Identifying Faces Across Variations in Lighting: Psychophysics and Computation

Cullen D. Jackson[*] (Cullen_Jackson@brown.edu)
Department of Psychology
Brown University

Michael J. Tarr (Michael_Tarr@brown.edu)
Department of Cognitive and Linguistic Sciences
Brown University

Athinodoros Georghiades
(athinodoros.georghiades@yale.edu)
Department of Electrical Engineering
Yale University

Humans have the ability to identify objects under varying lighting conditions with extraordinary accuracy. We investigated the behavioral aspects of this ability and compared it to the performance of the Illumination Cones (IC) model of Belhumeur and Kriegman (1998). In five experiments, observers learned 10 faces under a small subset of illumination directions. We then tested observers' recognition ability under different illuminations. Across all experiments recognition performance was found to be dependent on the distance between the trained and tested illumination directions. This effect was modulated by the nature of the trained illumination directions. Generalizations from frontal illuminations were different than generalizations from extreme illuminations. Similarly, the IC model was sensitive to whether the trained images were near-frontal or extreme. Thus, we find that the quality of the images in the training set affects how good an estimate of the complete illumination space is derived for both humans and the model. Beyond this general correspondence, the microstructure of the generalization patterns for both humans and the IC model were remarkably similar, suggesting that the two systems may employ related algorithms.

Introduction

Invariant three-dimensional object recognition has been an elusive goal in computer vision. Given the significant attention paid to this problem it is remarkable how far away we still are from a generic solution. At the same time, such a solution appears possible in that biological vision systems are capable of highly accurate object recognition across a wide range of image variability. Indeed, the apparent ability of humans to recognize objects in an invariant manner is often held up as an existence proof for the ultimate solvability of this problem.

Just how good is human object recognition? Human object recognition abilities are often characterized as being near-invariant with the few exceptions being limited to extreme cases, such as accidental views (Biederman & Gerhardstein, 1993). In particular, claims have been made for invariance across changes in viewpoint, position, and lighting, as well as possibly color, texture, and object configuration. Moreover, our intuitions tell us that we generally know what is out there despite a changing world. On the other hand, when such intuitions are evaluated empirically, they turn out to be false in many instances.

Consider the oft-examined problem of recognizing objects across rotations or observer movement. The standard approach for many years in computer vision assumed that viewpoint invariance was both desirable and attainable using recovered three-dimensional part-based models (Binford, 1971; Marr & Nishihara, 1978). Motivated in part by this stance, the most well known theory of biological object recognition has posited that objects are represented as collections of three-dimensional volumes ("Geons"; cylinders, cubes, etc.) that may be recovered in a viewpoint-invariant manner (Biederman, 1987). In support of this theory, its major proponents have claimed that human recognition performance is "typically" viewpoint invariant (Biederman & Gerhardstein, 1993).

Although there are conditions where this is true, they are few and far between, hardly typical, and only obtained by carefully following a "recipe" almost religiously (Tarr, Williams, Hayward, & Gauthier, 1998). Indeed, the results of many behavioral studies make plain that we cannot rely on our intuitions in assessing our object recognition abilities. For example, one type of experiment examined how observers generalized from one view of a simple three-dimensional volume – a Geon – to new views of the same Geon (Hayward & Tarr, 1997; Tarr et al., 1998). Despite Geons being very regular and highly distinctive from one another, across a variety of tasks and image conditions and with few exceptions, recognition of familiar Geons in new views has been found to be *viewpoint dependent*. That is, observers take progressively longer and are less accurate in recognition as a function of the rotation distance from the original view of the Geon. Similarly, "paperclip" objects (Poggio & Edelman, 1990; Bülthoff & Edelman, 1992) with single Geons inserted in the center position are also recognized in a viewpoint-dependent manner. Moreover, adding more Geons, something that theoretically should make the objects more distinctive from one another (Hummel & Biederman, 1992), actually dramatically increased the magnitude of viewpoint dependency (Tarr, Bülthoff, Zabinski, & Blanz, 1997). Illustrating the degree to which intuition fails us, following participation in a pilot version of the paperclip

[*]Department of Psychology, Brown University, Box 1853, Providence, RI 02912; tel: 401.863.2727; fax: 401.863.1300.
This publication was made possible by Grant no. 1R01EY12691 from the NEI/NIH. The authors wish to thank Peter Belhumeur, James Coughlan, Daniel Kersten, David Kriegman, and Alan Yuille for their comments and suggestions regarding this research.

experiment with single Geons, one author exclaimed that the experiment was "silly" in that he/she believed that his/her recognition of these objects was completely viewpoint invariant. Yet when their data was analyzed, their recognition performance was clearly viewpoint dependent! Thus, empirical data in one domain, viewpoint, has undermined the idea that human recognition performance is invariant and, by extension, that invariant performance should be an ultimate goal for computer vision systems. In part in response to such findings, many recent recognition algorithms intended to model biological vision rely on image-based views rather than viewpoint-invariant object models (Fukushima, 2000; Lowe, 2000; Riesenhuber & Poggio, 2000; Ullman & Sali, 2000). The critical property of all such models is that they derive object representations that preserve the appearance of object features as they appeared in the image. Thus, when new images are near to those used in the representation, recognition performance is better than when new images are distant from the original images. Hence, recognition is *not* invariant, but rather is sensitive to the manipulation of stimulus parameters such as pose or illumination.

Just what do we mean by invariant? In interpreting results from studies of human object recognition, it is important to understand what is meant by the term "invariant." One sense of "invariant recognition" literally means that performance in terms of response times and errors rates does not vary over changes in the input. A second sense implies that although response times and errors may be dependent on changes in the input, overall recognition abilities are good, with there being a high probability of identifying a given object regardless of how it is transformed. Potential confusions arise in translating human data to computer vision in that most behavioral and brain scientists use "invariant" in the first sense to characterize performance data in recognition tasks. That is, when they refer to "invariant recognition," they mean cases where there is little or no sensitivity to a stimulus manipulation. For example, obtaining the same response times and errors rates in identifying an object at different viewpoints (Biederman & Gerhardstein, 1993).

Conversely, when such behavioral and brain scientists refer to "dependent recognition," as in viewpoint dependent or illumination dependent, they are not referring to a condition where recognition completely fails given changes in the stimulus. Rather, they are characterizing the *mechanisms* whereby relatively invariant recognition is achieved. View-sensitive recognition mechanisms that take more time and are less accurate as a stimulus is rotated in depth away from a familiar view nevertheless generally support recognition across such transformations. Observers are simply a bit slower and less likely to be correct for the transformed as opposed to the original viewing conditions. Thus, although human recognition is not invariant in the first sense, it is invariant in the second sense. For purposes of linking human and machine vision this is a critical point – near-invariant recognition is attainable, but the *algorithms* whereby it is attained are not themselves invariant. As we shall see, it is precisely this lack of invariance in the mechanisms of recognition that informs us regarding the algorithms used by humans and allows us to compare human abilities to those of computer vision systems.

Illumination dependence in human object recognition. As mentioned earlier and discussed above, the study of visual object recognition in humans and other primates has focused largely on the problem of recognition across changes in viewpoint (for example, see Logothetis & Pauls, 1995; Tarr, 1995). Other manipulations that have been assessed in at least a few studies include transformations in size, position, mirror-reflection, and surface detail. For the most part, however, one of the most dramatic transformations of an image has been ignored – the recognition of objects over changes in lighting direction. Why is this so? Computer vision scientists know that a shift in illumination severely alters an image of an object – often to such a degree that the object becomes unrecognizable. Surely such a problem is obvious to behavioral and brain scientists?

