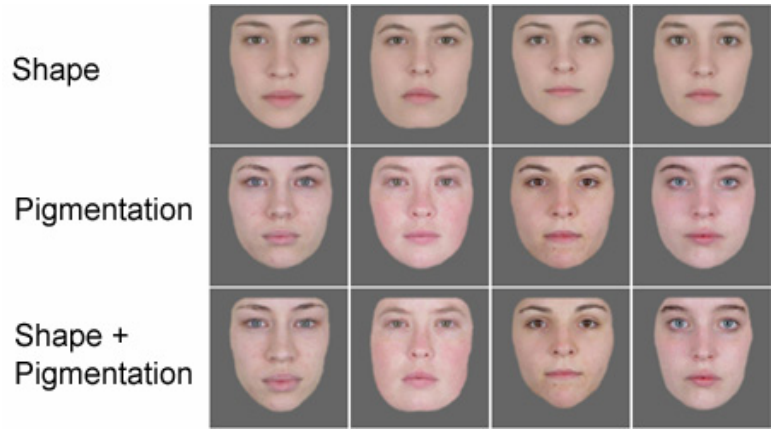

Figure 8. *Some of the faces from our comparison of shape and pigmentation cues. Faces along the top row differ from one another in terms of shape cues, faces along the middle row differ in terms of pigmentation cues, and faces along the bottom row differ in terms of shape and pigmentation cues (they are actual faces). The faces along the top row all have the same pigmentation, but they do not appear to be the same person. Similarly, the faces in the middle row do not look like the same person, despite all having the same shape. This suggests that both shape and pigmentation are used for face recognition by the HVS.*

One particular subcategory of pigmentation cues, color, has received a little extra attention, with researchers investigating whether color cues are used for face recognition. An early study reversed separately the hue and luminance channels of face images, and found that, while luminance reversal (contrast negation) caused a large decline in recognition performance, reversing the hue channel did not (Kemp et al 1996). Thus, faces with incorrect colors can be recognized as well as those with correct colors, a point that has also been noted with respect to the use of 'incorrect' colors by the early 20th century 'Fauvist' school of art (Zeki 2000). Another approach has been to investigate whether exaggerating color cues by increasing the differences in hue between an individual face image and an average face image improves recognition performance. The results suggest that this manipulation does improve performance, but only with fast presentation (Lee and Perrett 1997), and with a small bias toward color caricatures as better likenesses of the individuals than the veridical images (Lee and Perrett 1997). In our laboratory we have taken a more direct approach, comparing performance with full color and grayscale

images. In one study investigating familiar face recognition (Yip and Sinha 2002) and the study mentioned above investigating unfamiliar face matching (Russell et al 2004), we have found significantly better performance when subjects are viewing full color rather than grayscale images. Overall, color cues do seem to contribute to recognition.

To summarize the findings on the question of what cues are used in face recognition, there is evidence that all categories of cues that could potentially be used to recognize faces—illumination, shape, and pigmentation—do contribute significantly to face recognition. The human face recognition system appears to use all available cues to perform its difficult task. The system does this by making use of many weak but redundant cues. In the context of faces, the human visual system appears to be opportunistic, making use of all cues that vary across exemplars of this class to achieve its impressive face recognition skills.

An important question that arises at this juncture is whether humans are innately equipped with these skills and strategies, or have to learn them through experience. This is the issue we consider next.

3. What is the timeline of development of human face recognition skills?

Considering the complexity of the tasks, face spotting and recognition skills develop surprisingly rapidly in the course of an infant's development. As early as a few days after birth, infants appear to be able to distinguish their mother's face from other faces. Yet, in light of the rapid progression of face recognition abilities, it may be surprising to learn that child face processing abilities are in some ways still not adult-like, even up until the age of 10 *years*. This section explores the trajectory of development of face processing.

Bootstrapping Face Processing: Evidence from Neonates
As is the case for most visual skills, face processing must be bootstrapped with some primitive mechanism from which more advanced processes can be learned. A key question is whether infants are born with some innate abilities to process faces or are those abilities a consequence of general visual learning mechanisms?

To examine this issue, neonates (newborns) are assessed for various abilities as soon as is practical after birth. Three major findings have historically been taken as evidence for innate facial processing abilities: (1) The initial preference for faces over non-faces (2) The ability to distinguish the mother from strangers, and (3) imitation of facial gestures. We will look at each of these in turn.

Innate facial preference
Are infants pre-wired with a face detection algorithm? If infants knew how to locate the faces in an image, it would be a valuable first step in the subsequent learning of face recognition processes. Morton and Johnson (1991) formalized this idea into a theory called *CONSPEC and CONLERN*. CONSPEC is the structural information which guides newborn preferences for face-like patterns. CONLERN is a learning device which extracts further visual characteristics from patterns identified based on CONSPEC. CONLERN does not influence infant looking behavior until 2 months of age.

Some supporting evidence for this theory comes from the fact that newborns do indeed preferentially orient their gaze to face-like patterns. Morton and Johnson (1991) (also see Simion et al, 1996; and Goren et al 1975, Johnson et al, 1991; Valenza et al, 1996) presented face-like stimuli as in Fig. 9 (left) and the same display with internal features presented upside-down (right). The newborns gazed longer at the face-like pattern. Since the two patterns are largely identical except for their "faceness" and the subjects have had virtually no other visual experience prior to the presentation, the experimenters concluded that newborns have an innate preference for faces.

