

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600**

UMI[®]

**The Role of Illumination in Face Recognition: Evidence for Illumination
Dependent Representations Using Psychophysics and Computation**

by

Cullen Davis Jackson

B.A., Trinity University, 1996

B.S., Trinity University, 1996

Sc.M., Brown University, 1998

Thesis

**Submitted in partial fulfillment of the requirements for the Degree of Doctor
of Philosophy in the Department of
Psychology at Brown University**

PROVIDENCE, RHODE ISLAND

May 2002

UMI Number: 3050907

UMI[®]

UMI Microform 3050907

Copyright 2002 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

© Copyright 2002 by Cullen Davis Jackson

**This dissertation by Cullen Davis Jackson
Is accepted in its present form by the Department of
Psychology as satisfying the
dissertation requirements for the degree of Doctor of Philosophy**

Date 4-26-02



Dr. Michael J. Tarr, Director


Recommended to the Graduate Council

Date 4/17/02



Dr. Leslie Welch, Reader

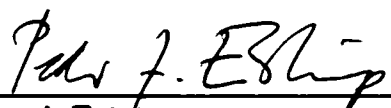
Date 4/17/02



Dr. Michael Black, Reader

Approved by the Graduate Council

Date 5/2/02



Dr. Peder J. Estrup
Dean of the Graduate School & Research

CURRICULUM VITAE

Cullen Davis Jackson

Education

Ph.D. May 2002	Experimental Psychology	Brown University
Sc.M. May 1998	Experimental Psychology	Brown University
B.A. May 1996	Psychology	Trinity University
B.S. May 1996	Computer Science	Trinity University

Professional Experience

Research:

1999 - 2001	<i>Research Trainee</i> NSF-IGERT Training Program	Brown University Department of Cognitive & Linguistic Sciences Providence, RI 02912
1998 - 1999	<i>Research Fellow</i> Brain & Behavior Fellowship	Brown University Departments of Psychology and Cognitive & Linguistic Sciences Providence, RI 02912

Teaching:

1998 & 2000	<i>Guest Lecturer</i> Perception and Mind	Brown University Prof. Michael Tarr
2000	<i>Guest Lecturer</i> Perception	Brown University Prof. Leslie Welch
1996 - 1998	<i>Teaching Assistant:</i> Introduction to Psychology Personality Sensory Processing Lab Quantitative Methods	Brown University Department of Psychology Providence, RI 02912

Awards and Honors

1997	ARVO/National Eye Institute Travel Grant	
1996	Associate Member	Sigma Xi: The Scientific Research Society

Scholarships and Fellowships

1999-2001	NSF-IGERT Training Fellowship	Brown University
1999	Summer Institute for Cognitive Neuroscience Fellowship	Dartmouth College
1998 - 1999	Brain & Behavior Fellowship	Brown University
1998	Summer Research Fellowship	Brown University
1992 - 1993	Jesse H. Jones Scholarship	Trinity University
1991 - 1996	President's Scholarship	Trinity University

Memberships

Association for Computing Machinery, Full Member
Sigma Xi: The Scientific Research Society, Associate Member-at-Large

Publications and Manuscripts

Jackson, C. D., Tarr, M. J., and Georghiades, A. (in revision). Identifying faces across variations in lighting: Psychophysics and computation. *International Journal of Computer Vision, Special Issue on Vision at Brown University.*

Jackson, C. D., and Welch, L. (in revision). How many temporal frequency filters contribute to speed perception? *Vision Research.*

Jackson, C. D., Laidlaw, D. H., and Prabhat. (in preparation). Examining hypercubes in a CAVE versus in a desktop VR system. *IEEE Computer Graphics and Applications.*

Jackson, C. D., Laidlaw, D. H., Karelitz, D. B., and Grossberg, D. (in preparation). Context matters: Evidence for the use of background texture in identifying shapes in immersive virtual reality. *IEEE Transactions on Visualization and Computer Graphics*.

Jackson, C. D., Tarr, M. J., Kersten, D. (in preparation). The impact of illumination on face recognition: Evidence for illumination dependent representations.

Jackson, C. D. (2002). *The role of illumination in face recognition: Evidence for illumination dependent representations using psychophysics and computation*. Unpublished Ph.D. dissertation, Department of Psychology, Brown University, Providence, RI.

Jackson, C. D. (1998). *How many temporal frequency channels contribute to speed perception?* Unpublished master's thesis, Department of Psychology, Brown University, Providence, RI.

Jackson, C. D. (1996). *RiGID: Rorschach Graphical Interface Development*. Unpublished bachelor's thesis, Department of Computer Science, Trinity University, San Antonio, TX.

Presentations and Abstracts

Jackson, C. D. Identifying faces across variations in lighting: Psychophysics and computation. Colloquium talk presented at the Smith-Kettlewell Eye Research Institute, San Francisco, CA, June 2001.

Jackson, C. D. Illumination and object recognition revisited: Out of the shadows come the answers. Talk presented at the Spring retreat for the NSF-IGERT program: "Learning and Action in the Face of Uncertainty: Cognitive, Computational, and Statistical Approaches." South Kingston, RI, May 2001.

Jackson, C. D., and Welch, L. The effects of double masking on speed perception. Paper presented at the annual meeting of the Optical Society of America, Providence, RI, October 2000.

Jackson, C. D., and Welch, L. The effects of masking duration on speed perception. Poster presented at the annual meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, FL, May 2000.

Jackson, C. D., and Welch, L. (2000). The effects of masking duration on speed perception. *Investigative Ophthalmology and Visual Science*, 41(4), S4210.

- Jackson, C. D., and Tarr, M. J. How does the brain recognize objects under extreme variations of illumination? Poster presented at the annual meeting of the Cognitive Neuroscience Society, San Francisco, CA, April 2000.
- Jackson, C. D. The role of illumination in object perception: Psychophysics and computation. Talk presented at the Spring retreat for the NSF-IGERT program: "Learning and Action in the Face of Uncertainty: Cognitive, Computational, and Statistical Approaches." Little Compton, RI, March 2000.
- Jackson, C. D., and Tarr, M. J. The impact of illumination on the recognition of faces I: Evidence for illumination dependent representations. Paper presented at the annual meeting of the Eastern Psychological Association, Providence, RI, April 1999.
- Tarr, M. J., and Jackson, C. D. The impact of illumination on the recognition of faces II: Modeling illumination dependency. Paper presented at the annual meeting of the Eastern Psychological Association, Providence, RI, April 1999.
- Jackson, C. D., and Welch, L. Temporal frequency channels and speed perception. Paper presented at the annual meeting of the Optical Society of America, Baltimore, MD, October 1998.
- Jackson, C. D., and Welch, L. (1998). Temporal frequency channels and speed perception. *Special Issue to Optics and Photonics News*, 9 (8), p. 111.
- Jackson, C. D., and Welch, L. How many temporal frequency filters for speed perception? Poster presented at the annual meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, FL, May 1997.
- Jackson, C. D., and Welch, L. (1997). How many temporal frequency filters for speed perception? *Investigative Ophthalmology and Visual Science (Suppl.)*, 38, S377.
- Jackson, C. D. RiGID: Rorschach Graphical Interface Development. Paper presented at the annual meeting of the National Conference for Undergraduate Research, Asheville, NC, April 1996.
- Jackson, C. D., and Craton, L. Shape from shadows. Poster presented at the annual meeting of the Southwestern Psychological Association, San Antonio, TX, April 1995.

Service

1998 – 2001	Student Representative, Graduate Council	The Graduate School Brown University
1997 – 2000	Representative, Graduate Student Council	Department of Psychology Brown University

Personal

Address: Brown University
Department of Psychology
89 Waterman Street
Providence, RI 02912-1853

Citizenship: United States of America

ACKNOWLEDGMENTS

Completing my dissertation has been like a personal “trial by fire.” I know that many Ph.D. candidates complete their theses every year, but for each one of us, it *is* personal. The past few years have culminated in this one work, and at this point in time, it seems a bit pointless. However, I persevere due to the tireless efforts of others who assure me that the end is in sight, and that I have the will and the ability to see this thing to the end. I’d like to thank those few people who have helped me along this tortuous path, either with words of encouragement, by leading the way themselves, or by giving intellectual and scientific assistance.

First, I’d like to thank my advisors and readers:

Michael Tarr, for teaching me how to think about a problem and how to formulate a strategy to study it; for encouraging me to pursue an interdisciplinary problem; for pointing me at excellent funding sources and helping me to attain them; for finding a postdoctoral appointment for me that would allow me to stay at Brown for a little while longer; and for never letting me quit.

Leslie Welch, for convincing me to come to Brown and accepting me into her lab as an untested first-year student; for being a scientific mentor and helping me to present my early work at a major conference only 8 months after starting my graduate career; as an advisor for my Master’s thesis; for being a good friend who was always there with encouragement and the occasional “barking;” for teaching me the ins and outs of visual psychophysics; and for reading this thing and telling me that the first draft was “pretty good.”

Michael Black, for listening to my ramblings within his first few weeks at Brown, for agreeing to serve as a reader, and for not making too many comments that required strenuous revision.

Second, I'd like to thank my fellow graduate students (and a faculty member):

Jason Machan and Galit Naor-Raz, for sticking it through and not letting our class's 73% attrition rate deter us; for always having words of encouragement when I needed them; for not letting my head get too big; and for continuing to be my good friends even though we're going our separate ways.

Jason Taylor, Elena Festa-Martino, and Bill Heindel, for letting me take up space in their lab this past year; for listening to me ramble about my research and complain sometimes; for going with me to get tremendous amounts of coffee at all times of the day; for numerous lunches and dinners; and for their continuing friendship and honesty.

Vlada Aginsky, for sharing an office with me and listening to my ramblings, "blah, blah, blah!"; for taking a chance and attempting to complete the Brown Entrepreneurship Program under unknown conditions with me; for her friendship; for her honesty; for trusting me; for leading the way and showing me how to get to the end of the road with a modicum of grace and dignity; and for never letting me quit.

Third, I'd like to thank those who helped with various scientific issues:

Athos Georghiades, for providing a Matlab implementation of the IC model and answering numerous questions about it, and for reading and commenting on parts of this work.

James Coughlan and Alan Yuille, for providing wonderful hospitality during my stay in San Francisco, and for giving insight into how we might improve on the IC model.

Daniel Kersten, for inspiring me to work on the questions of how the visual system deals with illumination in object recognition, and for his collaborative efforts in trying to answer some of these questions.

Fourth, I'd like to thank various members of the Brown community:

Michelle Ross and Patricia Devine, goddesses of the Psychology Department, for helping me keep up with what I needed to do to finish, and for being there when I needed a smiling face or a kick in the pants.

Bob Fifer and Bob Moore, for keeping the computers running and the software current.

Joan Lusk, for keeping the university bureaucracy at bay, for information concerning various and sundry rules, and for working out my stipend problems.

Finally, and foremost in my mind, I'd like to thank my family:

Robert and Elaine Jackson, my parents, for teaching me values; for always encouraging me to do my best and not give up; for believing in me; for letting me know that I can do whatever I put my mind to; for giving me the resources to achieve; for being the people they are; and for their unwavering love.

Richard and Melanie Moreno, my in-laws, for raising my most valuable resource; for sharing their families with me and allowing me to become part of theirs; for helping us when we need it; and for their love.

***Robert Jackson and Maggie McMahon*, my brother and sister-in-law, for not letting me get away with anything; for their love and support; and for helping me to keep my perspective and see the “big picture.”**

***Ophelia and Sebastian*, the “babies,” for not letting me take myself too seriously; for waking me up when they knew I needed to get up, and letting me sleep the rest of the time; and for their unconditional love.**

***Victoria*, my wife, for agreeing to share her life with me; for never letting me give up; for her loving encouragement; for following me up to Rhode Island; for sharing her family with me; for her tireless devotion; for listening to me talk about work; for reading stuff that bored her to tears; for sticking with me; for her laughter and her smile; for her hugs and kisses; for continuing to love me; for building a family with me; and for the rest of my life.**

This dissertation is dedicated to my family members who are no longer with us: Archie Moran, Pat Moreno, and Horatio. I love you and miss you!