There are several reasons why recognition across changes in lighting direction has been omitted from the extant behavioral literature. First, until recently computer graphics technology capable of creating realistic lighting effects (shading gradients, specularities, and soft shadows) was both difficult to use and expensive and therefore essentially inaccessible to most behavioral and brain scientists. Second, creating well-controlled manipulations in lighting direction in the physical world is tedious and time-consuming and, therefore, less appealing than many other potential transformations, e.g., moving a camera around an object. Third, although there has been awareness about the fact that lighting affects the shading gradients on a object's surfaces and that such information can be used to infer three-dimensional shape (Horn, 1975; Ramachandran, 1988), it has been less obvious that the effects of a particular illumination context might affect an object's *representation*. That is, although lighting clearly influenced processes involved in the derivation of representations of three-dimensional objects, it was not thought to impact the ultimate organization of such representations – these being illumination invariant.

The most salient version of this type of thinking was found in the hypothesis that object representations are edge-based (Marr, 1982; Biederman, 1987). The key idea was that an object's edges were stable across variations in the image and that from such canonical edge descriptions, completely invariant, three-dimensional object models could be derived. One strong critique of this approach is that edge maps are rarely stable over even relatively small changes in the image. Rather they are noisy and sensitive to variations in shading gradients and specularities. Thus, edge-based descriptions do not offer a likely basis for human object recognition (Sanocki, Bowyer, Heath, & Sarkar, 1998). Similarly, without a model of the lighting for a given scene, it is difficult, if not impossible, to discount edges that arise from shadows as opposed to object contours. Thus, the edge descriptions for an object under two different lighting directions may be drastically different from one another.

Given the non-viability of edge-based models, as well as other lighting-invariant representational schemes, Tarr, Kersten, and Bülthoff (1998) used computer graphics to explore the question of whether human object recognition was truly invariant with respect to variations in illumination. In part, they were motivated by the finding that cast shadows help constrain the perceived three-dimensional layout of a scene (Kersten, Knill, Mamassian, & Bülthoff, 1996; Kersten, Mamassian, & Knill, 1997). Three notable findings emerged from Tarr et al.'s study: 1) Novel objects learned under one lighting direction were more poorly recognized when shown under a new lighting direction; 2) This illumination dependence was obtained only when attached shadows were present in the scene; 3) Overall recognition performance, although lighting invariant, was worse in the absence of attached shadows. Thus, shadows and shadow edges seem to be included in object representations for one very good reason – although they produce some lighting dependence in recognition this is outweighed by the fact that shadows help to disambiguate the three-dimensional appearance of objects. It should be noted that the costs for changing lighting direction were relatively small and that overall recognition accuracy was quite high under both familiar and unfamiliar illumination conditions. Again, the key point is that it is the *pattern* of dependence in performance that informs us regarding the mechanisms underlying the human ability to attain near-invariant recognition in practice.

It is worth noting that this study was restricted to novel, relatively simple objects composed of a small number of three-dimensional volumes. Left open was the question of whether such effects of lighting direction would impact known object classes in a similar manner. Indeed, a study by Moore and Cavanagh (1998) suggested that familiarity with the identity of an object might facilitate invariant recognition over different lighting conditions. They found that the ability of observers to recognize illuminated three-dimensional objects rendered as two-tone or binary images depended on whether the objects were familiar or unfamiliar to the observers. When shown as two-tone images, known objects were nameable while unknown, novel objects were not (until observers were shown the unknown objects as shaded, photorealistic images). Therefore, both the sensitivity to lighting direction and the overall recognition advantage seen for objects rendered with attached shadows might break down for familiar objects.

Although the behavioral literature is sparse, there are hints that the recognition of some familiar object classes is lighting dependent. Most notably, Johnston, Hill, and Carman (1992) reported on the well-known horror-film phenomenon that human faces lit from below look very different from the same faces lit from above. Braje, Kersten, Tarr, and Troje (1998) explored this somewhat more systematically, finding that human faces lit from one side were recognized more poorly when shown with the light moved to the other side. Importantly this lighting dependence was obtained both with and without shadows on the faces. Thus, the representation and recognition of at least one highly familiar object class, human faces, is lighting dependent.

Illumination dependence in machine vision systems. As stated earlier, that illumination greatly influences the appearance of an object has not gone unnoticed in the computer vision community. Over the years many different approaches have been proposed to deal with the fact that changing the direction of lighting can impact mean illumination, shading gradients, shadows, and specularities. For whatever reason, particular emphasis has been placed on recognizing human faces across variations in lighting. The degree to which face recognition is affected by lighting was elucidated by Moses, Adini, and Ullman (1994) who pointed out that changes in illumination accounted for the greatest image variance in measuring the differences between images of faces, even more than the variance that arose between individual identities or changes in viewpoint. Despite the fact that variations in lighting dramatically change an image, lighting information may also facilitate the recovery of the shape and the structure of a face. Consequently, it would be less than optimal to deal with lighting variation by completely discarding it, thereby reducing the probability of correctly recognizing a face under ambiguous conditions. Thus, computational models of face recognition have often focused on how to compensate for illumination variability across multiple images in a manner that also allows for some representation of the lighting.

One approach to lighting variability that has recently become quite popular relies solely on two-dimensional images, rather than the explicit recovery of three-dimensional scene parameters. These image-based/appearance-based models reduce the dimensionality of the image space (each pixel value in an image being a coordinate in image space) by projecting it onto a lower-dimensional feature space. The object is then recognized by using a nearest neighbor classification scheme in the new feature space. Three of the most widely cited versions of this general method are referred to as Eigenfaces, Fisherfaces, and Illumination Cones.

Eigenfaces. A technique common to computer vision for the reduction of dimensionality is principal components analysis (PCA) also known as Karhunen-Loeve expansion. Given a set of sample images of different individual faces, PCA produces a linear projection that maximizes the determinant of the total co-variance matrix of the samples across all individuals. The resulting eigenvectors of this matrix have the same dimensionality as the sample images. Since the eigenvectors characterize the feature space, each individual face is represented by a linear combination of the eigenvectors. This technique is sometimes called "Eigenfaces" (Turk & Pentland, 1991a, 1991b). However, since PCA maximizes the total scatter in the sample images, both the between-individual variance and the within-individual variance are retained. For recognition, only the between-individual variance is useful. In fact, the retention of the within-individual information allows changes in illumination between images to influence the resulting feature space. Since variations due to illumination are usually larger than those due to individual identity, this can cause the feature space to become muddled and, consequently, recognition is difficult.

To compensate for the variability introduced by illumination, Belhumeur, Hespanha, and Kriegman (1997) and Georghiades, Kriegman, and Belhumeur (1998) point out that the first three principal components are primarily due to changes in illumination and might be discarded for purposes of recognition. In practice this makes explicit the fact that Turk and Pentland (1991a, 1991b) used only the most appropriate eigenvectors in their recognition algorithm. Georghiades and his colleagues also implemented this approach in their comparison of several models of face recognition under variations in illumination. Consistent with their observation, the Eigenface method did achieve better recognition performance without the first three principal components than it did with them. However, even with this addition, under some lighting conditions the Eigenface model failed 27% of the time and it showed a gain of only 16% over the Eigenface model with all principal components.

Fisherfaces. To address the poor performance of the Eigenface model across lighting variation, Belhumeur et al. (1997) proposed an alternative method that attempted to retain the benefits of linearly reducing the image space into a low-dimensional feature space, but avoid the problems of the Eigenface method. Belhumeur et al. (1997) used Fisher's Linear Discriminant to maximally separate the between-individual and within-individual variance in order to provide more consistent discrimination between individual faces. By maximizing the ratio between the determinants of the between-individual and within-individual scatter matrices, they were able to produce a set of eigenvectors with reduced dimensionality but without the confound of variability due to illumination. To avoid a singular within-class scatter matrix, Belhumeur et al. (1997) created a technique they refer to as "Fisherfaces". This method uses PCA to reduce the dimensionality of the image space, and then applies Fisher's Linear Discriminant to reduce further the dimensionality of the feature space and achieve a nonsingular within-class scatter matrix. The authors empirically demonstrated that this technique could successfully recognize individual faces over broad variations in lighting across the sample images. Indeed, under the same lighting conditions where Eigenfaces produced 27% errors, Fisherfaces resulted in only a 5% error rate. Thus, the application of Fisher's Linear Discriminant following PCA provided a significant gain in illumination invariance.