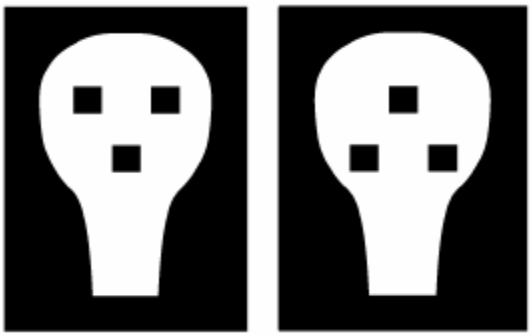

Figure 9. *Newborns preferentially orient their gaze to the face-like pattern on the left, rather than the one shown on the right, suggesting some innately specified representation for faces. (From Johnson and Morton, 1991).*

Recent studies, however, have called into question the structural explanation based on 'faceness' and focused more on lower-level perceptual explanations. Simion (2001) showed infants displays as in figure 10. These displays bore no resemblance to faces, yet newborns preferred the displays on the left. Simion and colleagues concluded that newborn preferences in this case were not for face-like stimuli per se, but rather for top-heavy displays. Similar experiments with real faces and real faces with scrambled internal features (Cassia 2004) had the same pattern of results. By three months, however, genuine face preferences appear to emerge. Turati (2005) performed similar experiments with 3-month old infants and found that, by this age, infants do indeed seem to orient specifically towards face-like stimuli and that this orientation cannot be so easily explained by the low-level attributes driving newborns' preferences.

The above pattern of experimental results shows that, although newborns seem to have some set of innate preferences, these mechanisms are not necessarily domain-specific to faces. However, even though the particular orienting heuristic identified here can match non-face stimuli, it may still be the case that the presence of this heuristic biases the orientation of attention towards faces often enough over other stimuli to target a face-specific learning process on stimuli that are usually faces. A more accurate computational characterization of these perceptual mechanisms for the orientation of attention could be applied to sample images of the world (adjusted to mimic the filter of an infant's relatively poor visual system) in order to determine whether these mechanisms do indeed orient the infant towards faces more often than to other stimuli. For now, it cannot be said whether in fact newborns prefer faces to other stimuli in their environment.

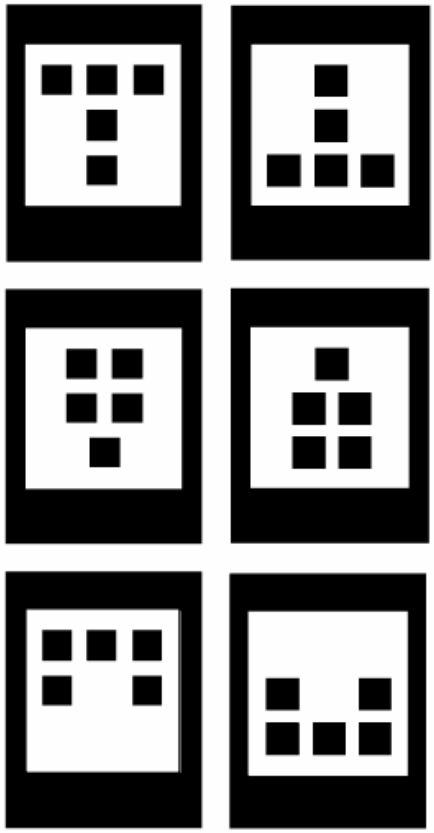

Figure 10. *As a counterpoint to the idea of innate preferences for faces, Simion (2001) has shown that newborns consistently prefer top-heavy patterns (left column) over bottom-heavy ones (right column). This may well account for their tendency to gaze longer at facial displays, without requiring an innately specified 'face detector'.*

Distinguishing the mother from strangers

Plausible evidence for an innate mechanism for facial processing comes from the remarkable ability of newborn infants within less than 50 hours of birth (Field 1984, Bushnell 1989, Walton 1992) to discriminate their mothers from strangers. Newborns suck more vigorously when viewing their mother's face on a videotaped image. They also are capable of habituating to a mother's image, eventually preferring a novel image of a stranger, showing a classic novelty preference. Even with Pascalis' (1995) qualification that discrimination can only be achieved when external features (mainly hair) are present, it seems that such performance can only be achieved if newborns already possess at least a rudimentary facial processing mechanism.

A counterargument to the conclusion that an innate mechanism is necessary is given by Valentin and Abdi (2001). These researchers point out that a newborn is only asked to discriminate among a few different faces, all of which are generally very different. Given a small number of face images degraded to mimic the acuity of a newborn, Valentin and Abdi attempted to train an artificial neural network to distinguish among them using an unsupervised learning algorithm. For small numbers of images comparable to the experiments performed on newborns, the network is successful. Since this network made no initial assumptions about the content of the images (faces or otherwise), an infant's visual system also need not necessarily have a face-specific mechanism to accomplish this task.

Imitation of facial expressions
An oft-cited "fact" is that newborns are able to imitate the facial expressions made by others. This would entail the infant's recognizing the expression, then mapping it to its own motor functions. Many studies have shown evidence of imitation (most notably, Meltzoff and Moore 1977 and 1983). A comprehensive review by Anisfeld (1991, 1996), however, revealed that there were more studies with negative than positive results. Moreover, there were consistent positive results only for one expression, namely, tongue protrusion. This action might be an innate releasing mechanism, perhaps to a surprising stimulus, or an attempt at oral exploration (Jones 1996). Thus, the action is a response to the stimulus, not a recognition of it followed by imitation. The coincidence only *seems* like imitation to the observer.

On the whole, the findings so far are equivocal regarding the innateness of face processing mechanisms. This is perhaps good news for computer vision. The machine-based systems do not necessarily have to rely on any 'magical' hardwired mechanisms, but can learn face processing skills through experience.

Behavioral Progression

Although infant face recognition proficiency develops rapidly within the first few months of life, performance continues to improve up to the age of 10 years or even later. The most well-studied progression in the behavioral domain is the use of featural versus configural cues.