TABLE OF CONTENTS

Signature Page	iii
Curriculum Vita	iv
Acknowledgments	ix
List of Tables	xv
List of Illustrations	xvi
Chapter 1: <i>Introduction</i>	1
Chapter 2: <i>Identifying Faces Across Variations in Lighting: Psychophysics and Computation</i>	
Introduction	24
General Methods	26
Results.....	36
Discussion	50
Chapter 3: <i>Lighting and Face Recognition: Evidence for Illumination Dependent Mechanisms</i>	
Introduction	57
General Methods	61
Experiment 1	
Introduction and Methods.....	66
Results and Discussion	68
Experiment 2	
Introduction and Methods.....	77

Results and Discussion.....	79
Experiment 3	
Introduction and Methods	85
Results and Discussion	87
Discussion	97
Chapter 4: <i>General Discussion</i>	101
References.....	109

LIST OF TABLES

Table 1: <i>Correlations (Pearson's r) between the IC model and psychophysically assessed human subject performance for Experiments 1 through 5.....</i>	49
Table 2: <i>The distribution of observers across the three experiments described in Chapter 3.</i>	62

LIST OF ILLUSTRATIONS

- 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.*..... 27
- 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).*..... 29
- Figure 3: *Schematic of the training phase for the human psychophysical experiments. The fixation cross was seen for 500 msec, followed by the training stimulus seen for 1000 msec, followed by a blank interval in which the subjects were given as much time as needed to make a response on the keyboard, followed by a 1000 msec inter-trial interval.* 30
- Figure 4: *Schematic of the testing phase for the human psychophysical experiments. The fixation cross was seen for 500 msec, followed by the testing stimulus seen for 500 msec, followed by a blank interval in which the subjects were given 3000 msec to make a response on the keyboard, followed by a 1000 msec inter-trial interval.* 31
- Figure 5: *Training sets for the human psychophysical and IC 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).*..... 32
- Figure 6: *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).* 35
- Figure 7: *Sampling of test images for Experiment 1, and their respective distances from the closest illumination direction 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°).* 37

Figure 8: Results for both human subjects and the Illumination Cone (IC) model for Experiment 1 (training with near-frontal lighting directions). The trained illumination directions were within 15° of the camera axis (0°, 0°). The circles represent mean percent correct for the human subjects. The error bars are the within-subject standard error of the mean. A single case of the IC model is represented by the squares..... 39

Figure 9: Results for both human subjects and the Illumination Cone (IC) model for Experiment 2 (training with extreme lighting directions). The trained illumination directions were within 15° of (75°, 0°). The circles represent mean percent correct for the human subjects. The error bars are the within-subject standard error of the mean. A single case of the Illumination Cone (IC) model is represented by the squares..... 40

Figure 10: Results for both human subjects and the Illumination Cone (IC) model for Experiment 3 (training with a single frontal lighting direction). The trained illumination direction was on the camera axis, (0°, 0°). The circles represent mean percent correct for the human subjects. The error bars are the within-subject standard error of the mean. A single case of the Illumination Cone (IC) model 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. 41

Figure 11: Results for both human subjects and the Illumination Cone (IC) model for Experiment 4 (training with a single extreme lighting direction). The trained illumination direction was (75°, 0°). The circles represent mean percent correct for the human subjects. The error bars are the within-subject standard error of the mean. A single case of the Illumination Cone (IC) model is represented by the squares. 42

Figure 12: Results for both human subjects and the Illumination Cone (IC) model for Experiment 5 (training with lighting directions along the horizontal axis of the lighting space). The trained illumination directions were along the horizontal meridian between (0°, 0°) and (75°, 0°). The circles represent mean percent correct for the human subjects. The error bars are the within-subject standard error of the mean. A single case of the Illumination Cone (IC) model is represented by the squares..... 43

Figure 13: Example of the training sequence in Experiment 1. Each of the two training images was displayed for 15 seconds separated by a 1 second blank. The order of presentation of the two images was randomized. The numbers in parentheses represent the horizontal and vertical displacement of lighting on the face from the center of the image..... 64

Figure 14: The test trials started with a fixation cross in the center of the screen displayed for 500 msec. The centered test image then was displayed for 150 msec. The test image shown has lighting rendered at -20° on the horizontal and 10° on the vertical axis. A mask was displayed after the test image for 500 msec. The same mask was shown across all trials and all observers. A 1 second blank screen followed each test trial. 65

Figure 15: A schematic of the lighting sphere used in Experiment 1 showing the relative positions of the testing conditions to the training illuminations. The dots indicate the trained lighting directions. The conditions in which the test illumination directions were grouped are shown: 1) lighting between the two training conditions (INTER); 2) lighting outside of the trained lighting directions (EXTRA); 3) lighting directions orthogonal to the axis defined by the training illuminations (ORTHO). 68

Figure 16: The effect of illumination condition on recognition performance in Experiment 1. The INTER condition includes lighting directions between the training illuminations. Lighting directions outside of the two training directions are in the EXTRA condition. The ORTHO condition contains lighting directions orthogonal to the axis defined by the training illuminations, which in this experiment are oriented vertically on the lighting sphere. Error bars are the within-subject standard error of the mean. 70

Figure 17: Recognition performance as a function of the lighting directions in the ORTHO condition for Experiment 1. The first number in the parentheses corresponds to horizontal position on the lighting sphere while the second number denotes vertical position (in degrees). The lighting coordinates to the right of the graph lie above the (10° , 10°) training illumination, while those coordinates to the left lie below the training illumination. Recognition performance is better when observers viewed images with lighting below the face than when viewing faces with lighting positioned above. 71

Figure 18: Recognition performance as a function of the tested illumination directions. The illumination directions on the x-axis are labeled as horizontal and vertical coordinates (in degrees) on the lighting sphere. The bolded labels (and lines) mark the INTER and EXTRA conditions as shown. The two trained lighting directions are labeled with "TRAIN." 72

Figure 19: Recognition performance as lighting deviated from either training point in Experiment 1. The distances are averaged over all three of the lighting conditions (INTER, EXTRA, and ORTHO). 73

Figure 20: Recognition performance as lighting deviated from the near-frontal (10° , 10°) training point in Experiment 1. The distances are grouped over all three of the lighting conditions (INTER, EXTRA, ORTHO). 75

Figure 21: The effect of illumination condition on recognition performance in Experiment 2. As observed in Experiment 1, the EXTRA condition is significantly different than the other two illumination conditions ($F(2, 46) = 3.88, p < 0.05, \epsilon = 0.74$). Error bars represent the within-subject standard error of the mean. 81

Figure 22: Recognition performance as a function of the lighting directions in the ORTHO condition in Experiment 2. Lighting coordinates to the right of the graph represent lighting directions that light the brow of the face while those to the left represent lighting directions that light the chin of the face. The error bars represent the within-subject standard error of the mean. 83

Figure 23: Recognition performance as lighting deviated from the near-frontal ($10^\circ, 10^\circ$) training point in Experiment 2. The distances are grouped over all three of the lighting conditions..... 85

Figure 24: The lighting sphere used in Experiment 3 showing the relative positions of the testing conditions with respect to the trained illumination directions. The dots indicate the trained lighting directions. The conditions in which the test illumination directions were grouped are shown: 1) lighting between the two training conditions (INTER); 2) lighting outside of the trained lighting directions (EXTRA); 3) lighting directions orthogonal to the axis defined by the training illuminations (ORTHO)..... 87

Figure 25: The effect of illumination condition on recognition performance in Experiment 3. Recognition performance in the INTER condition was significantly different from performance in the other two illumination conditions. Error bars represent the within-subject standard error of the mean. 89

Figure 26: Recognition performance for the lighting directions in the EXTRA condition in Experiment 3. The distances were measured from the near-frontal ($10^\circ, -10^\circ$) training point. The images represented by the 10° and 20° distances contained faces with chin-lit illumination and the images represented by the 50° and 60° distances contained faces with brow-lit lighting. 93

Figure 27: Recognition performance for the lighting directions in the INTER and ORTHO conditions in Experiment 3. The distances are measured from the near-frontal ($10^\circ, -10^\circ$) training point. The error bars represent the within-subject standard error of the mean..... 94

Figure 28: Recognition performance for the lighting directions in the INTER and ORTHO conditions in Experiment 3. The distances are measured from the near-frontal ($10^\circ, -10^\circ$) training point. The error bars represent the within-subject standard error of the mean..... 95

Figure 29: Recognition performance as a function of distance from either of the trained lighting directions in Experiment 3. A significant linear trend exists between the distances ($F = 5.231$, $p < 0.05$). The error bars represent the within-subject standard error of the mean. 96

CHAPTER 1

Introduction

“Efficient object recognition requires a mechanism whereby a set of two or more stimulus inputs are allocated to the same perceptual category. For example, we have the capacity to identify an object from an infinite variety of orientations, distances and luminances” (Warrington, 1982, p. 22).

Humans routinely exhibit real-time, highly accurate three-dimensional object recognition under widely varying illumination conditions. However, little is known of *how* humans are able to perform so well under so much variation in lighting. At the same time, recent progress in computer vision has produced recognition models that are quite good at compensating for lighting variations during object recognition (Hallinan, 1994; Belhumeur, Hespanha, & Kriegman, 1997; Belhumeur and Kriegman, 1998). These models use specific algorithms that compensate for changes in illumination across objects, and perform quite well at recognizing objects under lighting directions similar to previously learned

illumination directions. While the algorithms used by these models are well understood, their performance when trained on a variety of different illuminations is unknown. One method of investigating the ability of humans to perform accurate object recognition under changes in illumination is to examine the object recognition performance of both humans and computational models under a variety of lighting conditions. The goal of such an undertaking is to determine if illumination really plays a role in human object recognition, and if so, to determine how it is represented.

Evidence from previous behavioral research suggests that human object recognition is dependent on the lighting conditions in a scene, much like the dependence shown for recognition under varying object viewpoints. Neuropsychological and neurophysiological studies indicate neural systems that may be involved in object recognition under novel and ambiguous lighting conditions. Results from these various sources indicate that a certain type of model for object recognition is more likely than others: an image-based model. Several image-based computer vision models suggest a mechanism by which object recognition under illumination variations may be performed. These models show simulated behavior that appears lighting dependent. The present study compares performance data from human psychophysical studies and simulations using an image-based computer vision model under the same lighting conditions to answer the question of the function of lighting in human object recognition, as well as to explore the larger issues that arise when attempting such a comparison.

What is the Role of Illumination in Human Object Recognition?

Invariant three-dimensional object recognition has been an elusive goal in computer vision, and an ill-understood phenomenon in human vision. Given the significant attention paid to this problem, the inability to find a generic solution is remarkable. 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.

Changes in object orientation, or viewpoint, and changes in illumination on objects are two of the most obvious sources of variability present in images of objects. In fact, the visual system can recognize hundreds-of-thousands of unique objects at a large number of viewpoints. This remarkable ability to easily recognize objects regardless of orientation has been a subject of investigation for many years. Studies using humans (e.g., Rock & DiVita, 1987) and non-human primates (e.g., Logothetis & Sheinberg, 1996) have attempted to answer how a biological vision system compensates for viewpoint changes. Related to this question is whether the internal representation of objects is independent or dependent on the viewpoint of the observer to the objects. Two classes of models have arisen in order to explain how humans might represent objects internally: viewer-centered (e.g., Poggio & Edelman, 1990) and object-centered (e.g., Biederman, 1987; Ullman, 1989).

With respect to the problem of recognizing objects across multiple viewpoints, whether through object 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 hardly typical and only obtained by carefully following a strict “recipe” (Tarr, Williams, Hayward, & Gauthier, 1998). 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 was found to be *viewpoint dependent*. That is, observers took progressively longer and were 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 were also recognized in a

viewpoint-dependent manner. Moreover, adding more Geons, something that theoretically should have made the objects more distinctive from one another (Hummel & Biederman, 1992), actually dramatically increased the magnitude of viewpoint dependency (Tarr, Bülthoff, Zabinski, & Blanz, 1997).

Similar to changing the orientation of an object with respect to the viewer, the direction of illumination on an object can impact how well the object is recognized. Although Marr (1982) suggested that changes in illumination direction are only seen as changes in light source direction and not as an orientation change, Troje and Siebeck (1998) demonstrated that changes in illumination direction *can* provide cues that appear as changes in object orientation, which greatly affect an object's appearance and thus the ability to recognize it. Furthermore, when measured in terms of pixels, illumination changes account for the greatest image variance in measuring the differences between images of faces. In fact, illumination changes account for even more image variance than the variance described by individual identity *or* changes in viewpoint between images (Pentland, 1991; Moses, Adini, and Ullman, 1994). In order to accurately recognize an object under varying illumination conditions, both humans and computer vision models must compensate for the large variation caused by these lighting changes across images of the object. As with viewpoint invariance, there also are two types of models proposed to account for the ability to recognize objects under variable illumination conditions: edge-based (e.g., Biederman, 1987; Hummel & Biederman, 1992) and image-based (e.g., Hallinan, 1994; Shashua, 1997; Belhumeur & Kriegman, 1998).

Edge-based Models. The key idea behind these models is found in the hypothesis that object representations are edge-based (e.g., Marr, 1982; Biederman, 1987). The belief is that the edges of an object are stable across variations in the image, and that from such canonical edge descriptions completely invariant, three-dimensional object models can be derived. These edge-based models discount the surface characteristics (i.e., color, texture, illumination) of an object as secondary to recognition since the information about the object's shape is completely contained in the presumably stable edge maps. In support of the edge-based hypothesis, Biederman and Ju (1988) found that subjects showed no difference in performance for naming objects displayed both as line drawings and as color photographs. The authors believed that the color and texture elements of the objects shown in the photographs, but necessarily missing from the line drawings, did not contribute to the subjects' performance since no representation of these characteristics was ever accessed. Following up this work, Hummel and Biederman (1992) posited a neural network model for invariant object recognition under changes in viewpoint that used a structural description that was temporally bound, i.e., the model joined component parts when they were needed. The structural description in this model was edge-based and the authors again discounted the effects of illumination on the surface of the object, assuming that the surface properties of the object in question were easily described by its line drawing. In summary, edge-based models predict that changes in lighting over an object

will not differentially affect object recognition performance, i.e., recognition will be invariant.

Image-based Models. 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 an object. Consequently, it would be less than optimal to deal with lighting variation by completely discarding it, thereby reducing the probability of correctly recognizing an object (e.g., a face) under ambiguous lighting conditions, as is the case with edge-based models.

Many recent recognition algorithms intended to model biological vision rely on image-based views rather than edge-based (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 were displayed 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. Because the representations used in image-based models preserve the appearance of an object in the image, e.g., encoding the lighting information in the scene, the predicted performance of these models for recognition under varying lighting conditions should not be invariant per se. Instead, recognition performance should differ as the lighting conditions in the scene change from known to unknown conditions.

Evidence for the Role of Illumination in Human Object Recognition

One strong critique of the edge-based approach is that edge maps are rarely stable over even relatively small changes in the image. Unfortunately, changes in illumination across an object can create relatively large changes in an object's image. Rather than being stable over changes in lighting, the edge descriptions 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 (Bülthoff and Edelman, 1992; Sanocki, Bowyer, Heath, & Sarkar, 1998). Furthermore, without an explicit model of the lighting parameters for a given scene, it is difficult, if not impossible, to discount edges that arise from shadows as opposed to object contours. So, the edges seen in the image of an object would be suspect as to whether they were actually part of the object or an effect of the lighting in the scene, unless the three-dimensional scene parameters (including lighting) were known. This ambiguity could result in the edge descriptions for an object with two different lighting directions being drastically different from one another. Given that edge-based models predict lighting invariance and that image-based models predict some illumination dependence, is there a preponderance of evidence either way with respect to biological visual systems?