Illumination Cones. Following the development of the Fisherface model, Belhumeur and Kriegman (1998) proposed an even more effective image-based model for object recognition under variable lighting and viewpoint conditions. This model is referred to as the "Illumination Cone" (IC) method. An illumination cone is a convex polyhedral cone containing the set of images for an object under all possible point light sources. The dimensionality of the image space is reduced since it is projected onto object-specific illumination cones which have dimensionality equal to the number of distinct surface normals. Multiple images for each object are used as a basis set to construct individual illumination cones. As few as three distinct images for each object can determine a given object's cone. Critically, no explicit knowledge of the lighting parameters of the scene is required to construct the cone. While the IC method was designed for use with convex objects with Lambertian surface reflectance, several empirical studies have shown that the method is quite capable of performing excellent recognition with non-Lambertian, non-convex objects like faces (Georghiades et al., 1998, 2000). As a measure of the effectiveness of the IC model, a separate comparison of the Eigenface, Fisherface, and IC approaches produced error rates of 78%, 51%, and 37% respectively. Thus, with regards to illumination invariance, the IC model offers potentially far better performance than competing approaches.

Applying the IC Model to Human Performance. Given the superior performance of the IC model, it was used as the standard for comparisons with human vision. In particular, given that we assume human face recognition performance is at least as good across lighting variation as the best currently available computer vision algorithm, even the IC model may be at somewhat of a disadvantage. Adding to the unevenness of this comparison, human observers have years of experience at face recognition and presumably apply this class-level knowledge to the recognition of even entirely novel faces. In contrast, the IC model has no knowledge of faces beyond the training it receives at the beginning of each experiment. On the other hand, human observers are processing faces in the context of a wide array of potential objects, whereas the IC model knows only about faces.

Even considering these differences between humans and extant models, there is much to be learned by comparing the two. First, incredibly little is known about how humans generalize from known to unknown lighting conditions. Therefore, to the extent there is any correspondence between the biological and machine systems, we have learned something regarding the types of models that might be most effective in accounting for human performance. Second, there is the possibility that there will be significant correlations between human and model performance. In this case, we can not only draw stronger conclusions, but we can then refine future algorithms in the direction of biological plausibility.

In order to provide the most useful comparisons between human subjects and the IC model, we implemented similar training procedures for both cases. Moreover, to provide a more general picture of how both compensate for lighting variation, we chose to include experiments which used "extreme" training conditions, e.g., lighting directions in which the majority of the face was not illuminated, as well as more "standard" training conditions, e.g., frontal illumination. Interestingly, few, if any, computer vision studies have ever subjected their models to more than the standard cases during training.

Thus, the experiments presented here are useful for understanding the behavior of the IC model independent of the comparisons with human observers. Finally, there are almost no studies of how humans deal with extreme lighting during training, so again, the data obtained here is valuable in and of itself. We would hold, however, that the most informative analysis is the comparison between the IC model and human behavior. Such specific quantitative comparisons between a working model from computer vision and behavior are not frequent in the human psychophysical literature, yet they provide a promising method for furthering our understanding of algorithms in biological vision. We next review the specific methods used in training and testing both our human subjects and the IC model.

Methods

Human Psychophysics

Observers. The subjects for the five experiments were 106 human beings, mostly college students, between the ages of 18 and 22 years. There was a median of 20 subjects across the five experiments, with an equal number of males and females in each. All subjects had normal or corrected-to-normal vision. Subjects were naïve to the purpose of each experiment. When finished with the session, observers were informed of the intent of each experiment.

Apparatus and Stimuli. Stimuli were presented to the observers on one of three Apple PowerMac 8100s with NEC MultiSync XV15+ monitors. Connected to each were an Apple Extended Keyboard II and Apple Bus Mouse. The experiments were all programmed and run using the RSVP Experimental Control Software (Williams & Tarr, 2001).

All images and text in the experiments were displayed at 640x480 pixels of resolution. A strip of paper was placed above the numbers at the top of the keyboard. The strip was 22.7 cm long and 1.9 cm inches wide. The following ten names were placed horizontally from left to right along the center of the strip: Allen, Carla, David, Gary, Janet, Laura, Michael, Nigel, Robert, and Tony. The strip was positioned such that the first named (Allen) was placed above the '1' key and the last name (Tony) was placed above the '0' key. During the experiment, observers made responses by pressing the number keys on the keyboard under the name that the subject associated with the given stimulus on the screen.

A study sheet was given to each observer at the beginning of the experiments. The study sheet consisted of ten images printed on a sheet of paper. The images were placed in two rows on the sheet, five images along the top and five images along the bottom of the sheet. Each image was 295 x 338 pixels (screen size) and 2.9 x 3.4 cm printed. Names for the images were placed below the faces. The individuals in the images were facing forward and the illumination on each face was from the front (see Figure 1).

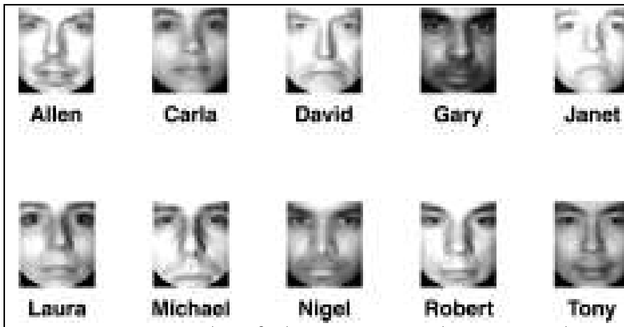

Figure 1. Example of the images and names subjects viewed prior to the start of the computer-based trials. Subjects viewed these images for 10 minutes in order to learn to associate the correct name to each face.

The images used in the computer-based trials were 651 images taken from the Harvard Face Database (Hallinan, 1994, 1995; available at: http://www.cog.brown.edu/~tarr/stimuli.html#ha). The images taken from the database represent 10 different individuals viewed under 66 different illumination conditions. The individuals are in a fixed frontal pose for all illuminations. The lighting space was sampled in 15° increments both horizontally and vertically to the right of the camera axis. A schematic of the illumination conditions is shown in Figure 2. While most of the individuals have 66 images, three individuals have missing or corrupted images and have less than the 21 images normally associated with the region 75° from the camera axis. This set of images was used in order to replicate the methods used for the computational experiment presented in Georghiades, Kriegman, and Belhumeur (1998) in a human psychophysical experiment.

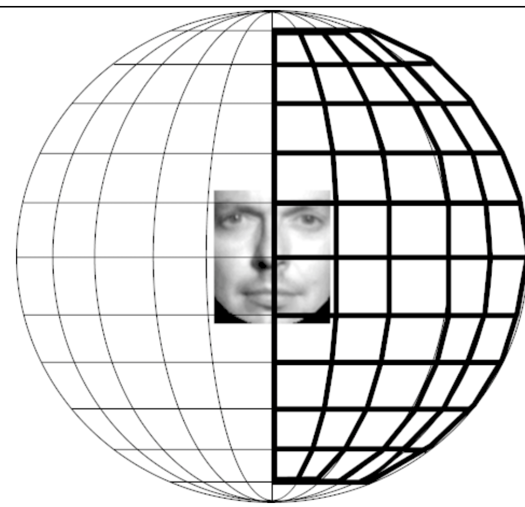

Figure 2. Schematic of the 66 different illumination conditions used in both the human psychophysical and IC model experiments. The illumination conditions correspond to the intersections of the longitudes and latitudes overlaid with bold lines. The center of the space is denoted (0°, 0°). Adapted from an illustration in Georghiades, Kriegman, and Belhumeur (1998).