Adults match upright faces more easily than inverted faces. This is believed to be a consequence of the disruption in configural processing in inverted faces. Interestingly, four month old infants do not show this decrement in performance (the so-called inversion effect) if the images to match are otherwise identical. However, the inversion effect does appear if, for instance, pose is manipulated at each presentation of the face (Schwarzer, 2000; Turati 2004). Thus, there is evidence that configural cues start to play some role in face processing at this early age, although the processing of these cues is still rudimentary. Processing based on features appears to play the primary role in infant facial recognition (Cohen and Cashon, 2001). Given the early onset of the usage of configural cues in child face recognition, rudimentary though it may be, one would expect that full maturation of such a fundamental system would ensue rapidly. However, numerous studies have found that, although face recognition based on features reaches near-adult levels by the age of 6 years, configural processing still lags behind until at least 10 years of age, with a gradual progression of greater accuracy and dependence on configural cues (Carey and Diamond 1977; Pellicano and Rhodes 2003; Maurer 2002; Mondloch 2002 and 2003; Hay and Cox 2000; Bruce 2000).

Is this developmental improvement in the use of configural cues an outcome of general learning processes or a maturational change? Some evidence for the latter argument comes from patients with congenital bilateral cataracts in infancy (Le Grand 2001, Geldart 2002). Even after more than 9 years of recovery, deprivation of patterned visual input within the first 2-6 months of life leads to a substantial deficit in configural processing (see figure 11). Children with a history of early visual deprivation are impaired in their ability to distinguish between faces differing in the positions of features. However, they show no such deficits when distinguishing between faces differing in their constituent features. Perhaps this early maturation is a critical period in the development of face recognition processes.

Figure 11. *Children with a history of early visual deprivation are impaired in their ability to distinguish between faces differing in the positions of features (top row). However, they show no such deficits when distinguishing between faces differing in their constituent features (bottom row). (After Le Grand 2001, Nature)*

Neuroimaging

Finally, we look at the recent contributions of neuroimaging to the understanding of changes in brain activation patterns to link the development of brain processes to changes in behavioral performance.

One neural marker used to study face-specific processes is the N170, which is an event-related potential (ERP) recorded using EEG. In adults, this signal generally occurs over the occipito-temporal lobe between 140 and 170 msec after the visual onset of a face image. De Haan et al. (2002) looked for the N170 in 6-month-old infants. They compared human to non-human primate faces, and upright to inverted faces. They report an "infant N170" (with a slower peak latency, a distribution of location more towards the medial areas of the brain, and a somewhat smaller amplitude as compared to the normal adult N170) which prefers human faces to non-humans, but which shows no sensitivity to inversion. This is evidence of the development of a possibly face-specific process, but this particular process, at least, seems to be insensitive to configural manipulation at this early age.

Another important neural marker in face research is activation in the fusiform gyrus during functional MRI (fMRI) scans while viewing faces. Aylward et al. (2005) searched for this signal in young children (8-10 years) and slightly older children (12-14 years). In a paradigm comparing the passive viewing of faces versus houses, younger children showed no significant activation in the fusiform gyrus, while older children did. Thus, although the neuroimaging data are very preliminary at the present, they do appear to be broadly consistent with the behavioral data showing continuing development of face processing well into adolescence.

Summary of developmental studies of face recognition
Although infant face recognition develops rapidly to the point where simple matching and recognition can take place, adult-like proficiency takes a long time to develop, at least until the age of 10. Early face processing seems to rely significantly on feature discrimination, with configural processing taking years to mature. Brain activation patterns echo this long maturational process. Interestingly, however, visual deprivation for even the first two months of life can cause possibly permanent deficits in configural face processing. The interaction of featural and configural processes may be one of the keys to fully understanding facial recognition.

4. What are some biologically plausible strategies for face recognition?

When developing computational methods for face recognition, it is easy to treat the recognition problem as a purely theoretical exercise, having nothing to do with faces per se. One may treat the input images as completely abstract data and use various mathematical techniques to arrive at statistically optimal decision functions. Such treatment will likely lead to the creation of a useful system for recognition, at least within the confines of the training images supplied to the algorithm.

At the same time, the machine vision researcher is very lucky to have access to a wealth of knowledge concerning the function of the human visual system. This body of knowledge represents an incomplete blueprint of the most impressive recognition machine known. By making use of some of the same computations carried out in the visual pathway it may be possible to create extremely robust recognition systems. Also, mimicking the computations known to be carried out in the visual pathway may lead to an understanding of what is going on in higher-level cortical areas. Building a model recognition system based on our admittedly limited knowledge of the visual pathway may help us fill in the gaps, so to speak. With this goal in mind, many laboratories have developed computational models of face recognition that emphasize biologically plausible representations. We shall discuss several of those models here to demonstrate how physiological findings have informed computational studies of recognition. Also, we shall discuss possible ways in which recent computational findings might inform physiology.

Early vision and face recognition

One of the most important findings in visual neuroscience was Hubel & Wiesel's discovery of orientation-specific cells in the feline primary visual cortex (Hubel & Wiesel, 1959). These cells were found to be tuned to stimuli along many different dimensions, including orientation, spatial frequency, motion, and color. The discovery of these cells led to the formulation of a hierarchical model of visual processing in which edges and lines were used to construct more complicated forms. An influential framework modeling how the visual pathway might combine low-level features in a cascade that leads to high-level recognition was put forth by Marr in his book *Vision* (Marr, 1982). This framework has inspired many recent models of face recognition, as reviewed below.

Face Recognition: The Malsburg Model - A strategy adopted by many researchers has been to begin constructing computational models of recognition that utilize these same low-level

visual primitives. In this way, the model is made to operate on data that roughly matches how an image appears to the early visual cortex. The receptive fields of cells in early visual cortex are most commonly represented as *Gabor functions* in these models. A Gabor function is simply a sinusoid that has been windowed with a Gaussian. The frequency, orientation, and size of the function can be easily manipulated to produce a range of different model receptive fields.