Behavioral Evidence. As mentioned before, the study of visual object recognition in humans and other primates has focused largely on the problem of recognition across changes in viewpoint (e.g., Rolls, Baylis, Hasselmo, & Nalwa, 1989; Bülthoff & Edelman, 1992; 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.

There are several reasons why the study of 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. This accessibility bottleneck made using high-end computer graphics to carefully answer questions about lighting and human object recognition essentially impossible for 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, there has been awareness that lighting affects the shading gradients on an object's surfaces, and that such shading information can be used to infer three-dimensional shape (e.g., Horn, 1975; Ramachandran, 1988). However, it has been less obvious that the effects of a particular illumination context might affect an object's *representation*. That is, although lighting clearly influences 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.

Given the non-viability of edge-based models (e.g., Sanocki, Bowyer, Heath, & Sarkar, 1998), as well as other lighting-invariant representational schemes, as stable representations for object recognition, 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, the finding that cast shadows helped constrain the perceived three-dimensional layout of a scene (Kersten, Knill, Mamassian, & Bülthoff, 1996; Kersten, Mamassian, & Knill, 1997) motivated the authors. 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 seemed to be included in object representations for one very good reason – although they produced some lighting dependence in recognition, this dependence was outweighed by the fact that the shadows helped to disambiguate the three-dimensional appearance of the 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. However, the key point was that the *pattern* in performance of the lighting dependence gave information regarding the mechanisms underlying the practical human ability to attain near-invariant recognition.

It is worth noting that the Tarr et al. (1998) 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, photo-realistic images). This suggested that 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 other 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 shown with lighting 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, was demonstrated to be lighting dependent.

The lighting dependent behavior exhibited in these various experiments suggests that the object representations of these subjects is not edge-based, since edge-based models posit that recognition performance should be invariant across illumination changes over an object. Instead, an image-based object representation that retains information about how the appearance of an object changes with respect to known illumination conditions is a more plausible explanation for the behavior of these subjects.

Neuropsychological Evidence. Results from several studies of patients with cortical deficits suggest that the processing of information concerning illumination conditions in scenes is lateralized to the right posterior cortex. Performance measures on a prototypical/non-prototypical lighting task, in which patients tried to identify objects shown with either a conventional (even illumination) or an unconventional (uneven lighting on the object) lighting condition (Warrington and Ackroyd, unpublished, as cited in Warrington, 1982), demonstrated that patients with damage to the right posterior cortex were more impaired in their recognition performance than patients with left posterior damage. The unevenly lighting in the unconventional illumination condition caused severe shadows in the image.

Etcoff, Freeman, and Cave (1991) also reported on a prosopagnosic patient, L.H., with right anterior temporal and frontal cortical brain loss, and dilation of the left temporal horn of the lateral ventricle. This patient performed poorly on a

task involving identifying objects under varying illumination conditions. While the authors suggested that the task performance was not due to any perceptual categorization deficit, since L.H. performed adequately on a task involving changing viewpoint, Warrington and James (1986) argued that the change in illumination on the objects in the task could have degraded the distinctive features of the objects more than merely rotating them in space. This evidence suggests that a representation that preserved the lighting parameters of the scene was used by L.H. for accurate object recognition under varying illumination conditions, and that this representation was unaffected by the right anterior temporal deficit. Edge-based models do not prescribe such a representation, as they discount surface characteristics in favor of shape contours; instead, the intact object representation was most likely image-based.

Neurophysiological Evidence. Neurophysiological studies also contribute to the evidence suggesting that variations in lighting are present in object representations to the extent that these representations correspond to particular neural codings in the cortex. A single-cell recording study was performed in macaque monkeys to study the effects of changing lighting conditions on faces for face-selective cells in the anterior upper bank of the superior temporal sulcus (STS) (Hietanen, Perrett, Oram, Benson, & Dittrich, 1992). The authors isolated the preferred face view for each cell they wished to study, and then tested the cells by displaying several images of the preferred face, each with different directions of lighting on the face. The illumination directions used were lighting from the front, from above, from below, and lighting from the side. Most

of the cells showed complete lighting direction generalization (to directions as disparate as 90° apart), meaning that the firing pattern of a cell did not change with variations in illumination direction over preferred face view of that cell. However, some of the recorded cells only responded to certain lighting directions, similar to the specificity found for viewpoint in IT and STS cells (Rolls, 1994; Logothetis & Pauls, 1995). The authors stated that the cells with some lighting specificity contributed to overall illumination generalization when a small proportion of the cells were considered together as a population. These neurophysiological studies indicate that representing lighting in the scene is important to object recognition, in so far as there are cortical cells that respond to changes in illumination.

What is *Invariant* Object Recognition?

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 Computer 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. Thus, computer vision models of object recognition, particularly 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 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 sample images of all individuals in the projected space. 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 projected 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. This information can cause errors in subsequent recognition, since variations between images due to illumination are usually larger than those due to individual identity (Moses, Adini, and Ullman, 1994).

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 in the Eigenface representation are primarily due to changes in illumination and may be discarded for purposes of recognition. Georghiades et al. (1998) also implemented this approach in their comparison of several models of face recognition under variations in illumination. Consistent with their observation, the Eigenfaces method did achieve better recognition performance without the first three principal components. However, even with this modification, under some lighting conditions the Eigenface model failed 27% of the time, and

showed a gain of only 16% over the Eigenface model with all principal components (Georghiades, Kriegman, and Belhumeur, 1998).

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 maximize the ratio of the projected between-class scatter to the projected within-class scatter in order to provide better discrimination between individual faces. By doing so, they were able to produce a set of eigenvectors with reduced dimensionality but without the confounding variability due to illumination. To avoid a singular within-class scatter matrix, Belhumeur et al. (1997) created a technique they referred to as "Fisherfaces." This method used PCA to first achieve a non-singular within-class scatter matrix by reducing the dimensionality of the image space down to $(N-c)$, where N is the number of sample images and c is the number of faces. It then applied Fisher's Linear Discriminant to reduce further the dimensionality of the feature space down to a $(c-1)$ dimensional feature space. 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 (Belhumeur et al., 1997). 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, a convex polyhedral cone in the image space whose apex coincides with the origin, contains the set of images of an object under all possible illumination conditions (where all light sources are at infinity). A small set of acquired images of an object is used as a basis set to construct its individual illumination cone; as few as three distinct images for each object can determine a given object’s cone in some cases. 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 (Georghiades et al., 1998, 2000). Thus, with regards to illumination invariance, the IC model offers potentially far better performance than competing approaches.

The Current Problem

The experiments presented here investigate how humans perform under conditions with varying illumination conditions. Observers were trained to recognize faces with lighting configurations designed to elicit different mental representations of lighting in each instance. For example, in one study the lighting direction varied from frontal to extreme (from one side). A second study used two well-separated lighting directions on either side of, or above and below, the faces. Several of the studies also used “extreme” lighting (illumination directions far from frontal, or typical, directions) during training, and then tested observers with more typical lighting conditions. Almost no studies have been performed of how humans deal with extreme lighting during training, so the results obtained from these studies are invaluable in and of themselves. Testing the Illumination Cone (IC) model (Belhumeur & Kriegman, 1998) with identical lighting conditions as used in the psychophysical experiments should provide the most useful comparisons between human subjects and computer vision models, specifically the IC model.

Of interest here is not simply the relative performance of human subjects and computer vision models, but what assumptions are made in order to make such comparisons and the evaluation of these comparisons. To provide the most useful comparison between human subjects and the IC model, we strived to closely mimic the experimental procedures used in the human psychophysical experiments in our execution of the IC model. However, the IC model in no way implements the large bulk of what we think of as vision. Therefore its output is in many ways derived under entirely different conditions

from the data obtained with humans, who necessarily bring their entire visual system into play in recognizing faces. Thus, the IC model may be at somewhat of a disadvantage, yet it performs as well as or better than our human observers under some conditions. In large part this may be due to the fact that the IC model only “knows” about a small subset of all possible images. In contrast, humans are equipped with a lifetime of experience and knowledge of 100,000’s of objects. This apparent disadvantage also has positive implications for human observers. Specifically, most humans are face experts, and thus have class-level knowledge regarding the appearance of faces *in general*. This knowledge allows us to rapidly learn and recognize entirely novel faces, as well as generalize from a single view of a face to an entirely new lighting (or viewpoint) context – the idea being that other faces have been seen under the new experimental conditions. In contrast, the IC model has no knowledge of faces beyond the training it receives, and it never generalizes between individual faces.

These same issues arise in nearly every extant computer vision model that is compared to human data. Nearly all address only a small part of the “vision problem”; in contrast, the human observer applies a complete vision system that includes filtering, sophisticated mid-level organization, and a rich representational space (and years of learning). It would be a mistake to claim that a given method does any more than model one specific mechanism of human vision. In the case of the IC model, that mechanism 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 a larger vision system, but it may be the particular component that determines how performance modulates across lighting variation. Therefore the *patterns* of generalization from known to unknown lighting conditions may be compared between the IC model and our human subjects. Similar comparisons are possible in many domains so long as one is willing to make explicit the assumptions used in both the computational model and the analogous psychophysical experiments. Indeed, we argue that such comparisons ultimately improve both sides of the problem – refining the algorithms used in computational implementations and constraining the space of solutions for explaining elements of human vision.

In the next two chapters, experiments are described that detail the previously discussed investigation of object recognition and illumination. The experiments in Chapter 2 were designed to explicitly examine how systematically changing the type of trained lighting conditions would affect the behavior of human observers and a computer vision algorithm, the Illumination Cone (IC) model. By presenting the same stimuli in both the human psychophysical experiments and in the computational simulations, certain inferences could be made concerning the specific mechanisms by which humans exhibit illumination invariant object recognition under varying illumination conditions.

Using a method conceived to study possible models of representing viewpoint in object recognition, the experiments in Chapter 3 were designed to

investigate interpolation in the representation of lighting direction with respect to object recognition. These experiments were also designed to answer questions concerning the effects of object geometry on any stored lighting representation.

CHAPTER 2

Identifying Faces Across Variations in Lighting: Psychophysics and Computation

Given the superior performance of the IC model (as described in the Introduction) when compared to other image-based computer vision models that somehow deal with illumination in the image, 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 computer vision 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 information that might account for human performance. Second, there is the possibility that there will be significant correlations between human and model performance. In this case, we can draw stronger conclusions, and may be able to 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 that used “extreme” lighting directions during training, i.e., lighting directions in which the majority of the face was not illuminated, as well as more “standard” lighting directions for training, e.g., frontal illumination. Interestingly, few 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 perform when trained to recognize images of objects with extreme lighting directions; so again, the data obtained here is valuable in and of itself. However, the most informative analysis here 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 understanding of algorithms in biological vision.

General 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 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 name (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 key 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 each experiment. 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 cm 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).

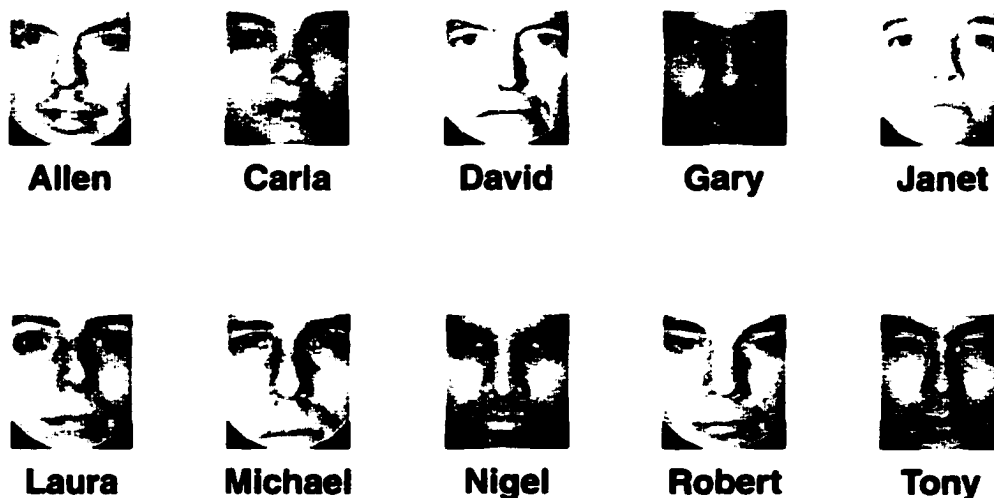


Figure 1. Example of the images and names subjects viewed prior to the start of the computer-based trials. Subjects viewed these images for 10 minutes in order to learn to associate the correct name to each face.

The images used in the computer-based trials were 651 images taken from the Harvard Face Database (see Hallinan, 1994; 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 images are cropped so that the hairline, the ears, and the necks of the people are missing. The cropping was done to eliminate these occluding contours because the surface reconstruction done by the models for which this image set was originally conceived could not handle the image gradients at these points. Since the database came with these cropped images, and the Illumination Cone model is one such model that performs a surface reconstruction, these cropped images were used for both the human psychophysical and the computer vision experiments. 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 had missing or corrupted images and had 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 Georgiades, Kriegman, and Belhumeur (1998) in a human psychophysical experiment.