Procedure. Each experiment consisted of three phases: name learning, training, and testing. In the first phase, observers were asked to study a sheet of 10 faces with corresponding names for 10 minutes. This time allowed subjects to learn to associate the correct name with each face. This phase was necessary because the observers were asked in subsequent computer trials to identify other images by the name associated with a face on the study sheet. We told observers not to consult the study sheet once the 10-minute study period ended. Observers were told to turn the study sheet over so that the blank side faced up and to set it aside during the remainder of the experiment.

The training phase familiarized observers with a small subset of illuminations for each face. This phase consisted of 60 computer-based trials. In each trial, observers first viewed a blank, white screen for 1000 ms followed by a 500 ms fixation point (+) in the center of the screen. The stimulus image then appeared centered on the screen for 1000 ms. Observers then had as much time as needed to correctly identify the image. Subjects used the number keys (1, 2, 3...0) at the top of the keyboard, each labeled with one of ten names, for their responses. A feedback sound (the Macintosh default system "beep") indicated an incorrect response and no beep indicated a correct response.

The testing phase of the experiment gauged how well subjects had learned the representations of each face during training. This part of the experiment contained 591 trials. Each trial was identical in design to a trial in the second phase, except the stimulus images were shown for 500 ms and observers only had 3000 ms to make a response before the trial timed out. If no response was made, it was recorded as such and was subsequently dropped form further analysis. No feedback was given during this phase.

Each of the five experiments only differed in the set of training images shown to subjects. These training sets are illustrated in Figure 3. The training sets for the first, second, and fifth experiments used six illuminations per face. The third and fourth experiments only contained one illumination for each face shown to subjects six times during training.

Model Simulations

While replicating the method used by Georghiades et al. (1998) with human psychophysical experiments, it also was necessary to test the Illumination Cone (IC) model under similar training conditions as used in the behavioral experiments. Subsequently, the IC model was trained using the same sets of illuminations as in the human behavioural experiments, with some modifications to help equate the experience of humans and the computational model.

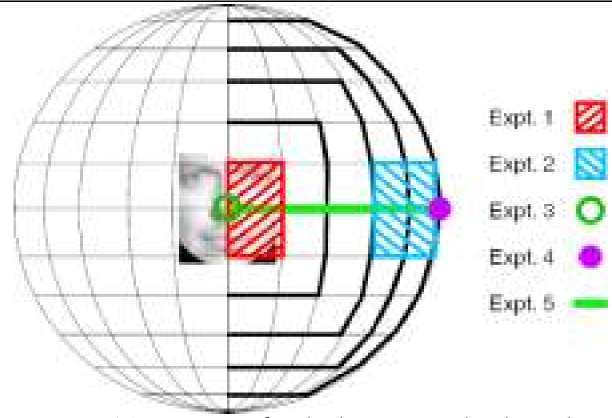

Figure 3. Training sets for the human psychophysical and computational model experiments. The training set for Experiment 1 contains illuminations within 15° of the camera axis. The training set for Experiment 2 is a mirror of Experiment 1 with extreme lighting directions. Experiments 3 and 4 only have one illumination condition each, (0°, 0°) and (75°, 0°), respectively, for training. The training set for Experiment 5 contains the illuminations along the horizontal meridian of the illumination space, from (0°, 0°) to (75°, 0°). Adapted from an illustration in Georghiades, Kriegman, and Belhumeur (1998).

In the rendering of the Illumination Cone model used, the training phase occurs with the construction of the illumination cone for each individual face (Belhumeur & Kriegman, 1998; Georghiades et al., 1998, 2000). These cones are labeled with the correct name of the individual face for later recognition. Thus, the IC algorithm combines the tasks of name learning and training which the human subjects performed during the comparable psychophysical experiments. Since human subjects learned the name for each face separately from the training task, they always received input about the frontal (0°, 0°) illumination condition separate from the illuminations in the training set. Due to this additional illumination component, the IC model was also given the (0°, 0°) condition during training for all of the experiments in which this illumination was not already a part of training, except for Experiment 2 in which the addition of the (0°, 0°) component seemed to hinder the performance of the IC model.

Another change in the training sets used with the IC model occurred for Experiments 3 and 4. In the corresponding human psychophysical experiments, subjects only received *one* illumination condition repeated six times during training. However, the IC model requires at least *three* different images in order to construct the illumination cone for each face. Moreover, humans already know a great deal about how the appearance of faces is generally affected by lighting direction. In order to fulfill the need for three different images and compensate for pre-existing knowledge in human subjects, in both Experiments 3 and 4 the IC model was trained using a set of seven different illumination conditions, of which six were randomly selected and the seventh was either (0°, 0°) or (75°, 0°) respectively. These training sets are illustrated in Figure 4. This was a departure from previous training since the other experiments used uniformly-defined light conditions, either within 15° of one single light or along the same lighting axis. Also, since faces are a known class to humans and observers usually generalize well to unknown faces, by randomly choosing the training conditions, we could determine how well the IC model would generalize to other unknown conditions given a non-uniform set of illuminations. An alternative method might be to choose random illuminations from several regions of the light sphere so that the entire light sphere would be represented in the training set. This method might provide a more robust generalization of the entire light space.

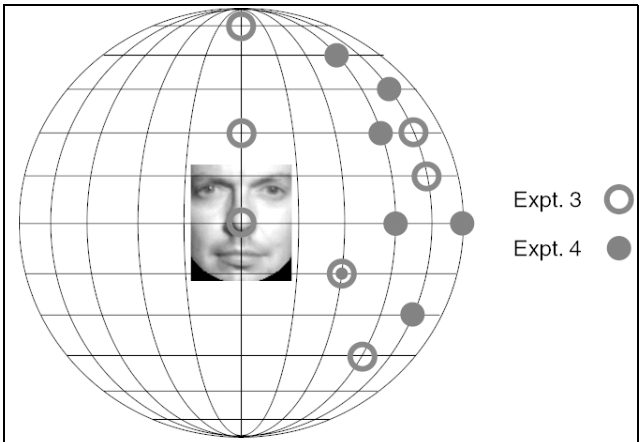

Figure 4. Training sets for Experiments 3 and 4 for the Illumination Cone (IC) model. Open circles are lighting coordinates for Experiment 3 and filled circles are coordinates for Experiment 4. Coordinate (30°, -15°) is a training illumination for both experiments. Adapted from an illustration in Georghiades, Kriegman, and Belhumeur (1998).

Another component of the IC model was building the basis matrix used to construct the illumination cone for each face. This process involved setting two parameters, one a saturation threshold and the other a shadow threshold. In order to build the three basis vectors for each face, the training images were input into the algorithm, reflected along the vertical axis (in order to double the number of images used in the calculations), thresholded (according to the parameters discussed above), and then reduced into three component vectors. These basis vectors were then used to construct and label the illumination cones for each individual face. Subsequently, the illumination cone built for each individual contained representations of all 66 illuminations in the lighting space for that face. This enabled the model to later identify the individual face in a novel image by comparing the image to the illumination cones and choosing the cone with the closest representation using a nearest-neighbor algorithm computed through a non-negative least squares solution.