Given these functions, there are various ways in which one can extract measurements from an image to perform recognition. One of the most useful strategies employed to date is the *Gabor jet* (also mentioned in section 2, above). A jet is simply a "stack" of Gabor functions, each one with a unique orientation, frequency tuning, and scale (Figure 12). The construction of a jet is meant to mimic the multi-scale nature of receptive field sizes as one moves upstream from V1. It also provides a wealth of information concerning each point to which it is applied, rather than just relaying one response. Jets have been used as the basis of a very successful recognition technique known as "elastic bunch graph matching." (Lades et al., 1993; Wiskott, Fellous, Kruger, & von der Malsburg, 1997)

Figure 12. *A Gabor jet. Linear filters are placed at a variety of scales and orientations.*

In this model, Gabor jets are applied to special landmarks on the face, referred to as *fiducial points*. The landmarks used are intuitively important structural features such as the corners of the eyes and mouth and the outline of the face. At each of these points, the Gabor jet produces a list of numbers representing the amount of contrast energy that is present at the spatial frequencies, orientations, and scales included in the jet. These lists of numbers from each point are combined with the locations of each landmark to form the final representation of the face. Each face can then be compared with any other with a similarity metric that takes into account both the appearance and the spatial configuration of the landmarks.

The model has been shown to perform well on the FERET database benchmark for face recognition (Phillips, Moon, Rizvi, & Rauss, 2000). Moreover, it has also been shown that the similarity values computed by the algorithm correlate strongly with human similarity judgments (Biederman & Kalocsai, 1997). This last result is particularly satisfying in that it suggests that these simple measurements capture perceptually important high-level information. The use of features that mimic visual system receptive fields appears to be a very useful strategy for representing faces when their outputs are combined in the right way.

Face Detection – A companion problem to face recognition (or individuation) is the challenge of face detection. In some ways, detecting a face in a cluttered scene presents deeper difficulties than determining the identity of a particular face. The main reason for this state of affairs is that when one is detecting a face in a complicated scene, one must determine a set of visual features that will always appear when a face is present and rarely appear otherwise. An additional layer of difficulty is added by the fact that one will likely have to scan a very large image to look for faces. This means that we will require fast computations at all stages of detection, as we may need to perform those measurements hundreds of times across a single image.

In this realm, as in face individuation, employing features that mimic those found in primary visual cortex has been found to be useful. We shall discuss two models that rely upon simple spatial features for face detection. The first makes use of box-like features that provide for exceptional speed and accuracy in large images containing faces. The second provides for impressive invariance to different lighting conditions by incorporating ordinal encoding, a non-linear computation that more closely resembles the response properties of V1 cells.

Viola & Jones – Our first example of a face detection model based on early visual processing was developed by Viola & Jones (Viola & Jones, 2001). As in the Malsburg model, the selection of primitive features was motivated by the finding that Gabor-like receptive fields are found in primary visual cortex. However, rather than using true Gabor filters to process the image, this model utilizes an interesting abstraction of these functions to buy more speed for their algorithm. Box-like features are used as a proxy for Gabors because they are much faster to compute across an image and they can roughly approximate many of the spatial characteristics of the original functions. By constructing an extremely large set of these box features, the experimenters were able to determine which computations were the most useful at discriminating faces from background. It should be noted that no individual feature was particularly useful, but the combined judgments of a large family of features provides for very high accuracy. The notion that many 'weak' classifiers can be 'smart' when put together is an instance of 'boosting.' Examples of the best features for face detection are shown in Figure 13.

Figure 13. *Examples of the box filters evaluated by Viola & Jones for face detection showing a selection of the most useful features for face detection. Note the resemblance to the receptive fields of V1 neurons. (after Viola & Jones, 2001)*

We see in Viola and Jones' model that very simple image measurements are capable of performing both detection and recognition tasks. This indicates that the early visual system may be capable of contributing a great deal to very complex visual tasks. Though this model is very good at detecting faces in cluttered backgrounds, it does break down when large variations in facial pose, expression or illumination are introduced. Human observers are capable of recognizing and detecting faces despite large variations of these kinds, of course. The challenge then is to produce a system equally capable of detecting faces despite these sources of variation.

We present in the next section a model from our laboratory which aims to uncover what processing strategies might support invariance to varying face illumination in particular. Our model overcomes the problem of varying illumination through a very simple modification of the basic edge-based representations mentioned previously. This modification, ordinal encoding, is particularly compelling in that it more closely models the behavior of V1 cells. This further bolsters the idea that biological systems may implicitly make use of computational principles that can inform our modeling.

Ordinal Encoding & Ratio-templates – The computations carried out by this model are very similar to those already discussed. We begin by noting a troubling difficulty associated with these simple measurements, however. As the lighting of a particular face changes, we note that the values produced by a simple box-like filter comparing two neighboring image regions may change a great deal. (Figure 14) This is deeply problematic if we wish to be able to detect this face consistently. We must either hope that we can find other measurements that will not behave this way, or find a new way of using these features to represent a face.