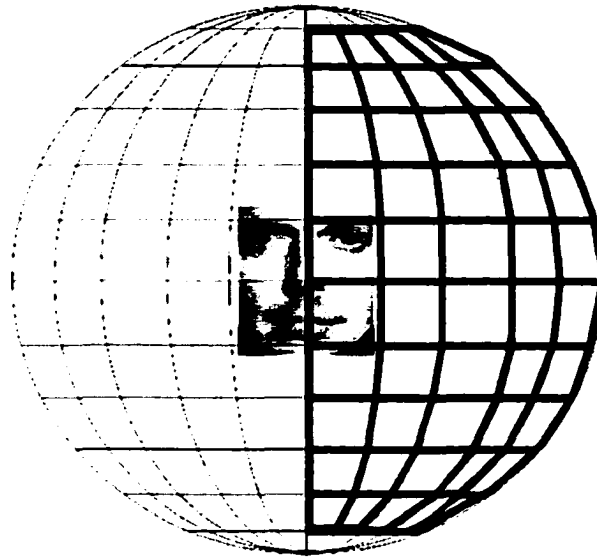


Figure 2. Schematic of the 66 different illumination conditions used in both the human psychophysical and IC model experiments. The illumination conditions correspond to the intersections of the longitudes and latitudes overlaid with bold lines. The center of the space is denoted $(0^\circ, 0^\circ)$. 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

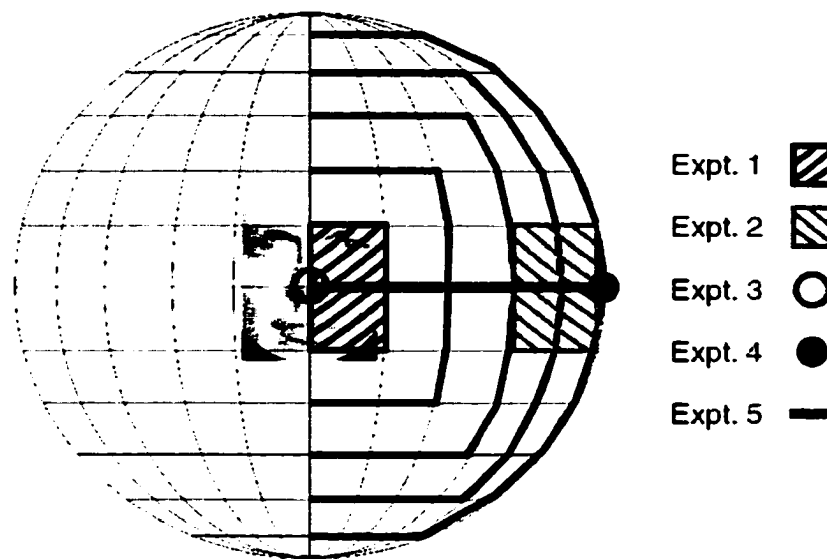


Figure 5. Training sets for the human psychophysical and IC model experiments. The training set for Experiment 1 contains illuminations within 15° of the camera axis. The training set for Experiment 2 is a mirror of Experiment 1 with extreme lighting directions. Experiments 3 and 4 only have one illumination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

Model Simulations

While replicating the method used by Georgiades 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 behavioral experiments, with some modifications to help equate the experience of humans and the computational model.

In the rendering of the Illumination Cone model used, the training phase occurred with the construction of the illumination cone for each individual face (Belhumeur & Kriegman, 1998; Georgiades et al., 1998, 2000). As previously described, an illumination cone contains all possible images of an object under

an arbitrary number of point light sources at infinity. An illumination cone is described by the following equation:

$$C = \{x : x = \sum \max(Bs_i, 0), \forall s_i \in \mathbf{R}^3, \forall k \in \mathbf{Z}^+\}$$

where x is an image in the illumination cone; B is a matrix whose rows represent the product of the albedo and a unit surface normal directed inward from a surface point projecting to a certain pixel in the image; s is a column vector representing the product of the light source strength with the light source direction (as a unit vector); \mathbf{R}^3 is the set of real numbers (3-space); and \mathbf{Z}^+ is the set of positive integers. Belhumeur and Kriegman (1998) present a complete derivation and proof of the Illumination Cone (IC) model.

The cones constructed according to the trained lighting directions 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 that 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 ($0^\circ, 0^\circ$) illumination condition separate from the illuminations in the training set. Due to this additional illumination component, the IC model was also given the ($0^\circ, 0^\circ$) 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 ($0^\circ, 0^\circ$) 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 $(0^\circ, 0^\circ)$ or $(75^\circ, 0^\circ)$, respectively. These training sets are illustrated in Figure 6. This was a departure from previous training, since the other experiments used uniformly defined lighting 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, how well the IC model would generalize to other unknown conditions given a non-uniform set of illuminations could be determined. 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.

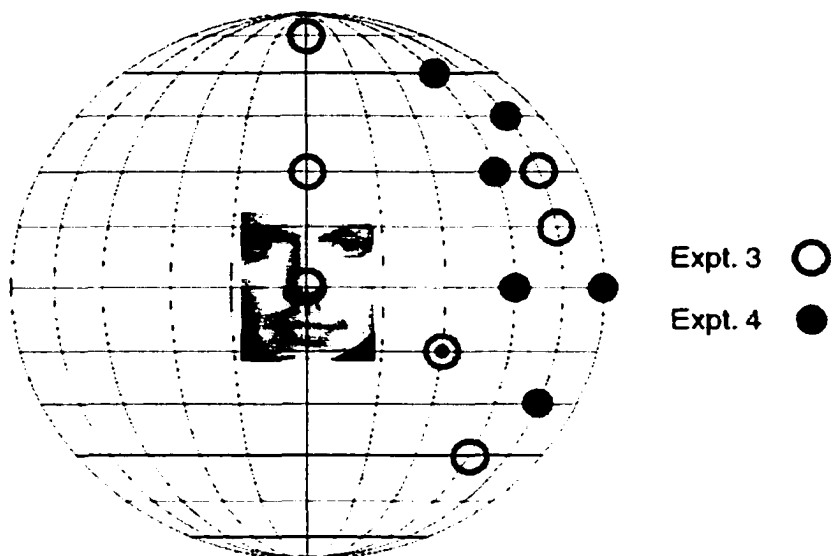


Figure 6. Training sets for Experiments 3 and 4 for the Illumination Cone (IC) model. Open circles are lighting coordinates for Experiment 3 and filled circles are coordinates for Experiment 4. Coordinate $(30^\circ, -15^\circ)$ is a training illumination for both experiments. Adapted from an illustration in Georghiades, Kriegman, and Belhumeur (1998).

Another component of the IC model was building the basis vectors 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 and reflected along the vertical axis (in order to double the number of images used in the calculations). Thresholding was performed on the images according to the saturation and shadow parameters previously defined, and then the image set was 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 a representation 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.

Results

Similar analyses were run on the data from the human psychophysical and the IC model experiments. Any differences in the analyses between the two are explained 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 7 shows an example of the different illumination conditions that comprised these five bins for the training set from Experiment 1. The dependent variable across all experiments was percent correct recognition, i.e., the ability to correctly identify each individual face, for each lighting condition for both human subjects and the IC model.

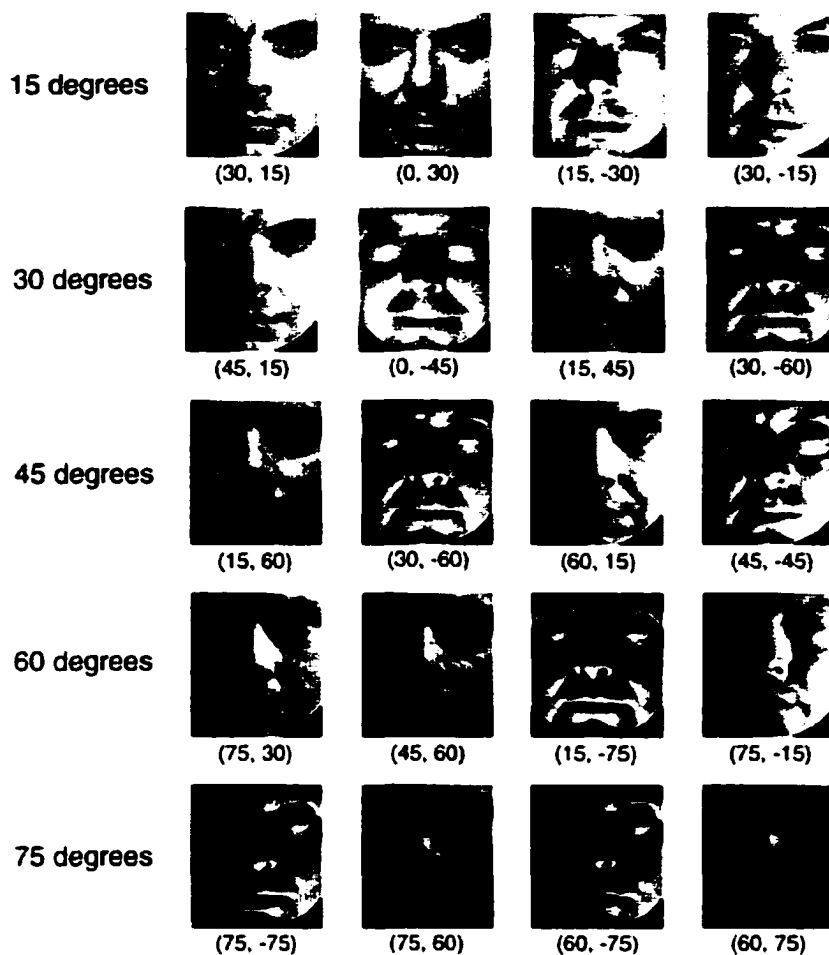


Figure 7. Sampling of test images for Experiment 1, and their respective distances from the closest illumination direction in the training set. The training set for Experiment 1 consisted of illuminations at the following coordinates: $(0^\circ, 0^\circ)$, $(0^\circ, 15^\circ)$, $(0^\circ, -15^\circ)$, $(15^\circ, 0^\circ)$, $(15^\circ, 15^\circ)$, $(15^\circ, -15^\circ)$.

Human Psychophysics

Subjects that failed to respond to more than 10% of the test trials were removed from the study. A trial with no response indicated that the trial timed out (i.e., 3000 msec elapsed) before the subject responded. 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 8 through 12 (Experiments 1 through 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, (see Figures 8, 10, and 12) where the images in the training sets contained mostly frontal or near-frontal illuminations or, in the case of Experiment 5, the trained lighting directions 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 (see Figures 9 and 11), where the training sets contained extreme illumination directions that produced images with pronounced shadows.

One explanation for the fall-off in performance with distance in Experiments 1 and 3 (see Figures 8 and 10) is that the 60° and 75° lighting conditions included images with extreme illuminations. Because so much of the face was in shadow, subjects had little information available to discern the identity of the face. In contrast, in Experiments 2 and 4 (see Figures 9 and 11), 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° test lighting

conditions in these two experiments were comprised of near-frontal illuminations. The near-frontal lighting made establishing the identity of the faces easy, despite the unfamiliarity of the lighting directions, compared to the analogous test lighting directions in Experiments 1 and 3.

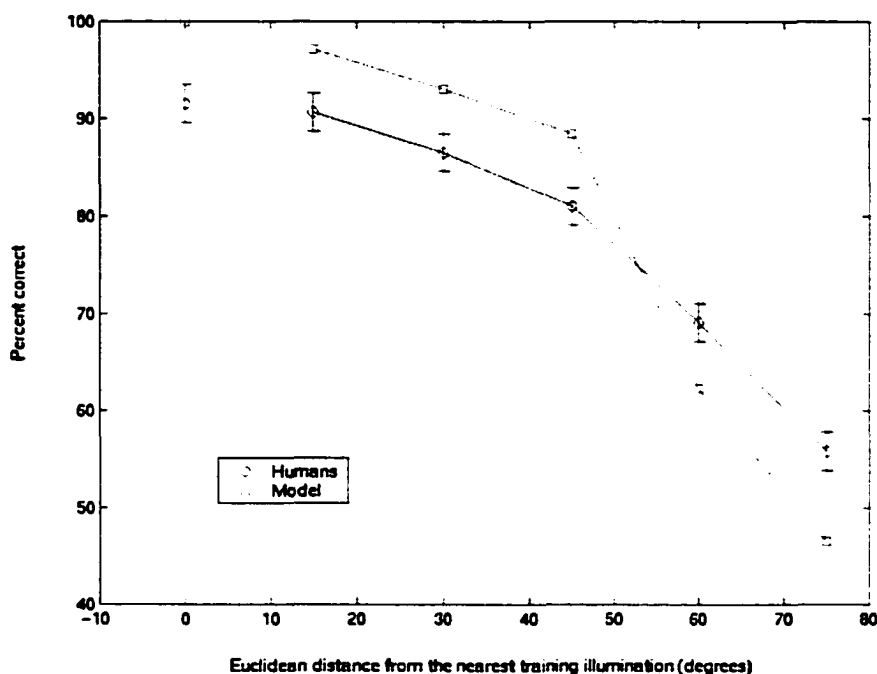


Figure 8. Results for both human subjects and the Illumination Cone (IC) model for Experiment 1 (training with near-frontal lighting directions). The trained illumination directions were within 15° of the camera axis ($0^\circ, 0^\circ$). The circles represent mean percent correct for the human subjects. The error bars are the within-subject standard error of the mean. A single case of the IC model is represented by the squares.

Note that although during the initial name association task subjects had the benefit of viewing the frontal ($0^\circ, 0^\circ$) 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 (see Figures 9 and 11), 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).

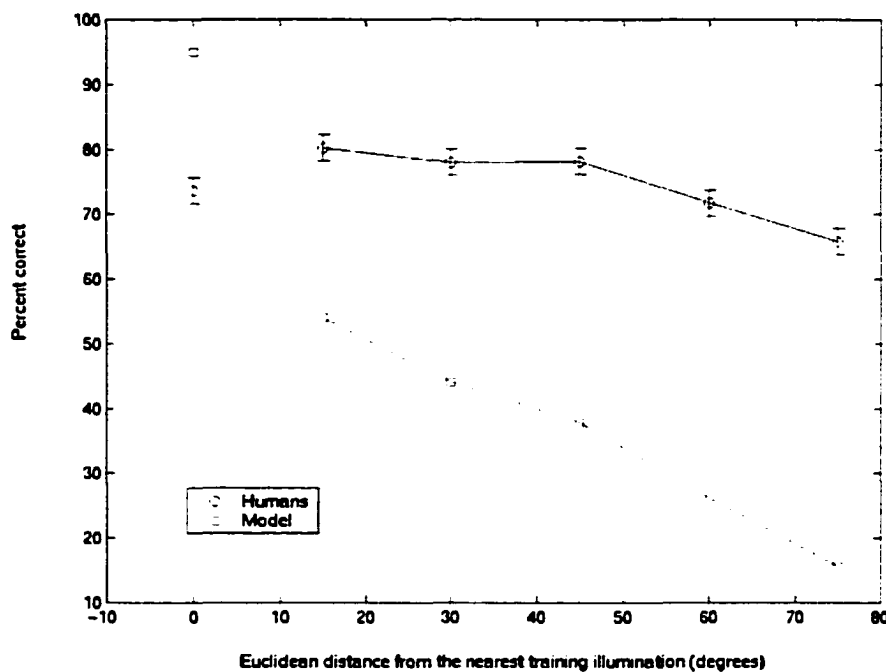


Figure 9. Results for both human subjects and the Illumination Cone (IC) model for Experiment 2 (training with extreme lighting directions). The trained illumination directions were within 15° of (75°, 0°). The circles represent mean percent correct for the human subjects. The error bars are the within-subject standard error of the mean. A single case of the Illumination Cone (IC) model is represented by the squares.