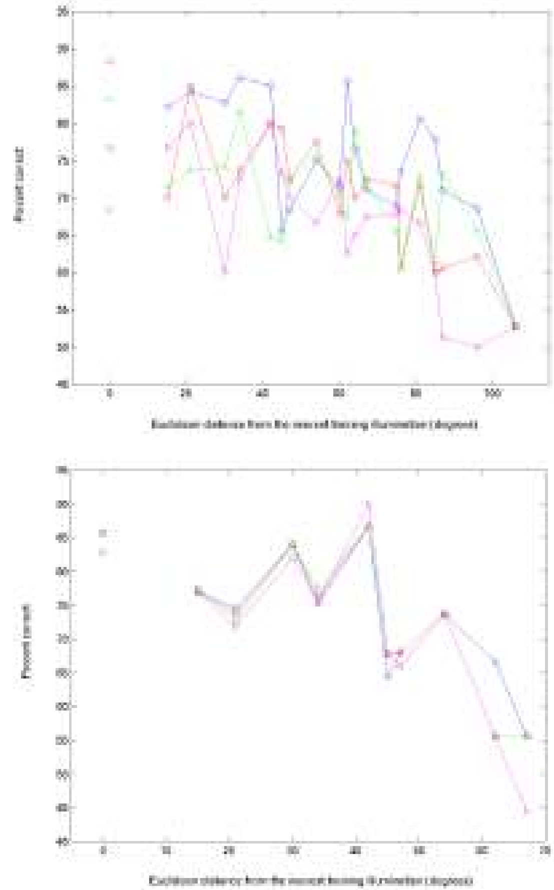

Figure 5. Differences between individual human subjects are similar to differences between subtle changes in the IC model parameters. Each connected line in the top graph shows an individual human subject. Each connected line in the bottom graph is a separate run of the IC model with slightly changed saturation thresholds. The shadow thresholds were not changed due to model constraints.

While the performance of the IC model does not vary across multiple instances of the same experiment, it does vary according to the threshold parameters discussed above, saturation and shadow. As previously discussed, changing these parameters for an individual face changes the creation of the illumination cone for that face. Thus, the subtle manipulation of these parameters is analogous to having different individual human subjects in the psychophysical experiments. This phenomenon is illustrated in Figure 5. The graph at the top of the figure shows four human subjects in Experiment 3, and the graph at the bottom shows four different saturation thresholds for the IC model in the same experiment. The differences between the four saturation thresholds are within several hundredths of each other as this parameter varies between 0 and 1.

Results and Discussion

Similar analyses were run on the data from the human psychophysical and the IC model experiments. Any modifications between the two are presented below. For all of the experiments, the lighting coordinates for each image were recorded and the Euclidean distance from the nearest training illumination to that coordinate was computed. For clarity of presentation, these distances were then mapped to the most appropriate lighting condition bin: 15°, 30°, 45°, 60°, or 75°. Figure 6 shows an example of the different illumination conditions that comprise these five bins for the training set from Experiment 1. The dependent variable across all experiments was percent correct recognition, that is, identifying the individual face, for each lighting condition for both human subjects and the IC model.

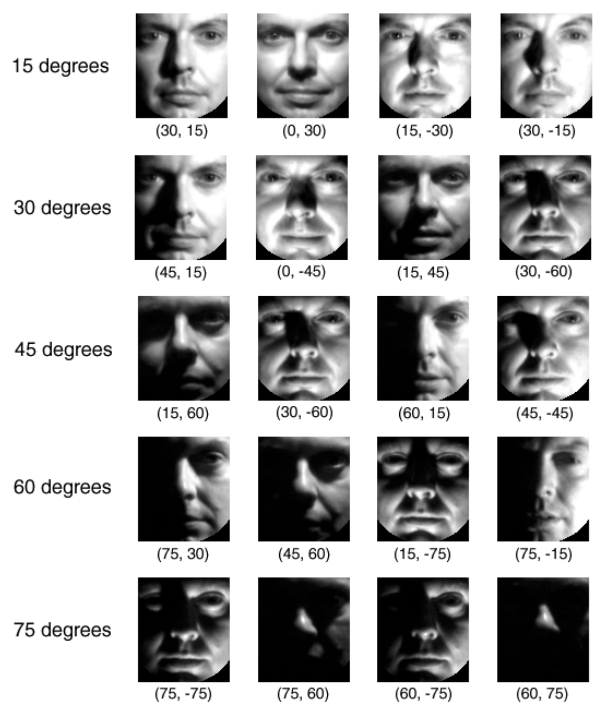

Figure 6. Sampling of test images for Experiment 1 and their respective distances from the closest illumination in the training set. The training set for Experiment 1 consisted of illuminations at the following coordinates: (0°, 0°), (0°, 15°), (0°, –15°), (15°, 0°), (15°, 15°), (15°, –15°).

Human Psychophysics

Subjects that failed to respond to more than 10% of the test trials were removed from the study. This procedure ensured that the results contained only observers who made actual responses to a majority of the test trials. This reduction in the data changed the median number of subjects across experiments from 20 to 18. For the remaining subjects, trials with no response were dropped from all analyses. The mean percent correct recognition for each test lighting condition (15°, 30°, 45°, 60°, and 75°) across subjects and the within-subject standard error of the mean were computed and are illustrated as the human psychophysical data marked as circles in Figures 7-11 (Experiments 1-5 respectively).

Across all five experiments, as the distance between the test illuminations and the training set increased, the identification performance of the subjects decreased. This performance drop-off was most pronounced in Experiments 1, 3, and 5, where the images in the training sets contained mostly frontal or near-frontal illuminations or, in the case of Experiment 5, the trained lighting conditions were all along the horizontal meridian of the lighting sphere. In contrast, the performance decrease across lighting conditions was less apparent in Experiments 2 and 4 where the training sets contained extreme illuminations that produced images with pronounced shadows.

One explanation for the fall-off in performance with distance in Experiments 1 and 3 is that the 60° and 75° lighting conditions included images with extreme illuminations. Because so much of the face was in shadow, subjects could not discern the identity of the face. In contrast, in Experiments 2 and 4, subjects actually saw these extreme illuminations during training and were therefore able to identify the individual faces at test in these otherwise difficult-to-recognize lighting conditions. Furthermore, the 60° and 75° lighting conditions were comprised of near-frontal illuminations, therefore identity was easy to establish despite the unfamiliarity of the lighting direction.

Note that although during the initial name association task subjects had the benefit of viewing the frontal (0°, 0°) illumination condition before the start of each experiment, this experience alone did not dramatically help them in the subsequent recognition tasks. For example, in Experiments 2 and 4, they were still worse for this lighting condition compared to the extreme lighting conditions that were seen during training. This is surprising because the most typically encountered "real-world" image, e.g., canonical, of a face is likely to be the frontal view with frontal illumination (from above); consequently we would expect human subjects to perform better with familiar near-frontal illuminations (e.g., the images in the 60° and 75° lighting conditions in Experiments 2 and 4).

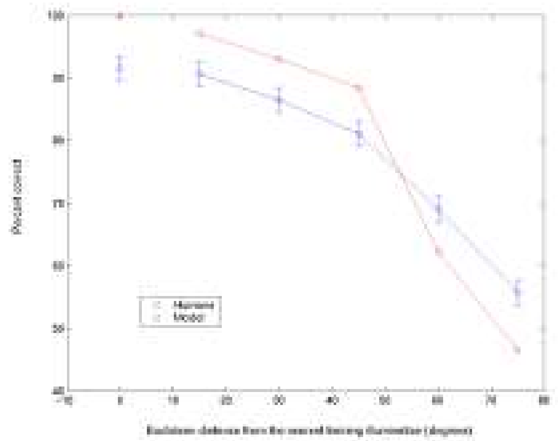

Figure 7. The circles represent mean percent correct for the human subjects in Experiment 1. The error bars are the within-subject standard error of the mean. A single case of the Illumination Cone (IC) model for Experiment 1 is represented by the squares.