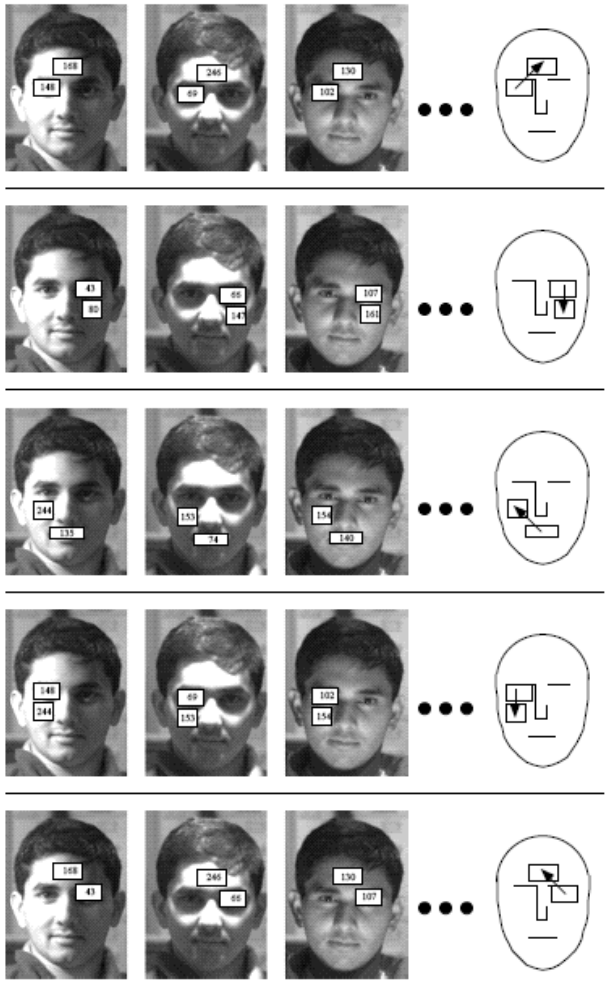

Figure 14. *An example of how changes in ambient lighting can greatly alter the brightness values in neighboring regions of a face. We point out that although the absolute difference changes, the direction of the brightness gradient between some regions stays constant. (from Sinha, 2002)*

We shall adopt the latter strategy, motivated by the observation that though the numerical values of the two regions may change, the sign of their relative magnitudes stays constant. This is precisely the kind of stable response we can use to perform face detection. Rather than maintaining precise image measurements when comparing two regions, we will only retain information about which region was the brightest. We are throwing away much quantitative information, but in a way that benefits the system's performance. We also note that this is much more representative of the behavior of a typical neuron in the visual system. Such cells are generally not capable of representing a wide range of contrasts, but rather saturate rapidly (Figure 15), providing an essentially binary signal (either "different" or "not different).

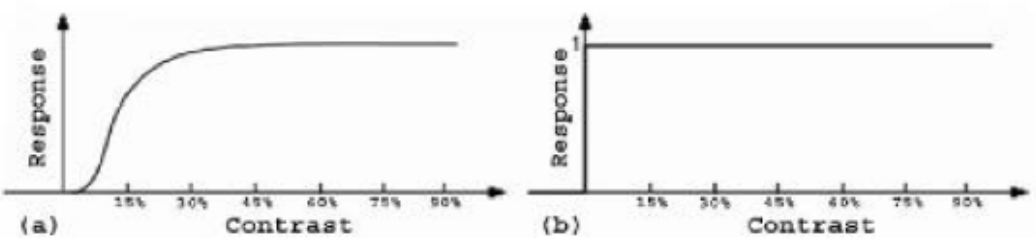

Figure 15. *A demonstration of how the response of a typical neuron in the primary visual cortex asymptotes rapidly with increasing contrast. An idealized version of this neuron which is a pure step function is presented at right. (from Sadr et al. 2002)*

This kind of measurement constitutes an ordinal encoding, and we can build up a representation of what we expect faces to look like under this coding scheme for use in detection tasks. We call this representation a "ratio-template" because it makes explicit only coarse luminance ratios between different face regions (Sinha, 2002). The model performs very well in detection tasks (figure 16 shows some results), and we have also shown that relatively high-fidelity reconstructions of images are possible from only local ordinal measurements (Sadr, Mukherjee, Thoresz, & Sinha, 2002). This means that despite the coarse measurements we encode via ordinal measurements, the original image can be recovered very accurately.

Figure 16. *Results of using a qualitative representation scheme to encode and detect faces in images. Each box corresponds to an image fragment that the system believes is a face. The representation scheme is able to tolerate wide appearance and resolution variations. The lower detection in the central image is a false positive.*

We have seen in these three models, that using computations similar to those carried out early in the visual system can provide good performance in complex visual tasks. At the heart of all three of these models is the Gabor function, which captures much of the tuning properties of neurons in the striate cortex. In the Malsburg model, these functions are applied in *jets* over the surface of a face to provide information about spatial features and configuration for individuation. In Viola & Jones' detection algorithm, box-filters that approximate these functions are rapidly computed across a complex scene and combine their outputs to determine if a face is present. Finally, by incorporating the non-linear nature of V1 cells into an ordinal code of image structure, our own algorithm provides for useful invariance to lighting changes. Taken together, these three models suggest that turning to biologically plausible mechanisms for recognition can help enhance the performance of computational systems.

Looking downstream: Can computational models inspire physiological research?
Up to this point, we have discussed how our knowledge of early visual cortex has driven the creation of various models for face recognition and detection. In this section, we shall speculate on how computational studies of face recognition might help drive physiologists to look for particular structures in higher-level visual areas.

It has been known for some time that the stimuli visual neurons are tuned to, increase in complexity as we move into higher stages of the visual pathway. While edge-like features are found in V1, cells that respond selectively to hands and faces have been found in primate infero-temporal cortex (Gross, Rocha-Miranda, & Bender, 1972). Likewise, functional MRI studies of the human visual pathway reveal a possible "face area" in the fusiform gyrus (Kanwisher,

McDermott, & Chun, 1997). Given the impressive selectivity in these areas, the obvious challenge is to determine what features or processes produce such specificity for complex objects. This is a very difficult endeavor, and many research groups have attempted to describe the tuning properties of various populations of high-level neurons. The chief difficulty is that the space of possible image features to choose from is very large, and one must limit oneself to some corner of that space in order to make progress. In so doing, however, it is impossible to know if there is another image feature outside of that space that would be truly "optimal" for a given neuron. The current picture of what stimuli higher-level neurons are tuned to is very complex, with reports of tuning for various forms of checkerboard gratings (Gallant, Connor, Rakshit, Lewis, & Van Essen, 1996) and curvature (Pasupathy & Connor, 2002) in V4 through parametric studies. Non-parametric stimulus-reduction techniques have also led to a possible characterization of IT into complex object 'columns' (Tanaka, 2003).