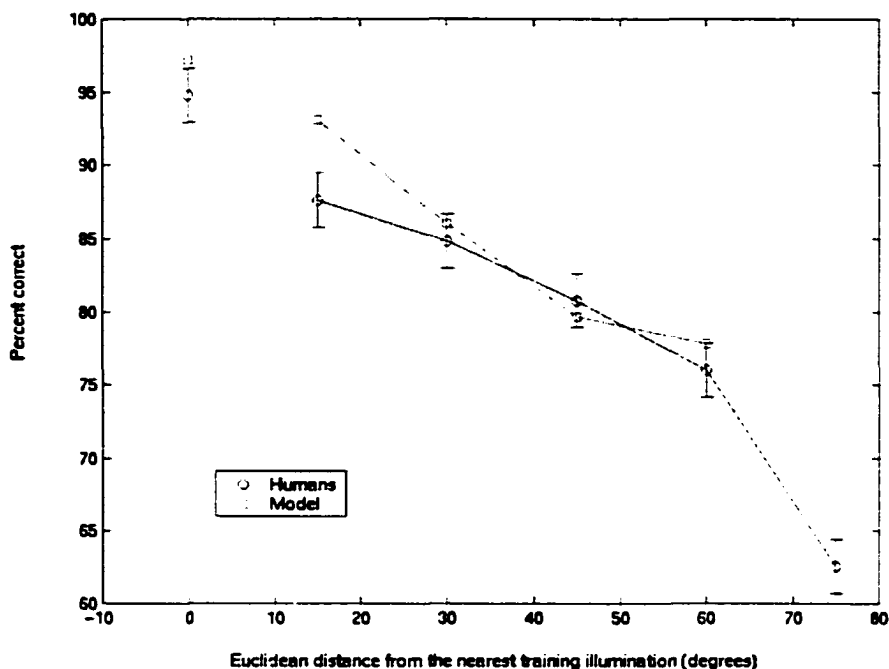


Figure 10. Results for both human subjects and the Illumination Cone (IC) model for Experiment 3 (training with a single frontal lighting direction). The trained illumination direction was on the camera axis, $(0^\circ, 0^\circ)$. The circles represent mean percent correct for the human subjects. The error bars are the within-subject standard error of the mean. A single case of the Illumination Cone (IC) model 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.

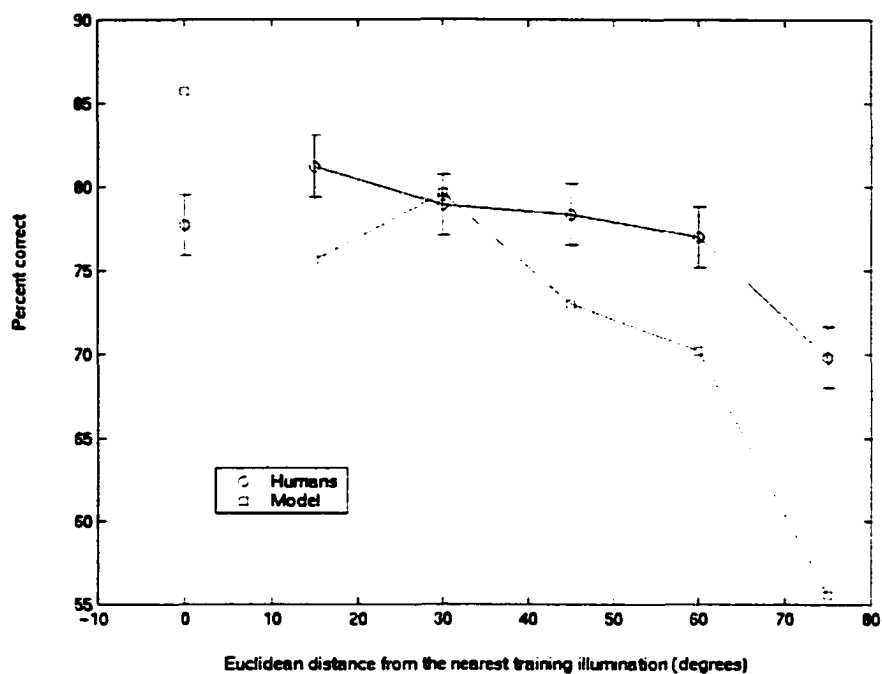


Figure 11. Results for both human subjects and the Illumination Cone (IC) model for Experiment 4 (training with a single extreme lighting direction). The trained illumination direction was (75°, 0°). The circles represent mean percent correct for the human subjects. The error bars are the within-subject standard error of the mean. A single case of the Illumination Cone (IC) model is represented by the squares.

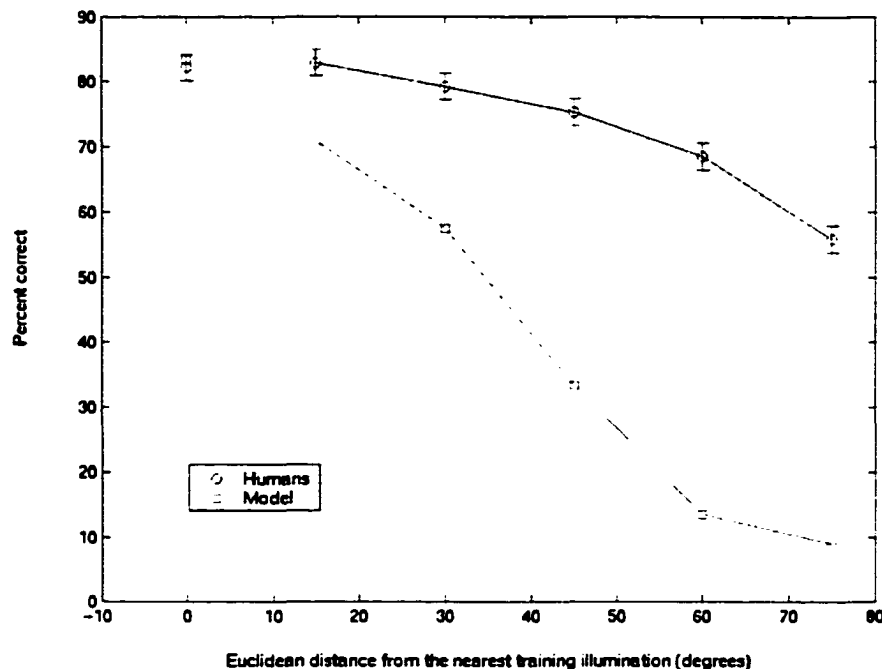


Figure 12. Results for both human subjects and the Illumination Cone (IC) model for Experiment 5 (training with lighting directions along the horizontal axis of the lighting space). The trained illumination directions were along the horizontal meridian between $(0^\circ, 0^\circ)$ and $(75^\circ, 0^\circ)$. The circles represent mean percent correct for the human subjects. The error bars are the within-subject standard error of the mean. A single case of the Illumination Cone (IC) model is represented by the squares.

Illumination Cone (IC) Model

The data points marked with squares in Figures 8 through 12 illustrate the recognition performance of the IC model and can be compared to the performance of the human observers in Experiment 1 through 5. An attempt was made to maximize the performance of the IC model by trying a variety of saturation and shadow thresholds for each face in each experiment. The saturation and shadow thresholds were adjusted for each face until the model was able to successfully construe basis vectors from the training set. These adjustments were made several times, and the parameters that provided the

best performance were chosen. The resulting model performance in each experiment represents the best attempt at manually 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 is seen 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. However, the performance of the IC model during training was always better than human performance in the same experiments, suggesting that the model was performing adequately.

The parameter search could have been done differently. The IC model could have been modified to search for the best parameter fit based on maximizing the amount of information in the images (i.e., keeping as much saturation and shadow as possible), while optimizing performance on the training set. This procedure would allow only those thresholds for which the IC model exhibited optimum (100% correct) performance on the training set. With the model exhibiting this level of performance during training, its performance on the test illumination directions might be more valid. Anecdotally, the experimenter noticed that the saturation and shadow thresholds for which the IC model exhibited better performance for the training set did not necessarily guarantee the best performance for the model on all of the tests sets. This suggests that the best strategy for finding the optimum saturation and shadow thresholds

might involve finding those thresholds that boosted performance for all illumination conditions.

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 occurred in Experiment 4 (see Figure 11); 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 (see Figure 6), the illuminations comprising the 30° bin may actually have mapped 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” (see Figure 6). 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 performance 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 an attempt was made to approximate human experience by including six additional randomly selected lighting directions (for the specific images used, see Figure 6). 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.

These results suggest 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 (see Figure 12).

Comparing Psychophysical and Computational Results

How do we compare the performance of a putative computational 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 computer vision model did approach the level of human performance, it would most likely be for very different reasons. Thus, the fact that the IC model actually does come close in many cases 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-à-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 it is intriguing that model accuracy and human accuracy are sometimes quite close).

The overall general pattern 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 (i.e., performance during the training phase), while the second set does not contain these data. The correlations without the training sets are probably more representative of the similarity between human performance and the IC model. 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. By comparison, 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.

Experiments	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 through 5.

The Microstructure of Generalization. Beyond the fact that as test images were further and further from training images in terms of illumination direction 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 (defined here as a monotonic change of equal magnitudes in performance) 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 8 through 12 because we “binned,” or grouped, 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. That is, for lighting directions less than 45°, there was a less pronounced performance falloff for both humans and the model, while for lighting directions greater than 45°, the performance falloff was more pronounced. Another specific similarity between humans and the IC model is that both showed poorer overall performance in Experiment 5 relative to their own performance in Experiments 1 through 4. Thus, both human subjects and the IC model appear sensitive to an accidental alignment of all lighting directions during training.

This microstructure analysis of lighting generalization is the only way to compare human and model performance because the pattern of responses is what matters. If human and model performances degrade in similar ways, then inferences about how the systems compensate for changes in illumination can be made with respect to each other. Such microstructure comparisons are important for understanding exactly how humans and computational systems handle lighting variability and should be explored in more detail in future studies.

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 revealed that humans 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 allowed for the recognition of objects across varying lighting conditions (Hallinan, 1994; Belhumeur & Kriegman, 1998).

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 for humans 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, this effect was observed 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 a direct comparison 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, 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, the generalization performance of human observers and the IC model were tested 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 between human subjects and the IC model.

Critically, the nature of the training images was manipulated in each experiment. Experiment 1 used a set of near-frontal lighting directions, Experiment 2 used a set of extreme lighting directions (opposite to those used in Experiment 1), Experiment 3 used a single frontal lighting direction, Experiment 4 used a single extreme illumination direction (opposite to that used in Experiment 3), and Experiment 5 used a training set that spanned the horizontal meridian of the lighting space from frontal to extreme lighting. Across these different training conditions the following results were obtained:

- 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 the trained 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.

Of course, some of the above differences are inherent in the comparison being made between the full vision system of humans and the 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 to the expectation for human observers to display both better generalizations across all unfamiliar illumination conditions and dramatically better generalizations 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 (10 individuals in this implementation) for an individual face under the trained illumination conditions, while humans have 1000's of possible matches (due to experience), might lead to the expectation that the model should 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 those findings are extended 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 computer vision systems. Specifically, a computational model was employed 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 comparing human performance to recognition systems that address the question of lighting variability using different algorithms from the IC model. For example, the approaches implemented both in the Lades, Vorbruggen, Buhmann, Lange, von der Malsburg, Wurtz, & Konen (1993) and Atick, Griffin, & Redlich (1996) computer vision 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 may provide additional constraints on the extraction of a lighting model for the scene, as well as the structure of the object.

CHAPTER 3

Lighting and Face Recognition:

Evidence for Illumination Dependent Representations

The experiments discussed in the previous chapter provide evidence that humans exhibit illumination dependence, and this suggests that a model of illumination is used by the human visual system. The results, from identical methods used in both human psychophysical and computer vision experiments, suggest that the algorithms used in the Illumination Cones (IC) computational model (Belhumeur & Kriegman, 1998) are similar to the mechanisms responsible for human object recognition under variable lighting conditions.

The previous chapter's results show that the Illumination Cone approach does not predict perfect generalization from known to unknown lighting conditions. As the lighting direction becomes more distant from those used during training, the recognition performance of the algorithm suffers. The IC model only provides a mechanism for compensating for changes in illumination. Although the model constructs a representation of the entire lighting space, this

representation is based on the images input to the model during training, and these training images greatly influence the quality of the representation, as one would expect with an image-based model. The representation is more robust for lighting directions close to the trained lighting directions since those images contain scene parameters similar to the trained ones and, therefore, the model predicts better recognition performance for lighting directions near the known (trained) illumination conditions. This performance bias to lighting directions near the trained illumination conditions is similar to the prediction of models that use interpolation to compensate for changes in object viewpoint (e.g., Poggio and Edelman, 1990).