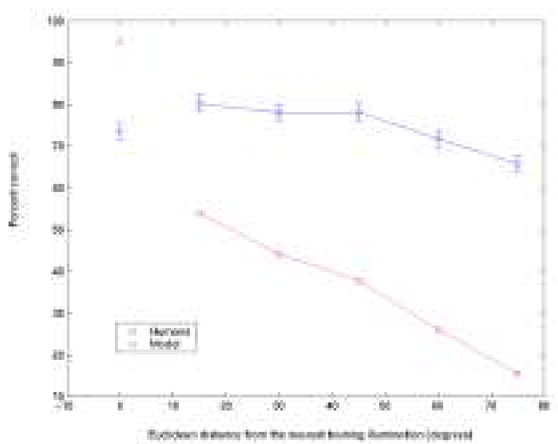

Figure 8. The circles represent mean percent correct for the human subjects in Experiment 2. The error bars are the within-subject standard error of the mean. A single case of the Illumination Cone (IC) model for Experiment 2 is represented by the squares.

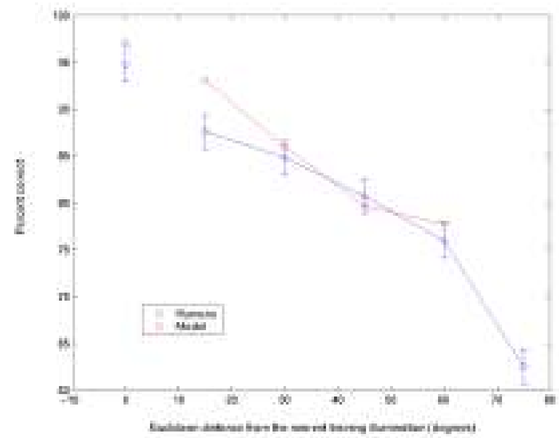

Figure 9. The circles represent mean percent correct for the human subjects in Experiment 3. The error bars are the within-subject standard error of the mean. A single case of the Illumination Cone (IC) model for Experiment 3 is represented by the squares. Note that the IC model was only tested with four new illumination types (bins) because of the manner in which lighting directions in the training set were randomly selected. See text for an explanation for this selection procedure.

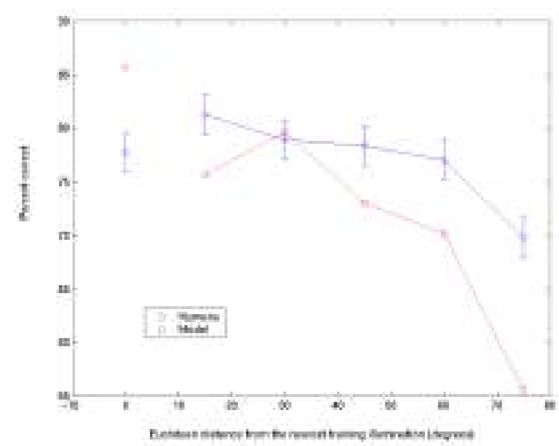

Figure 10. The circles represent mean percent correct for the human subjects in Experiment 4. The error bars are the within-subject standard error of the mean. A single case of the Illumination Cone (IC) model for Experiment 4 is represented by the squares.

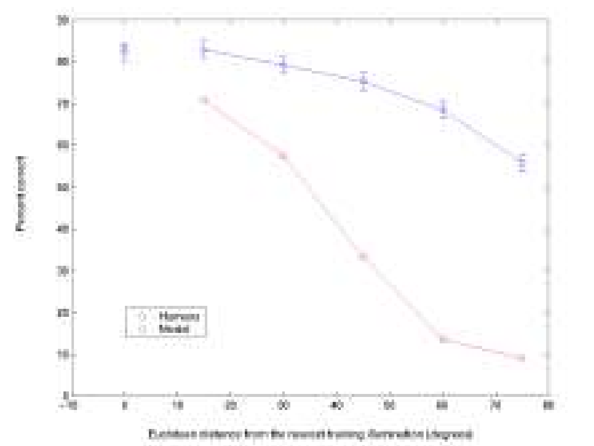

Figure 11. The circles represent mean percent correct for the human subjects in Experiment 5. The error bars are the within-subject standard error of the mean. A single case of the Illumination Cone (IC) model for Experiment 5 is represented by the squares.

Illumination Cone (IC) Model

The data points marked with squares in Figures 7-11 illustrate the recognition performance of the IC model and can be compared to the performance of the human observers in Experiment 1-5. We attempted to maximize the performance of the IC model by trying a variety of saturation and shadow thresholds for each face in each experiment. The resulting model performance in each figure represents the best attempt at manipulating these thresholds. Because of the variable nature of the IC model due to these parameters, the recognition performance shown for each experiment may not actually represent optimal performance. Evidence for this exists in the recognition performance of the model for the training illuminations. Only in Experiment 1 does the model correctly recognize all of the faces at the training illuminations.

Importantly, as with our human subjects, the recognition performance of the IC model in each experiment decreased as the distance of each test illumination from the training set increased. The only minor exception to this pattern occurs in Experiment 4; in this experiment, recognition performance actually increased by 5% between 15° and 30° and then resumed its downward trend. This change in pattern was probably due to the span of the training illuminations across the lighting sphere. Since the training conditions were randomly selected for this experiment, the illuminations comprising the 30° bin may actually map onto the average of the illuminations in the training set. Another consequence of using a randomly selected set of lighting directions for the training set was that Experiment 3 included only four test lighting condition "bins". Specifically, when the distances of each test illumination from the training set were computed, none of the lighting directions extended beyond the 60° bin (the greatest distance being 54°). Similarly in Experiment 4, the greatest distance of any tested illumination from the training set was 67°. However, because this lighting direction was equidistant between the 60° and 75° bins, and its performance was poor in relation to the 54° and 62° lighting directions, it was placed into the 75° bin.

As with our human subjects, the IC model was sensitive to whether it was trained with near-frontal or extreme illuminations. Specifically, in Experiment 1, training with near-frontal illuminations produced 45% correct performance at its worst. In contrast, in Experiment 2, training with extreme illuminations produced 15% correct at its worst. The results of Experiments 3 and 4, analogous to Experiments 1 and 2, but with single lighting directions at training, were less clear-cut in that we attempted to approximate human experience by including six additional randomly selected lighting directions (for the specific images used, see Figure 4). Even still, performance for Experiment 3, 78% correct at its worst, was better than for Experiment 4, 72% correct at its worst, again indicating that the IC model is sensitive to the quality of the images with which it is trained.

Our interpretation of these results is that, as in the human psychophysical experiments, extreme shadows in the training images do not allow the IC model to construct a robust representation of the lighting space for each face. In contrast, when the faces are clearly illuminated during training, the IC model is apparently able to create a much better approximation of the actual lighting space.

Finally, Experiment 5 used training lighting directions that spanned both near-frontal and extreme lighting directions, all lying along a horizontal meridian of the lighting sphere. Here performance was remarkably poor at all but the test lighting condition closest to the training set. This result suggests that the IC model is quite bad at creating a generic lighting model for an object when the trained lighting directions are all accidentally aligned. Apparently when lighting changes only interact with a singular aspect of an object's geometry, insufficient information is available regarding how the surfaces of an object will appear when illuminated from an orthogonal direction. This is exemplified by the fact that the IC model was quite bad at identifying faces when the lighting direction was shifted vertically relative to all training illuminations.

Comparing Psychophysical and Computational Results

How do we compare the performance of a putative model of human performance with actual observations of human performance? One approach would be to take the comparison at face value and simply assess where the model performance is better, where human performance is better, and where the two are essentially the same. However, this sort of comparison is rife with peril in two respects. First, nearly every extant computational model of vision deals only with a small part of the "vision problem"; in contrast, the human observer is always applying a complete vision system which includes massive early filtering, sophisticated mid-level organization, and a remarkably rich representational space. Thus, there is little reason to believe that a model should perform anywhere near as well as a human (let alone a pigeon or rat!). Second, even if a model did approach the level of human performance, it would most likely be doing so for very different reasons. Thus, the fact that the IC model actually does in many cases come close to the absolute performance of our subjects is not particularly informative.