We suggest that determining what complex features are useful for high-level tasks such as face detection and recognition may help physiologists determine what kind of selectivities to look for in downstream visual areas. In this way, computational models can contribute to physiology in much the same way as physiology has enhanced our computational understanding of these processes. We present here one very recent result from our lab that may suggest an interesting modification of our current models of cortical visual analysis based on computational results.

Dissociated Dipoles: Image representation via non-local operators

We have noted thus far that many computational models of face recognition rely on edge-like analyses to create representations of images. However, one of the primary computational principles that appears to underlie the emergence of such cells in the visual cortex is that of image reconstruction. That is, edge-like cells appear to be an optimal solution to the problem of perfectly reconstructing image information with a small set of features (Bell & Sejnowski, 1997; Olshausen, 1996; Olshausen & Field, 1997). However, it is not clear that one needs to perfectly reconstruct an image in order to recognize it. This led us to ask if such features are truly the most useful for a task like face recognition. We set out to determine a useful vocabulary of image features for face recognition through an exhaustive consideration of two-lobed box filters. In our analysis, we allowed lobes to be overlapping, neighboring, or completely separate. Each feature was then evaluated according to its ability to accurately identify faces from a large database (Balas & Sinha, submitted).

Under the modified criterion of recognition, rather than reconstruction, two distinct families of model neurons appeared as good features. These "optimal" features were cells with a center-surround organization and those with two spatially disjoint lobes (Figure 17). The former are akin to retinal ganglion cells, but the latter do not resemble anything reported to date in the primate or human visual pathway. We have since found that despite the computational oddities they present, these non-local operators (which we call "dissociated dipoles") are indeed very useful tools for performing recognition tasks with various face databases. (Balas & Sinha, 2003)

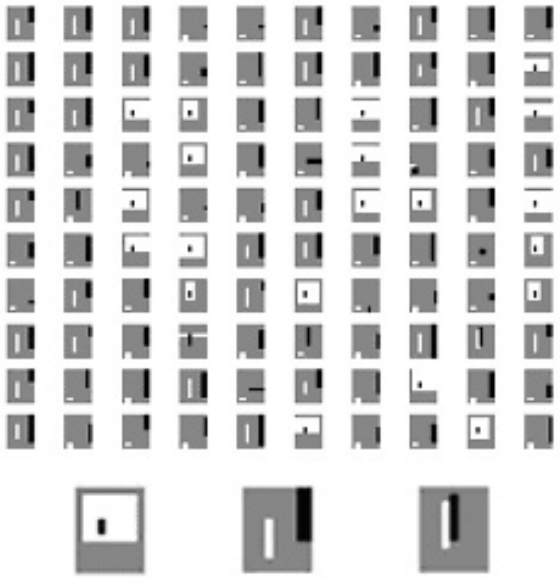

Figure 17. *Examples of the best differential operators for face recognition discovered by Balas & Sinha. Note the prevalence of center-surround and non-local receptive fields. We call the non-local operators "Dissociated Dipoles" in reference to their spatially disjoint receptive fields.*

Could such computations be carried out in the visual processing stream? They appear to be found in other sensory modalities, such as audition and somatosensation (Chapin, 1986; Young, 1984) meaning that they are certainly not beyond the capabilities of our biological machinery. We suggest that it might be useful to look for such operators in higher-level visual areas. One way in which complexity might be built up from edges to objects may be through incorporating information from widely separated image regions. While there is still no evidence of these kinds of receptive fields to date, this may be one example of how a computational result can motivate physiological inquiry.

Conclusion

Face recognition is one of the most active and exciting areas in neuroscience, psychology and computer vision. While significant progress has been made on the issue of low-level image representation, the fundamental question of how to encode overall facial structure remains largely open. Machine based systems stand to benefit from well-designed perceptual studies that can allow precise inferences to be drawn about the encoding schemes used by the human visual system. We have reviewed here a few results from the domain of human perception that provide benchmarks and guidelines for our efforts to create robust machine based face recognition systems. It is important to stress that the limits of human performance do not necessarily define upper bounds on what is achievable. Specialized identification systems (say those based on novel sensors, such as close-range IR cameras) may well exceed human performance in particular settings. However, in many real-world scenarios using conventional sensors, matching human performance remains an elusive goal. Data from human experiments can not only give us a better sense of what this goal is, but also what computational strategies we could employ to move towards it and, eventually, past it.

References

Anisfeld, M. (1991). Neonatal imitation. *Developmental Review,* **11**, 60-97.

Anisfeld, M. (1996). Only tongue protrusion modeling is matched by neonates. *Developmental Review,* **16**, 149-161.

Aylward, E. H., Park, J. E., Field, K. M., Parsons, A. C., Richards, T. L., Cramer, S. C., et al. (2005). Brain activation during face perception: Evidence of a developmental change. *Journal of Cognitive Neuroscience,* **17**(2), 308-319.

Bachmann, T. (1991). Identification of spatially quantized tachistoscopic images of faces: How many pixels does it take to carry identity? *European Journal of Cognitive Psychology*, **3**, 85-103.

Balas, B. J., & Sinha, P. (2003). *Dissociated Dipoles:Image representation via non-local operators* (AIM-2003-018). Cambridge, MA: MIT AI Lab.

Balas, B. & Sinha, P. (2005) Receptive Field Structures for Recognition. MIT CSAIL Memo, AIM-2005-006, CBCL-246.

Bell, A. J., & Sejnowski, T. J. (1997). The "Independent Components" of Natural Scenes are Edge Filters. *Vision Research,* **37**(23), 3327-3338.

Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society of London , Series B.,* **352**(1358), 1203-1219.