These viewpoint-interpolation models use a small number of images, each of which contains the object at a different viewpoint, to build a representation that describes the object at unknown viewpoints by synthesizing them through interpolation of the known viewpoints, much in the same way that the Illumination Cones (IC) model builds a lighting representation. Similar also to the performance of the IC model in recognizing objects with illuminations close to the trained illumination conditions, these viewpoint-interpolation representations predict better object recognition performance at unknown viewpoints that are close to the trained viewpoints.

Poggio and Edelman (1990) proposed a viewpoint-interpolation model that used a network that learned to recognize three-dimensional objects from two-dimensional views. They suggested that encoding a sufficient number of two-dimensional views of an object was equivalent to having the specific three-

dimensional structure of the object. They derived this idea from the work of Ullman (1979), who specified a model of structure-from-motion to specify the three-dimensional structure of an object by a set of feature points (at least 5) present on two perspective views of the object. Poggio and Edelman's model uses a regularization scheme based on Generalized Radial Basis Functions (GRBF) that approximate an object's three-dimensional structure from any perspective view given a small number of basis functions (two) and a set of "familiar" views (10-40 views per object). The GRBFs are set in the middle layer of the network and compute the distance of the input view from a fixed standard view, represented by the center of the basis function, and this value is applied to a weighted distance function. The resulting value is the activity of the GRBF. The output of the network is a linear superposition of the activities of all the basis units in the network, and a weighted combination provides an output standard view of the object. This model is able to accomplish 3-D object recognition from 2-D image representations fairly well by comparing the input image to this interpolated viewpoint surface of all possible viewpoints. This method is similar to the illumination representation constructed by the Illumination Cone (IC) model. The IC model uses several training images to construct three orthogonal basis vectors that are used to synthesize images of the object under all possible lighting directions. The object representation created by this viewpoint-interpolation method predicts that unknown object viewpoints that are close to the trained viewpoints will create output views of the object that will be similar to the standard image elicited by the training

views. Bülthoff and Edelman (1992) performed several psychophysical experiments designed to determine if this viewpoint-interpolation model predicted human performance.

The study by Bülthoff and Edelman (1992) investigated whether objects were represented as three-dimensional models or as two-dimensional “snapshots.” They looked at the predictions of three models: 1) the model proposed by Ullman (1989) that compared the input image of an object with the projection of a stored three-dimensional object model, which they called “recognition by alignment”; 2) Ullman and Basri’s (1991) model of recognition that linearly combined several two-dimensional views of an object; and 3) the model discussed above that performs recognition through approximation across an interpolated hypersurface of all possible two-dimensional views (Poggio & Edelman, 1990; Poggio & Girosi, 1990). The results of the Bülthoff and Edelman (1992) experiments showed that observers committed fewer errors in recognizing “paperclip” objects when the object viewpoints shown during testing were *between* two trained viewpoints, as opposed to being outside of the trained viewpoints, or orthogonal to the axis defined by the training viewpoints. They argued that this good performance was due to the observers interpolating between the two trained viewpoints to accurately recognize the objects.

As stated previously, the architectural assumptions of the viewpoint-interpolation model of Poggio and Edelman (1990) and the Illumination Cone (IC) model of Belhumeur and Kriegman (1998) are quite similar. Specifically, both models build object representations that are hypersurfaces of all possible

views or lighting conditions, respectively. The models also both predict that performance will be better for views or lighting directions, respectively, that are closer to those known to the representations of the models. Given these similarities between the viewpoint-interpolation model and the Illumination Cones (IC) model, the present set of experiments has two goals: 1) Examine the specifics of how a model of illumination might function given the need to use interpolation in order to achieve adequate object recognition; 2) Determine whether object geometry and its relation to lighting direction affect how the illumination model is computed.

To achieve these goals, the present study uses the methodology prescribed by Bülthoff and Edelman (1992) in their study of object representation and viewpoint to determine the role of interpolation in models of illumination. Instead of varying the views of the faces, the direction of lighting will vary on the faces while the view remains constant.

General Methods

Observers

Seventy-seven people participated across four experiments. These subjects were mostly college students between the ages of 18 and 23 years. The allocation of males and females was close to equal in each experiment, except for Experiment 1 in which twice as many females participated. Table 2 shows the distribution of subjects in each of the four experiments. All subjects had normal or corrected-to-normal vision. Subjects were naïve to the purpose of the

experiments. When finished with the session, observers were informed of the intent of each experiment.

	Females	Males	Total
Experiment 1	12	6	18
Experiment 2	12	12	24
Experiment 3	10	9	19

Table 2. *The distribution of observers across the three experiments described in Chapter 3.*

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 sticker was placed over the 'v' and 'm' keys on each keyboard. The sticker over the 'v' key stated 'diff' and the sticker over the 'm' key stated 'same'. Observers used the 'v' ('diff') key to indicate that the face presented during the trial was a different identity from the face shown during training. Conversely, they used the 'm' ('same') key to specify that the face shown during the trial was the same as the identity learned during training.

The face images used in all of the trials were 396 images (370 images were used in Experiment 3) taken from a database of 1800 images of 100 faces (50

males and 50 females) produced at the Max-Planck Institute for Biological Cybernetics under the supervision of Heinrich Bülthoff. The images were collected as 3D head models and associated color maps using a Cyberware 3D scanner and then modified using a morphing algorithm developed at the Max-Planck Institute. The database is available from the Institute website (<http://faces.kyb.tuebingen.mpg.de/>). The images were created such that each face had 18 different illumination conditions associated with it. The lighting directions changed in 10° increments both horizontally and vertically. The experiments presented here used 16 of the possible 18 illumination conditions, except for Experiment 3, which used 15 lighting conditions. Two of the lighting directions were used during training with the others shown during testing. Sixteen individual faces were used as targets (8 male and 8 female) and 10 individual faces were used as distracters. An amalgamation of randomly sized and positioned face pieces was used as a masking stimulus. Unless otherwise noted, the faces were presented in an upright orientation, i.e., eyes above mouth. The faces in the images were also looking to the left (the right of the computer screen as the image was shown). The orientation of the face was approximately 10° off the center of the camera axis. Examples of the stimuli are shown in Figure 13.

Procedure

For each experiment, observers were randomly assigned to one of six groups. All groups received the same conditions but in different orders. In each group, stimulus images were presented in three 20-minute sessions. Self-

regulated breaks were provided in between the sessions. Observers performed five to six sets of randomized trials per session. Each set of trials consisted of two training images followed by 84 test images. Of these 84 images, 14 corresponded to the target (previously trained face) and 70 were distracters. Observers viewed the two training images at the beginning of each set of trials for 15 seconds each with a 1 second blank between the two images. The order of the two lighting directions represented in the images was randomized. The faces were illuminated from one of two directions, depending on the experiment. Observers were told to study the images of the face carefully because they would “be asked to identify the face in the next phase.” The training procedure for Experiment 1 is illustrated in Figure 13.

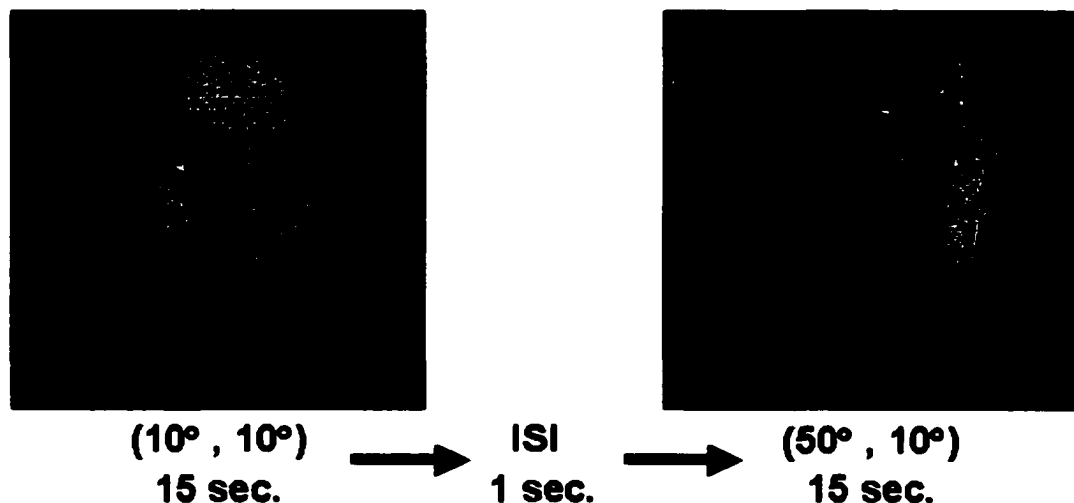


Figure 13. Example of the training sequence in Experiment 1. Each of the two training images was displayed for 15 seconds separated by a 1 second blank. The order of presentation of the two images was randomized. The numbers in parentheses represent the horizontal and vertical displacement of lighting on the face from the center of the image.

For the testing portion of the experiment, observers were instructed to respond as to whether the test image was that of the face previously seen during training, or a different face. Each test trial started with a 500 msec fixation cross in the middle of the screen, followed by the stimulus for 150 msec, the mask for 500 msec, and then a blank interval for 1000 msec before the beginning of the next trial. No feedback was provided during the test trials. This testing procedure is illustrated in Figure 14.

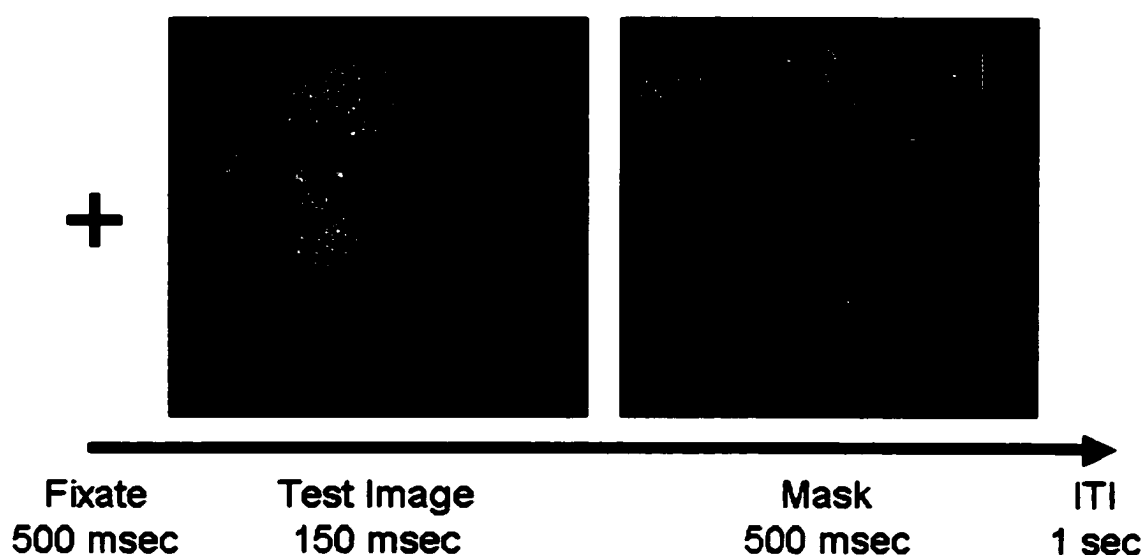


Figure 14. The test trials started with a fixation cross in the center of the screen displayed for 500 msec. The centered test image then was displayed for 150 msec. The test image shown has lighting rendered at -20° on the horizontal and 10° on the vertical axis. A mask was displayed after the test image for 500 msec. The same mask was shown across all trials and all observers. A 1 second blank screen followed each test trial.

In all of our experiments, two lighting conditions were used for training on either side of the object to approximate the two different viewpoints that Bülhoff and Edelman (1992) used in their study of viewpoint. Instead of using 15° increments between lighting directions, as they did with their viewpoints, we

used 10° increments between lighting directions. We also varied the lighting directions in the EXTRA condition in two directions (away from both trained illuminations) instead of just the near-frontal direction. While this was a departure from their method, it provided a measure of performance from two training points (either separately or together), instead of only one. However, a disadvantage of this method was that we lost power in our analysis of the effect of distance from training, since we had fewer distances over which to perform the analysis.

Experiment 1

The present experiment investigated whether learning faces under a specific lighting direction affected the recognition of the same faces under different directions. These studies complement those reported in Chapter 2, but differ in a significant way. The purpose of this experiment, and those that follow, was to examine the underlying object representations used for object recognition with respect to lighting, given that interpolation seems to be needed to achieve good object recognition, as described by Bülthoff and Edelman (1992). The experiments in Chapter 1 mainly tried to confirm if human object recognition and a reasonable computer vision model for handling object recognition under varying illumination were lighting dependent and, if so, to see how the behavior of the two vision systems was similar regarding specific lighting directions.

As such, the present experiments were based on studies of viewpoint-dependency in object recognition run by Bülthoff and Edelman (1992). As in

their earlier work with viewpoint, the lighting directions (rather than viewpoint) used during testing were grouped into one of three conditions: 1) lighting positioned between the two trained illumination directions (INTER); 2) lighting positioned outside of the two trained lighting directions (EXTRA); 3) lighting orthogonal to the axis defined by the two trained illumination directions (ORTHO). In this experiment, the lights in the ORTHO condition were oriented along the vertical axis of the faces.

Methods

The training illuminations were situated at $(10^\circ, 10^\circ)$ and $(50^\circ, 10^\circ)$. The first number in the parentheses is the latitude and the second number is the longitude on the lighting sphere. All of the lighting directions that varied horizontally (longitudinally) were situated on the $+10^\circ$ latitude of the lighting sphere. During the test trials, 14 other lighting directions were shown. The range of the test lighting directions spanned from -20° to 70° on the horizontal and -20° to 40° on the vertical axes of the illumination space. These lighting directions were grouped according to their positions relative to the training illumination directions. The INTER condition consisted of lighting between the two trained lighting directions. Test illuminations outside of the trained illuminations were part of the EXTRA condition. Illumination directions, orthogonal to the axis defined by the two trained lighting directions, were grouped in the ORTHO condition. An illustration of this is shown in Figure 15. The axis defined by the lighting directions in the ORTHO condition always

intersected one of the training illumination directions. In this experiment, the ORTHO condition intersected the $(10^\circ, 10^\circ)$ trained lighting direction.