Given this context, is it possible to make any statement about models of computer vision vis a vis human vision? Indeed it is. Specifically, a good computer vision model intended to capture some aspect of human vision is responsible only for *that* aspect. That is, it would be a mistake to claim that a given model does any more than model one specific mechanism of human vision. In the case of the IC model, that specific mechanism is how a recognition system generalizes from known to unknown lighting conditions for a given image of an object. This mechanism is but one factor that mediates the overall performance of the larger system, but it is the particular element that mediates how performance modulates across light variation. This means that the *patterns* of generalization from known to unknown light conditions may be compared for the IC model and our human subjects (although we are intrigued by the fact that model accuracy and human accuracy are sometimes quite close!).

The overall general gradient of performance between the IC model and humans is illustrated in the correlations shown in Table 1. The first set of correlations represent the data with the training sets included in the calculations, while the second set does not contain this data. The correlations without the training sets are probably more representative of the pattern between human performance and the IC model. As mentioned earlier, because the model is not a perfect representation of actual human vision but merely a possible representation of a specific chunk, it always performs very well on images that were previously viewed, i.e., training images. On the other hand, humans can apply a great deal of class knowledge to the recognition of faces, so although they are always poorer at recognizing training images, they are typically much better at test images than the IC model. That is, humans generally know how the appearance of a human face will change with changes in lighting direction and they can use this general information to make inferences about the appearance of specific faces.

Experiment	With Training Data	Without Training Data
1	0.992	0.991
2	0.388	0.962
3	0.966	0.944
4	0.796	0.941
5	0.896	0.905

Table 1. Correlations (Pearson's *r*) between the IC model and psychophysically assessed human subject performance for Experiments 1-5.

The Microstructure of Generalization. Beyond the fact that as test images were further and further from training images, both human observers and the IC model exhibited a general decrease in recognition accuracy, there is the question of how specific lighting directions affected performance. A general characteristic of the human-model comparison is the degree of similarity in the deviations from linearity in both recognition functions. In certain cases, when there was a deflection in the response of humans, a similar deflection was found for the model; other times this was not the case. However, some of the more subtle similarities between the two patterns of performance are not visible in Figures 7-11 because we "binned" the data into five qualitative categories for purposes of clarity of presentation. In particular, in the raw data, there was generally an inflection point in performance at the 45° distance from training for both human observers and the IC model. Another specific similarity between humans and the IC model is that both showed poorer overall performance in Experiment 5 relative to their own respective performance in Experiments 1-4. Thus, both human subjects and the IC model appear sensitive to an accidental alignment of all lighting directions during training. Such microstructure comparisons are important for understanding exactly how humans and computational systems compensate for lighting variability and should be explored in more detail in future studies.

General Discussion

To date there has been surprising little work on how biological systems compensate for variations of lighting in a scene. To some extent this stems from a failure to recognize the difficulty of the problem and the assumption that edge-based models are able to produce lighting-invariant descriptions. Other factors include an inability to readily generate stimuli under varying lighting conditions and a lack of models that make any concrete predictions about the representation of lighting information and its impact on recognition. At the same time, because they were often intended as working, real-world systems, recognition models in computer vision ran head on into the problem of lighting. As the lighting direction shifts, mean illumination, shading gradients, shadows, and specularities on an object may all change in dramatic fashion.

Recent work brings together these two threads. First, several studies of human recognition under varying lighting conditions reveal that we are indeed sensitive to changes in illumination (Tarr et al., 1998), even for highly-familiar classes such as faces (Braje et al., 1998). Second, unlike some earlier models within computer vision (e.g., Turk & Pentland, 1991a, 1991b), a new image-based approach to object recognition allows for the recognition of objects across varying lighting conditions (Hallinan, 1994; Belhumeur & Kriegman, 1998). Consider the fact that human vision is sensitive to variations in lighting. Although the inclusion of lighting parameters in high-level object representations may seem inefficient at first glance, there is evidence that such information is critical in the disambiguation of three-dimensional structure, particularly for unfamiliar objects (Tarr et al., 1998) or under-constrained scenes (Kersten et al., 1996; Kersten et al., 1997). Thus, not only do human observers derive shape information from shading gradients and surface orientation from specularities, but we also draw on shadows to provide constraints on the otherwise ambiguous three-dimensional layout of a scene. However, there is some cost to relying on such information – recognition performance, which without the presence of such information might be lighting invariant, becomes lighting sensitive. That is, the object representations we remember and use for recognition include information about the particular lighting conditions under which objects were actually seen. Therefore, changing the lighting from a familiar to an unfamiliar configuration will negatively affect recognition. As mentioned, we have observed this effect for both novel and familiar objects. Examination of the *pattern* of this illumination sensitivity is the first step towards understanding the specific algorithms being used by the human visual system to compensate for variations in lighting.

The results of the present study provide one of the first direct comparisons between the performance of human observers and a functional computer vision recognition system. Although neither the behavioral task used here, the recognition of static views of faces, nor the implemented algorithm used for recognition, the Illumination Cone (IC) model, address the question of how generic object recognition is achieved, both the task and the model capture critical aspects of the recognition process. Specifically, we were interested in how biological and machine vision systems compensate for the dramatic changes in the appearance of objects that occur with variable lighting. Human faces were used as the stimulus domain because they offer a paradigmatic recognition problem that is both complex and of great interest. Building on recent work in both research communities we tested the generalization performance of human observers and the IC model under similar training conditions. In each of five experiments observers and the model learned the identity of ten faces under a small subset of lighting directions and were then tested with the same faces appearing under new lighting directions. The ability to generalize from familiar to unfamiliar illumination conditions was then compared across human subjects and the IC model.

Critically, we manipulated the nature of the training images in each experiment. Experiment 1 used a set of near-frontal lighting directions, Experiment 2 used the mirror of this set, creating extreme illumination conditions, Experiment 3 used a single frontal lighting direction, Experiment 4 used the mirror of this, a single extreme direction, and, finally, Experiment 5 used a set that spanned the horizontal meridian of the lighting sphere. Across these different training conditions we obtained the following results:

- Although the IC model exhibited higher accuracy for the exact images shown in training, it often performed worse than humans for the same faces under new lighting directions.
- Humans were much better at generalizing from extreme lighting directions than was the IC model. On the other hand, recognition performance for subjects and the model was similar when generalizing from near-frontal lighting directions.
- Humans were able to perform at a more constant level with new illuminations distant from the training set when the training set was comprised of extreme lighting directions. In contrast, when the training set was comprised of near-frontal directions, generalization fell off rapidly with distance from training images.
- When the training set was comprised of lighting directions along the horizontal meridian, humans were far better than the IC model at generalizing to test images arrayed vertically around this horizontal axis.

We wish to reiterate that some of the above differences are inherent in the comparison we are making between the full vision system of humans and extremely limited vision system implemented in the IC model. Moreover, although humans must recognize faces in the context of their familiarity with 1000's of similar objects (in particular other faces), they may also use their knowledge of the general geometry of faces as a class to make inferences regarding the appearance of new faces under novel lighting directions (for a similar class-level mechanism for making inferences about novel viewpoints, see Tarr & Gauthier, 1998). These factors lead us to expect human observers to display both better generalization across all unfamiliar illumination conditions and dramatically better generalization for lighting directions far from the training set as compared to the IC model. At the same time, the fact that the IC model has few competitors for an individual face under the trained illumination conditions, while humans have 1000's, leads the model to perform better than humans for the exact images used in training.