Biederman I, Ju G, 1988 “Surface versus Edge-Based Determinants of Visual Recognition” *Cognitive Psychology* **20** 38-64

Braje W L, Kersten D, Tarr M J, Troje N F, 1998 “Illumination Effects in Face Recognition” *Psychobiology* **26** 371-380

Bruce V, Healey P, Burton M, Doyle T, Coombes A, Linney A, 1991 “Recognising facial surfaces” *Perception* **20** 755-769

Bruce V, Hanna E, Dench N, Healey P, Burton M, 1992 “The importance of 'mass' in line drawings of faces” *Applied Cognitive Psychology* **6** 619-628

Bruce, V., & Young, A. (1998). In the Eye of the Beholder: The Science of Face Perception. Oxford: Oxford University Press.

Bruce, V., Campbell, R. N., Doherty-Sneddon, G., Import, A., Langton, S., McAuley, S., et al. (2000). Testing face processing skills in children. *British Journal of Developmental Psychology,* **18**, 319-333.

Bushnell, I. W. R., Sai, F., & Mullin, J. T. (1989). Neonatal Recognition of the Mothers Face. *British Journal of Developmental Psychology,* **7**, 3-15.

Carey, S., & Diamond, R. (1977). From Piecemeal to Configurational Representation of Faces. *Science,* **195**(4275), 312-314.

Cassia, V. M., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science,* **15**(6), 379-383.

Chapin, J. K. (1986). Laminar Differences in Sizes, Shapes, and Response Profiles of Cutaneous Receptive Fields in Rat SI Cortex. *Experimental Brain Research,* **62,** 549-559.

Cohen, L. B., & Cashon, C. H. (2001). Do 7-month-old infants process independent features or facial configurations? *Infant and Child Development,* **10**(1-2), 83-92.

Collishaw, S. M., & Hole, G. J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception*, **29**(8), 893-909.

Costen, N. P., Parker, D. M., & Craw, I. (1994). Spatial content and spatial quantization effects in face recognition. *Perception*, **23**, 129-146.

Davidoff J, Ostergaard A, 1988 "The role of color in categorical judgments" *Quarterly Journal of Experimental Psychology* **40** 533-544

Davies G M, Ellis H D, Sheperd J W, 1978 "Face recognition accuracy as a function of mode of representation" *Journal of Applied Psychology* **63** 180-187

DeAngelis, G., Ohzawa, I. and Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology*, **69**(4): 1091-1117.

de Haan, M., Pascalis, O., & Johnson, M. H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience,* **14**(2), 199-209.

Duda, R., and Hart, P. (1973). Pattern classification and scene analysis. Wiley: NY.

Enns J T, Shore D I, 1997 "Separate influences of orientation and lighting in the inverted-face effect" *Perception & Psychophysics* **59** 23-31

Field, T. M., Cohen, D., Garcia, R., & Greenberg, R. (1984). Mother-Stranger Face Discrimination by the Newborn. *Infant Behavior & Development, 7*(1), 19-25.

Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *Journal of Neuroscience, 76*(4), 2718-2739.

Geldart, S., Mondloch, C. J., Maurer, D., de Schonen, S., & Brent, H. P. (2002). The effect of early visual deprivation on the development of face processing. *Developmental Science, 5*(4), 490-501.

Goren, C. C., Sarty, M., & Wu, P. Y. K. (1975). Visual Following and Pattern-Discrimination of Face-Like Stimuli by Newborn-Infants. *Pediatrics, 56*(4), 544-549.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology, 35*(1), 96-111.

Harmon, L. D. & Julesz, B. (1973a). Masking in visual recognition: Effects of two-dimensional noise. *Science*, 180, 1194-1197.

Harmon, L. D. (1973b). The recognition of faces. *Scientific American*, **229**(5), 70-83.

Hay, D. C., & Cox, R. (2000). Developmental changes in the recognition of faces and facial features. *Infant and Child Development, 9*, 199-212.

Hill H, Bruce V, 1996 "Effects of lighting on the perception of facial surfaces" *Journal of Experimental Psychology: Human Perception and Performance* **22** 986-1004

Hubel, D., & Wiesel, T. (1959). Receptive Fields of Single Neurons in the Cat's Striate Cortex. *Journal of Physiology, 148*, 574-591.

Humphrey G, 1994 "The role of surface information in object recognition: studies of visual form agnosic and normal subjects" *Perception* **23** 1457-1481

Jarudi, I. and Sinha, P. (2005). Contribution of internal and external features to face recognition. (Submitted)

Johnston A, Hill H, Carman N, 1992 "Recognising faces: effects of lighting direction, inversion, and brightness reversal" *Perception* **21** 365-375

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns Preferential Tracking of Face-Like Stimuli and Its Subsequent Decline. *Cognition, 40*(1-2), 1-19.

Jones, S. S. (1996). Imitation or exploration? Young infants' matching of adults' oral gestures. *Child Dev, 67*(5), 1952-1969.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302-4311.

Kemp R, Pike G, White P, Musselman A, 1996 "Perception and recognition of normal and negative faces: the role of shape from shading and pigmentation cues" *Perception* **25** 37-52

Lades M, Vortbruggen J C, Buhmann J, Lange J, von der Malsburg C, Wurtz R P, Konen W, 1993 "Distortion invariant object recognition in the dynamic link architecture" *IEEE Transactions on Computers* **42** 300-311

Leder H, 1999 "Matching person identity from facial line drawings" *Perception* **28** 1171-1175

Lee K J, Perrett D, 1997 "Presentation-time measures of the effects of manipulations in colour space on discrimination of famous faces" *Perception* **26** 733-752

Lee, T. S. (1996). Image representation using 2D Gabor wavelets. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(10): 959-971.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Neuroperception - Early visual experience and face processing. *Nature, 410*(6831), 890-890.