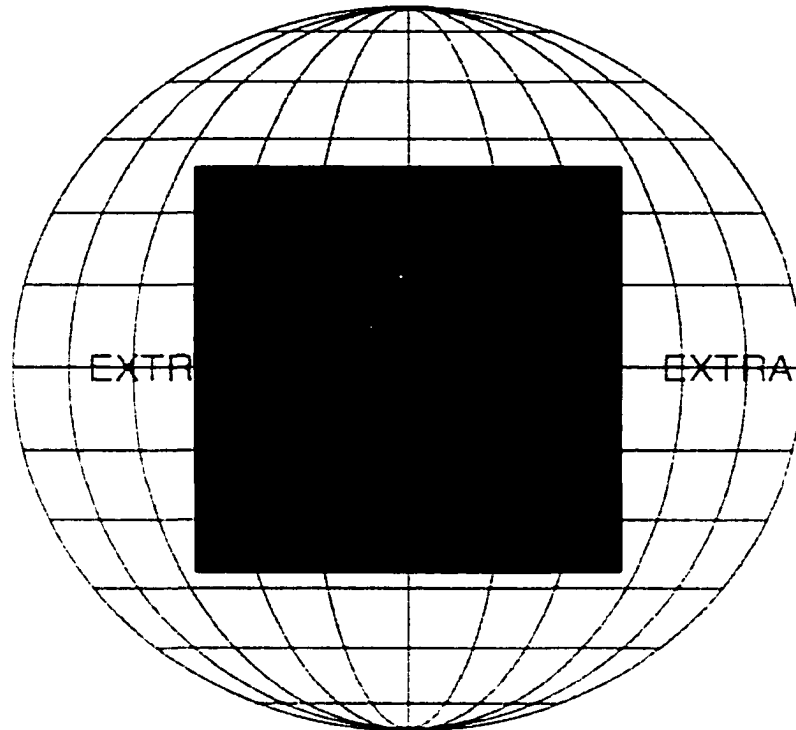


Figure 15. A schematic of the lighting sphere used in Experiment 1 showing the relative positions of the testing conditions to the training illuminations. The dots indicate the trained lighting directions. The conditions in which the test illumination directions were grouped are shown: 1) lighting between the two training conditions (INTER); 2) lighting outside of the trained lighting directions (EXTRA); 3) lighting directions orthogonal to the axis defined by the training illuminations (ORTHO).

Results and Discussion

For all of the test images, the lighting coordinates for each image were recorded, and the minimum Euclidean distance from the nearest trained illumination direction to that tested coordinate was computed, as well as the distance from the test illumination to the near-frontal $(10^\circ, 10^\circ)$ training condition. The dependent variable was mean percent correct recognition

calculated as the ratio of correct old/new identifications to the total number of test trials for a particular individual trained face. Although test trials were timed-out if no response was made within three seconds, subjects responded to more than 95% of the total trials.

With two factors in this within-subjects design (lighting condition and lighting direction), the most logical analysis would be the two-way analysis of variance (ANOVA). However, this analysis is not possible with this data set due to an unequal number of levels in one of the factors. The number of lighting directions within the three conditions (INTER, EXTRA, ORTHO) is not equal. No matter which training directions are used in order to calculate the minimum distance of each test illumination direction from training, there were always an unequal number of lighting directions per condition. Due to this inequity, a two-way within-subjects ANOVA was unfeasible. Instead, we opted to perform several one-way within-subject ANOVAs to discover the main effects of condition and lighting direction on recognition performance separately. Where necessary in analyses with more than one degree-of-freedom in the numerator, the Greenhouse-Geisser (1959) correction was used and is denoted by reporting the epsilon (ϵ) correction value with the F statistic.

The main effect of lighting condition (INTER, EXTRA, ORTHO) was statistically significant with $F(2, 34) = 6.087, p < 0.01, \epsilon = 0.897$. This result is shown in Figure 16. As the graph illustrates, the majority of the effect is present in the EXTRA condition. While a decrease in performance was expected for the ORTHO condition, performance here was on par with the INTER lighting

condition. While this may seem unusual given the orientation of the lighting directions to the trained lighting directions, this result is likely due to the fact that lighting from above is typically seen on faces in the real world. However, when the lighting directions in the ORTHO condition were separated out according to position above or below the faces, there was no clear preference for lighting above the faces in the recognition task. In fact, performance for lighting directions below the face was better on average, as shown in Figure 17. However, the difference between the two groups (above and below) in the one-way within-subjects ANOVA was not significant ($F = 2.924, ns$).

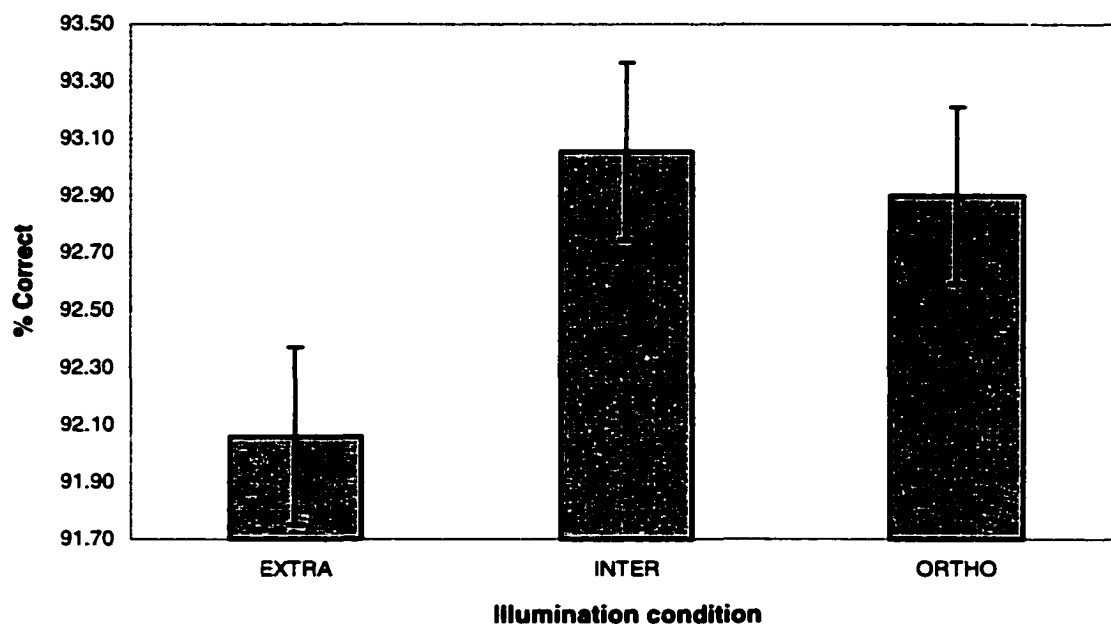


Figure 16. The effect of illumination condition on recognition performance in Experiment 1. The INTER condition includes lighting directions between the training illuminations. Lighting directions outside of the two training directions are in the EXTRA condition. The ORTHO condition contains lighting directions orthogonal to the axis defined by the training illuminations, which in this experiment are oriented vertically on the lighting sphere. Error bars are the within-subject standard error of the mean.

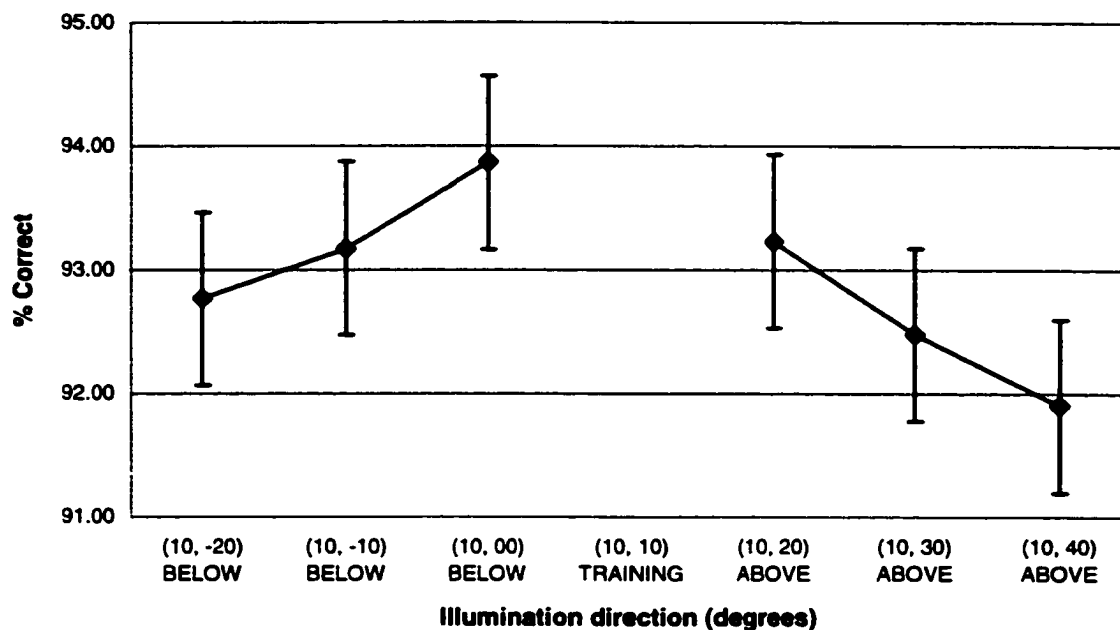


Figure 17. Recognition performance as a function of the lighting directions in the ORTHO condition for Experiment 1. The first number in the parentheses corresponds to horizontal position on the lighting sphere while the second number denotes vertical position (in degrees). The lighting coordinates to the right of the graph lie above the (10°, 10°) training illumination, while those coordinates to the left lie below the training illumination. Recognition performance is better when observers viewed images with lighting below the face than when viewing faces with lighting positioned above.

The distance of the test lighting directions from training were calculated in two ways. In the first instance, the minimum Euclidean distance from *either* of the two training illuminations was found. Using this measure, the INTER condition contains two distances (10° and 20°), while the EXTRA and ORTHO conditions both have three associated distances (10°, 20°, and 30°). Looking at recognition performance within each lighting condition across the distances from either training point, the only significant differences occurred within the ORTHO condition ($F = 3.628$, $p < 0.05$, $\epsilon = 0.954$). However, this result was misleading because the effects of deviating from both of the training points are grouped together.

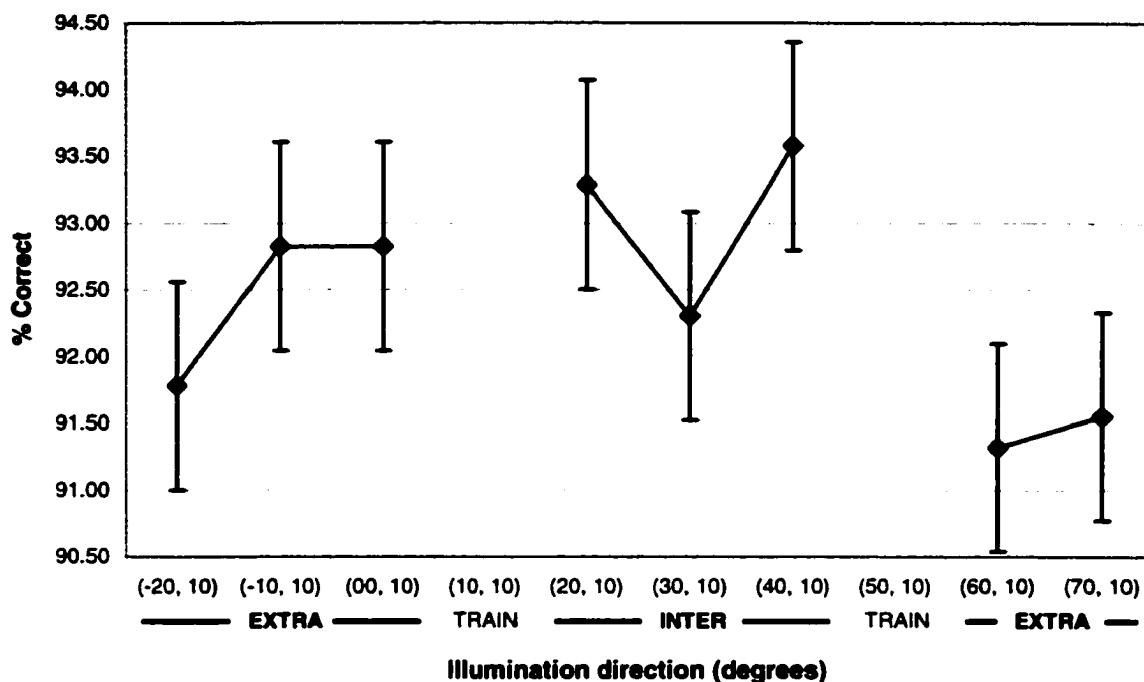


Figure 18. Recognition performance as a function of the tested illumination directions. The illumination directions on the x-axis are labeled as horizontal and vertical coordinates (in degrees) on the lighting sphere. The bolded labels (and lines) mark the INTER and EXTRA conditions as shown. The two trained lighting directions are labeled with "TRAIN."

As Figure 18 shows, as the lighting in the test images diverges from the two training points, recognition performance decreases, especially for the test illuminations closest to the (50°, 10°) training point. There are two items to note here: 1) recognition performance for the (40°, 10°) illumination direction (10° from the training point with extreme lighting, but in the INTER condition) is excellent; 2) performance for the (60°, 10°) lighting direction (also 10° from the training point with extreme lighting, but in the EXTRA condition) is poor. While both of these test illuminations are only 10° from the same training point, observers' performance while viewing them was drastically different. The reason for this dichotomous behavior is probably that the observers' lighting representations for the faces was influenced by their extensive prior knowledge

concerning the relationship of lighting on human faces, i.e., observers were more used to seeing faces with lighting close to frontal.