Even given these differences, there is remarkable similarity in the performance of our human subjects and the IC model. It is worth remembering that even the fact that humans show any systematic lack of lighting invariance is somewhat contrary to "standard" thinking in the psychophysical literature. To date, all studies of illumination dependence in human object recognition have only compared changed to unchanged lighting in a qualitative manner – never examining how recognition performance varies as a function of distance from known illumination conditions. Under these circumstances it is difficult to infer much about the computational algorithms used to compensate for lighting variability, even more so because most qualitative comparisons have revealed only small effects of changing lighting direction (Braje et al., 1998; Tarr et al., 1998). Here we extend these findings in a more systematic fashion, exploring not only how performance varies as the lighting direction is moved further and further from the original training conditions, but how well human vision generalizes across both standard and unusual lighting conditions, for instance, when most of the face is in shadow due to extreme lighting directions. A second important feature of the present study is the execution of analogous experiments in both humans and machine vision systems. Specifically, we employed a computational model specifically designed to account for lighting variability in scenes. The performance of this model in each experiment was then directly compared to the generalization pattern obtained from human observers. These comparisons are summarized above, but overall it is clear that both humans and the IC model show a similar sensitivity to lighting direction, although specific effects are mitigated somewhat by the highly-familiar nature of faces for human subjects.

Such results indicate that one important future study involves extending the present methods to entirely novel object classes for which neither humans nor any computational model would have pre-existing knowledge. A second important direction is to compare human performance to recognition systems that address the question of lighting variability using different algorithms from the IC model. For example, the approaches implemented in both the Lades et al. (1993) and Atick, Griffin, & Redlich (1996) systems should be considered among others. In terms of the conditions under which these and other models are compared to human observers, there are also more complex lighting manipulations that might be implemented. One of the most important is the inclusion of multiple simultaneous light sources for each image. Such complexity may make images more difficult to interpret, but also provide additional constraints on the extraction of a lighting model for the scene, as well as the structure of the object. Finally, we should consider that our human subjects benefited from their prior experience with faces. Currently the IC model (as well as most other recognition models that address lighting variability) does not represent information about an object class – rather a separate and unique illumination cone is constructed for each individual face. However, it is apparent that class-level knowledge about how illumination generically affects the appearance of members of the class is a desirable feature to incorporate into future models. More generally, this last point illustrates that a consideration of human visual abilities in the context of models drawn from computer vision is not a one-way street. Both approaches benefit from the comparison and ultimately more robust machine vision systems and better accounts of biological vision will result.

References

Atick, J. J., Griffin, P. A., & Redlich, A. N. (1996). Statistical approach to shape from shading: reconstruction of three-dimensional face surfaces from single two-dimensional images. *Neural Computation, 8*(6), 1321-1340.

Belhumeur, P. N., & Kriegman, D. J. (1998). What is the set of images of an object under all possible illumination conditions? *International Journal of Computer Vision, 28*(3), 245-260.

Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class-specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(7), 711-720.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115-147.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance, 19*(6), 1162-1182.

Binford, T. O. (1971, December). *Visual perception by computer.* Paper presented at the IEEE Conference on Systems and Control, Miami, FL.

Braje, W. L., Kersten, D., Tarr, M. J., & Troje, N. F. (1998). Illumination effects in face recognition. *Psychobiology, 26*(4), 371-380.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc. Natl. Acad. Sci. USA, 89*, 60-64.

Fukushima, K. (2000). Active and adaptive vision: Neural network models. In S.-W. Lee, H. H. Bülthoff & T. Poggio (Eds.), *Biologically Motivated Computer Vision* (Vol. 1811, pp. 623-634). Berlin: Springer-Verlag.

Georghiades, A., Kriegman, D., & Belhumeur, P. (1998). Illumination cones for recognition under variable lighting: Faces. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 52-59.

Georghiades, A. S., Belhumeur, P. N., & Kriegman, D. J. (2000). From few to many: Generative models for recognition under variable pose and illumination. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition,* 277-284.

Hallinan, P. (1994). A low-dimensional representation of human faces for arbitrary lighting conditions. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 995-999.

Hallinan, P. L., Gordon, G. G., Yuille, A. L., Giblin, P., & Mumford, D. (1999). Two- and Three-Dimensional Patterns of the Face. Natick, MA: A K Peters.

Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 23*(5), 1511-1521.

Horn, B. K. P. (1975). *Obtaining Shape from Shading Information*. New York: McGraw-Hill.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*(3), 480-517.

Johnston, A., Hill, H., & Carman, N. (1992). Recognising faces: Effects of lighting direction, inversion, and brightness reversal. *Perception, 21*, 365-375.

Kersten, D., Knill, D., Mamassian, P., & Bülthoff, I. (1996). Illusory motion from shadows. *Nature, 379*, 31.

Kersten, D., Mamassian, P., & Knill, D. C. (1997). Moving cast shadows induce apparent motion in depth. *Perception, 26*(2), 171-192.

Lades, M., Vorbruggen, 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.

Logothetis, N. K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representation in the primate. *Cerebral Cortex, 3*, 270-288.

Lowe, D. G. (2000). Towards a computational model for object recognition in IT Cortex. In S.-W. Lee, H. H. Bülthoff & T. Poggio (Eds.), *Biologically Motivated Computer Vision* (Vol. 1811, pp. 20-31). Berlin: Springer-Verlag.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. of Lond. B, 200*, 269-294.

Moore, C., & Cavanagh, P. (1998). Recovery of 3D volume from 2-tone images of novel objects. *Cognition, 67*(1-2), 45-71.

Moses, Y., Adini, Y., & Ullman, S. (1994). Face recognition: The problem of compensating for changes in illumination direction. In J.-O. Eklundh (Ed.), *Lecture Notes in Computer Science, Vol. 800: Computer Vision – ECCV '94* (pp. 286-296). Berlin: Springer-Verlag.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343*, 263-266.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature, 331*(14), 163-166.

Riesenhuber, M., & Poggio, T. (2000). CBF: A new framework for object categorization in cortex. In S.-W. Lee, H. H. Bülthoff & T. Poggio (Eds.), *Biologically Motivated Computer Vision* (Vol. 1811, pp. 1-9). Berlin: Springer-Verlag.

Sanocki, T., Bowyer, K. W., Heath, M. D., & Sarkar, S. (1998). Are edges sufficient for object recognition? *Journal of Experimental Psychology: Human Perception and Performance, 24*(1), 340-349.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review, 2*(1), 55-82.

Tarr, M. J., Bülthoff, H. H., Zabinski, M., & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science, 8*(4), 282-289.

Tarr, M. J., & Gauthier, I. (1998). Do viewpoint-dependent mechanisms generalize across members of a class? *Cognition, 67*(1-2), 71-108.

Tarr, M. J., Kersten, D., & Bülthoff, H. H. (1998). Why the visual system might encode the effects of illumination. *Vision Research, 38*(15/16), 2259-2275.

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint-dependent. *Nature Neuroscience, 1*(4), 275-277.

Turk, M., & Pentland, A. (1991a). Eigenfaces for recognition. *Journal of Cognitive Neuroscience, 3*(1), 71-86.

Turk, M., and Pentland, A. (1991b). Face recognition using eigenfaces. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 586-591.

Ullman, S., & Sali, E. (2000). Object classification using a fragment-based representation. In S.-W. Lee, H. H. Bülthoff & T. Poggio (Eds.), *Biologically Motivated Computer Vision* (Vol. 1811, pp. 73-87). Berlin: Springer-Verlag.

Williams, P. and Tarr, M. J. *RSVP: Experimental control software for MacOS* [Online]. Available: http://psych.umb.edu/rsvp/ [2001, February 8].