Liu C H, Collin C A, Burton A M, Chaurdhuri A, 1999 "Lighting direction affects recognition of untextured faces in photographic positive and negative" *Vision Research* **39** 4003-4009.

Marr, D. (1982). *Vision*. New York: W.H. Freeman and Company.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*(6), 255-260.

McMullen P A, Shore D I, Henderson R B, 2000 "Testing a two-component model of face identification: Effects of inversion, contrast reversal, and direction of lighting" *Perception* **29** 609-619

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of Facial and Manual Gestures by Human Neonates. *Science,* **198**(4312), 75-78.

Meltzoff, A. N., & Moore, M. K. (1983). Newborn-Infants Imitate Adult Facial Gestures. *Child Development,* **54**(3), 702-709.

Mondloch, C. J., Geldart, S., Maurer, D., & Le Grand, R. (2003). Developmental changes in face processing skills. *Journal of Experimental Child Psychology,* **86**(1), 67-84.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception,* **31**(5), 553-566.

Morton, J., & Johnson, M. H. (1991). Conspec and Conlern - a 2-Process Theory of Infant Face Recognition. *Psychological Review,* **98**(2), 164-181.

Moses Y, Adini Y, Ullman S, 1994 "Face recognition: the problem of compensating for illumination changes" *Proceedings of the European Conference on Computer Vision* 286-296

Olshausen, B. A. (1996). Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images. *Nature,* **381**(607-609).

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research,* **37**(23), 3311-3325.

Ostergaard A, Davidoff J, 1985 "Some effects of color on naming and recognition of objects" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **11** 579-587

O'Toole A J, Vetter T, Blanz V, 1999 "Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: an application of three-dimensional morphing" *Vision Research* **39** 3145-3155

Palmer S E, 1999 *Vision Science: Photons to Phenomenology* (Cambridge, Massachusetts: MIT Press)

Pascalis, O., Deschonen, S., Morton, J., Deruelle, C., & Fabregrenet, M. (1995). Mothers Face Recognition by Neonates - a Replication and an Extension. *Infant Behavior & Development,* **18**(1), 79-85.

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience,* **5**(12), 1332-1338.

Pellicano, E., & Rhodes, G. (2003). Holistic processing of faces in preschool children and adults. *Psychological Science,* **14**(6), 618-622.

Phillips, P. J., Moon, H., Rizvi, S., & Rauss, P. (2000). The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* ***22*****,** 1090-1104.

Price C J, Humphreys G W, 1989 "The effects of surface detail on object categorization and naming" *Quarterly Journal of Experimental Psychology* **41A** 797-828

Rhodes G, Brennan S E, Carey S, 1987 "Recognition and ratings of caricatures: Implications for mental representations of faces" *Cognitive Psychology* **19** 473-497

Russell R, Sinha P, Biederman I, Nederhouser M, 2004 "The importance of pigmentation for face recognition" *Journal of Vision* **4** 418a

Sadr, J., Mukherjee, S., Thoresz, K., & Sinha, P. (Eds.). (2002). The Fidelity of Local Ordinal Encoding*, Advances in Neural Information Processing Systems* (Vol. 14): MIT Press.

Schwarzer, G. (2000). Development of face processing: The effect of face inversion. *Child Development,* **71**(2), 391-401.

Simion, F., Cassia, V. M., Turati, C., & Valenza, E. (2001). The origins of face perception: Specific versus non-specific mechanisms. *Infant and Child Development,* **10**(1-2), 59-65.

Sinha, P. and Poggio, T. (1996) I think I know that face…, *Nature*, **384**, 404.

Sinha, P. (2002). Qualitative representations for recognition, *Lecture Notes in Computer Science* (Vol. LNCS 2525, pp. 249-262): Springer-Verlag.

Tanaka J, Weiskopf D, Williams P, 2001 "The role of color in high-level vision" *Trends in Cognitive Sciences* **5** 211-215

Tanaka, K. (2003). Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar bit slightly different stimulus selectivites. *Cerebral Cortex,* **13**(1), 90-99.

Turati, C., Sangrigoli, S., Ruel, J., & de Schonen, S. (2004). Evidence of the face inversion effect in 4-month-old infants. *Infancy,* **6**(2), 275-297.

Turati, C., Valenza, E., Leo, I., & Simion, F. (2005). Three-month-olds' visual preference for faces and its underlying visual processing mechanisms. *Journal of Experimental Child Psychology,* **90**(3), 255-273.

Ullman S, 1996 *High-Level Vision: Object Recognition and Visual Cognition* (Cambridge, Massachusetts: MIT Press)

Valentin, D., & Abdi, H. (2001). Face recognition by myopic baby neural networks. *Infant and Child Development,* **10**(1-2), 19-20.

Valenza, E., Simion, F., Cassia, V. M., & Umilta, C. (1996). Face preference at birth. *Journal of Experimental Psychology-Human Perception and Performance,* **22**(4), 892-903.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In: *Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, IEEE Computer Society Press, Jauai, Hawaii, December 8-14.

Walton, G. E., Bower, N. J. A., & Bower, T. G. R. (1992). Recognition of Familiar Faces by Newborns. *Infant Behavior & Development,* **15**(2), 265-269.

Wiskott, L., Fellous, J. M., Kruger, N., & von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **19**(7), 775-779.

Wurm L e a, 1993 "Color improves object recognition in normal and low vision" *Journal of Experimental Psychology: Human Perception and Performance* **19** 899-911

Yip A, Sinha P, 2002 "Contribution of color to face recognition" *Perception* **31** 995-1005

Young, E. D. (1984). Response Characteristics of Neurons of the Cochlear Nucleus. In C. I. Berlin (Ed.), *Hearing Science Recent Advances* . San Diego: College Hill Press.

Zeki S, 2000 *Inner Vision* (New York, New York: Oxford University Press).