Grouping the distances from the training points between the three lighting conditions resulted in significant differences between the three distances, 10°, 20°, and 30° ($F(2, 34) = 3.689, p < 0.05, \epsilon = 0.973$), and a significant decreasing linear trend with increasing distance from either training point ($F(1, 17) = 6.209, p < 0.05$). By grouping the three lighting conditions, the overall effect of increasing the illumination angle on the faces away from the training points is seen more clearly (see Figure 19).

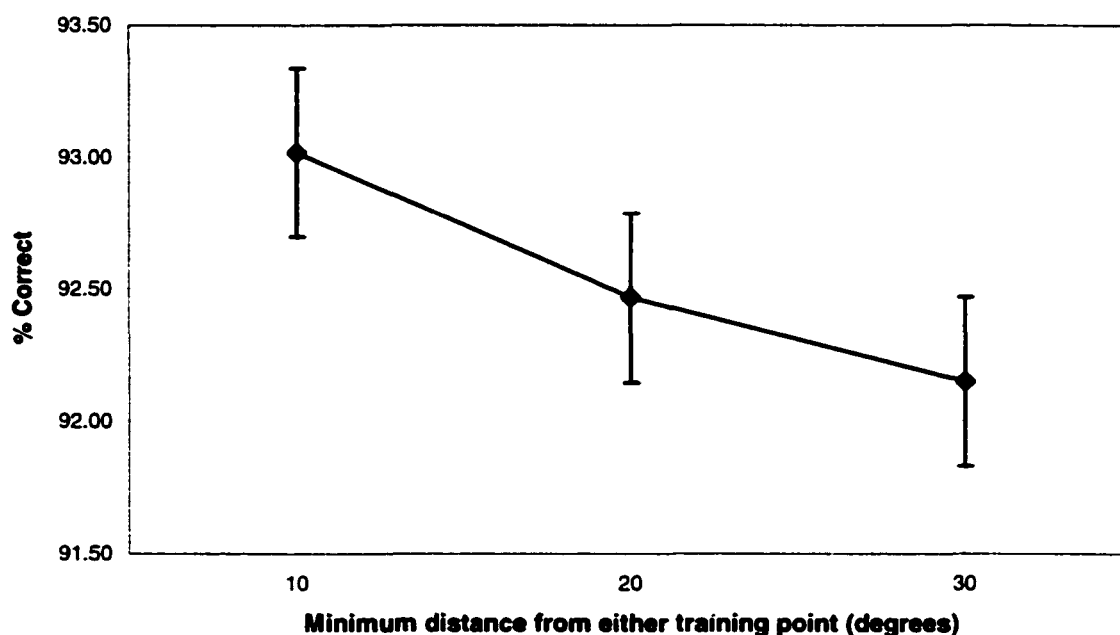


Figure 19. Recognition performance as lighting deviated from either training point in Experiment 1. The distances are averaged over all three of the lighting conditions (INTER, EXTRA, and ORTHO).

When measuring the minimum distance of the lighting direction in the test trials from the near-frontal (10°, 10°) training illumination, the simple effect of

distance on performance was only seen in the lighting directions associated with the ORTHO condition (10°, 20°, 30°), $F(2, 34) = 3.628$, $p < 0.05$, $\epsilon = 0.954$. A significant linear trend also was found across the three lighting directions for the ORTHO condition, $F(1, 17) = 6.772$, $p < 0.05$. This trend suggested that the decrease in performance with increasing eccentricity from the trained illuminations might also exist in the other conditions (i.e., INTER and EXTRA) if a more powerful manipulation were used. While there is a trend towards decreasing performance with increasing eccentricity from training in the lighting directions of the EXTRA testing condition, this trend did not reach statistical significance.

When the test illumination directions are averaged over the test conditions, five lighting directions emerge: 10°, 20°, 30°, 50°, and 60°. Figure 20 shows how these lighting deviations affect recognition of the images. Performing a one-way within-subjects ANOVA, the effect of increasing the distance of the test lighting direction from the (10°, 10°) trained lighting direction on recognition performance was significant, $F(4, 68) = 3.501$, $p < 0.05$, $\epsilon = 0.657$. The linear trend associated with these lighting directions also was significant, $F(1, 17) = 7.319$, $p < 0.05$. Again, as was shown when the illumination angles for deviation from either training point were analyzed together, as the illumination angle increased from the near-frontal trained lighting direction, recognition performance for the faces decreased.

Performing the analysis on recognition performance for deviations in lighting direction either from *both* of the trained illumination directions or from the *near-*

frontal (10°, 10°) trained lighting direction seemed logical. Measuring from *either* trained illumination allowed for inferring how well observers were able to build a representation of the lighting model for the observed faces. Likewise, measuring from the near-frontal (10°, 10°) trained lighting direction also incorporated the integration of information with regards to lighting between the two training views (how could it not?) but with the additional parameter of using a usual view of lighting on the face (near-frontal) as the baseline for the measurement.

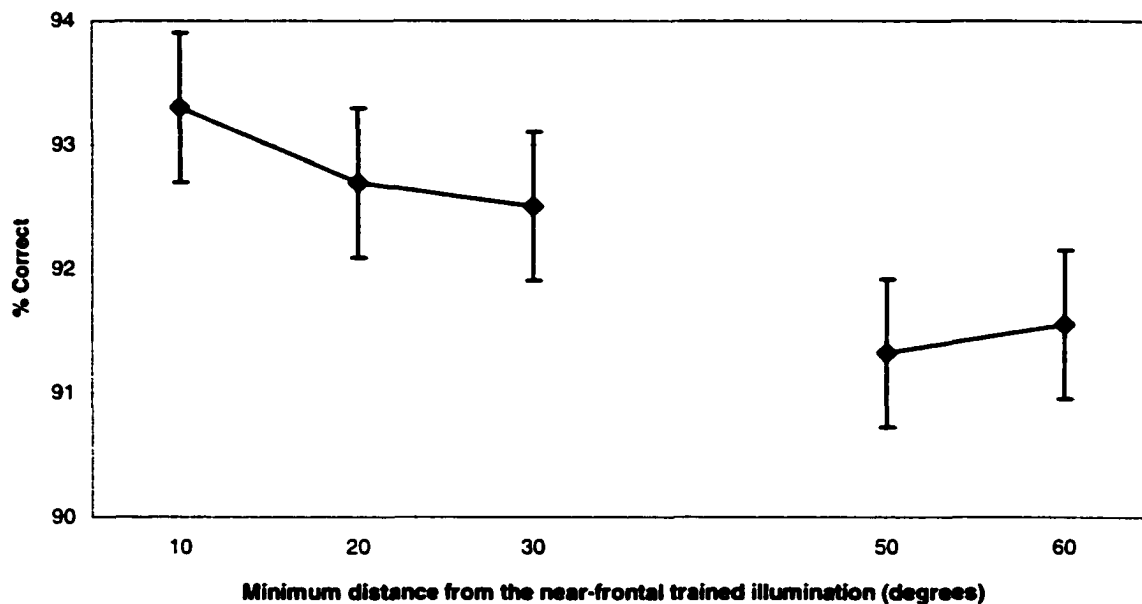


Figure 20. Recognition performance as lighting deviated from the near-frontal (10°, 10°) training point in Experiment 1. The distances are grouped over all three of the lighting conditions (INTER, EXTRA, ORTHO).

The near-frontal (10°, 10°) trained lighting direction was the better measure of performance due to the larger range of distances that this measurement provided by increasing the distance between the given training condition and all

test lighting directions. Results from the experiments presented in Chapter 2 showed that observers who viewed near-frontal lighting on faces during training performed similarly to performance on the trained lighting directions, and to images with lighting that ranged from 15° to 30° in distance from the trained illuminations. For distances greater than these, performance decreased dramatically. Using the near-frontal training point as the baseline for measuring test distance is also most similar to the technique employed by Bülthoff and Edelman (1992) in their study of viewpoint dependence, whose methods the present study is emulating in terms of lighting variation over objects. For completeness, lighting distances were also averaged with respect to the training condition with extreme lighting (50° , 10°). Using this metric, subject performance actually increased with increasing distance from the trained lighting direction. This result was probably due to the increase in lighting direction with respect to the trained direction (50° , 10°), i.e., the lighting in the images was actually approaching near-frontal lighting directions. This result is consistent with the results from Chapter 2, and other research, suggesting that the recognition of faces is easier with lighting that is typical (e.g., Johnston, Hill, & Carman, 1992; Tarr, Kersten, & Bülthoff, 1998). Since faces often are seen with frontal illumination, any lighting model of faces should include a robust representation of this illumination condition. The remaining experiments will use the near-frontal (10° , 10°) training point as the baseline for measuring recognition performance for the test stimuli.

Because the responses of all of the subjects were close to ceiling, the significant differences found in several of the analyses may be due to artificially small variance. We believe that the effects between conditions are real and that with a more sensitive measure, the effects would become more pronounced.

Experiment 2

The results of the previous experiment show that the face recognition performance of observers is better when the test lighting directions are between the two trained lighting directions. These results suggest that interpolation between the trained illumination directions is a possible explanation for the good recognition performance of subjects in the INTER condition as compared to their performance in the EXTRA condition. If subjects were building a complete three-dimensional representation of the scene, i.e., the shape and reflectance of the face, and of the lighting parameters of the scene, i.e., encoding the image variability accounted for by changes in the lighting of the scene, performance should have been the same across the three illumination conditions. The results from the previous chapter suggest that observers may be able to encode the illumination variations across a scene (*not* including the strength or direction of the lighting, but rather the changes in the appearance of the image as the lighting in the scene changed), but may not accurately derive the shape and reflectance properties of the objects. The results of the previous experiment also suggest that the lighting directions that were orthogonal to the axis defined by the trained illumination conditions provided sufficient information

for highly accurate recognition of the faces shown. While these results seem unusual when compared to the results of the Bülhoff and Edelman (1992) viewpoint-interpolation studies (in which the views in the ORTHO condition were the most difficult to recognize), viewing illuminations oriented along the vertical axes of faces is not unusual for observers in the real world.

The current experiment attempted to investigate the effects of inverting the faces in the training and test images on recognition. The images were inverted such that the faces were still turned in the same direction along the horizontal axis as in the previous experiment, but upside-down. Past studies have shown that absent of any cues to illumination on a face, recognizing inverted faces, even well-known faces, is more difficult than recognizing the same faces upright (Hochberg & Galper, 1967; Yin, 1969, 1970; Valentine, 1988). This suggested that inverting the face somehow affected the ability of a subject to extract the necessary geometric information about the features on the face to accurately identify the person. By inverting the faces and the lighting on the faces, the results of this experiment should indicate whether changing the geometry of the faces changes how the illumination algorithms use interpolation to achieve adequate recognition of the faces.

Methods

This experiment was identical to Experiment 1 with the exception that the stimuli were all inverted so that the faces were upside-down, i.e., eyes below mouth, but not rotated. This resulted in the faces looking in the same direction as in the images in Experiment 1. The lighting on the faces followed the

orientation of the faces, so that a face that was lit from above normally was lit from below when inverted and vice versa. The training lighting directions again were situated at $(10^\circ, 10^\circ)$ and $(50^\circ, 10^\circ)$ and 14 lighting directions comprised the testing stimuli, as in Experiment 1.

Results and Discussion

The results from Experiment 1 suggested using the near-frontal $(10^\circ, 10^\circ)$ training point as the best baseline for measuring the minimum Euclidean distance of the test lighting directions from training to measure recognition performance. Again, the dependent variable was the mean percent correct recognition for each illumination direction on the lighting sphere. This measure was calculated as the ratio of correct recognition to the total number of test trials for each of the 14 lighting directions. For instance, for images with lighting at $(30^\circ, 10^\circ)$, the distance from the near-frontal training point was 20° . The mean percent correct recognition for a particular lighting direction was the total number of correct recognition trials for that illumination divided by the total number of test trials. As well as calculating the mean percent correct recognition for distances deviating from the near-frontal training point, analyses also were performed on the responses grouped according to the three lighting conditions as outlined before: INTER, EXTRA, ORTHO. To reiterate what test lighting directions comprised these conditions: 1) the INTER condition consisted of those test lighting directions between the two training points; test illumination directions outside of the training points were in the EXTRA condition; 3) the ORTHO condition contained lighting angles orthogonal to the lighting axis

defined by the two training points; in this case, an axis vertical to the 10° horizontal axis. Since, all subjects responded to more than 99% of the total trials, no subjects were used in the analyses. As in the previous experiment, for analyses with more than one degree-of-freedom in the numerator, the Greenhouse-Geisser (1959) correction was used and is denoted by reporting the epsilon (ϵ) correction value with the F statistic.

The first result that can be inferred from Figure 21 is that the overall mean recognition performance (83.43%) is down from that of Experiment 1 (92.67%). One explanation for this overall decrease in recognition performance is the “inverted face” effect. This principle stems from several studies that found that upside-down faces were more difficult to recognize than upright faces (Hochberg & Galper, 1967; Yin, 1969, 1970; Valentine, 1988). However, these results did not take into account the possible interaction of lighting on the inverted faces. Several recent studies have shown that lighting direction (above or below the face) can interact with the effects of face inversion in recognition or classification tasks (Johnston, Hill, & Carmen, 1992; Enns & Shore, 1997). However, this interaction does not occur when the task is specific identification of a face, i.e., naming the given face using one of several choices (Enns & Shore, 1997).

As Figure 21 illustrates, recognition performance in the EXTRA lighting condition was significantly worse than in the other two lighting conditions ($F(2, 46) = 3.88, p < 0.05, \epsilon = 0.74$). The figure also shows, supported by subsequent analysis, that observers' recognition performance in the INTER and ORTHO

conditions were similar ($F(1, 23) = 0.13, ns$). Since no difference between the INTER and ORTHO lighting conditions in Experiment 1 was observed, no differences were anticipated in this experiment since vertically-oriented lighting did not seem to be an unusual condition on human faces.

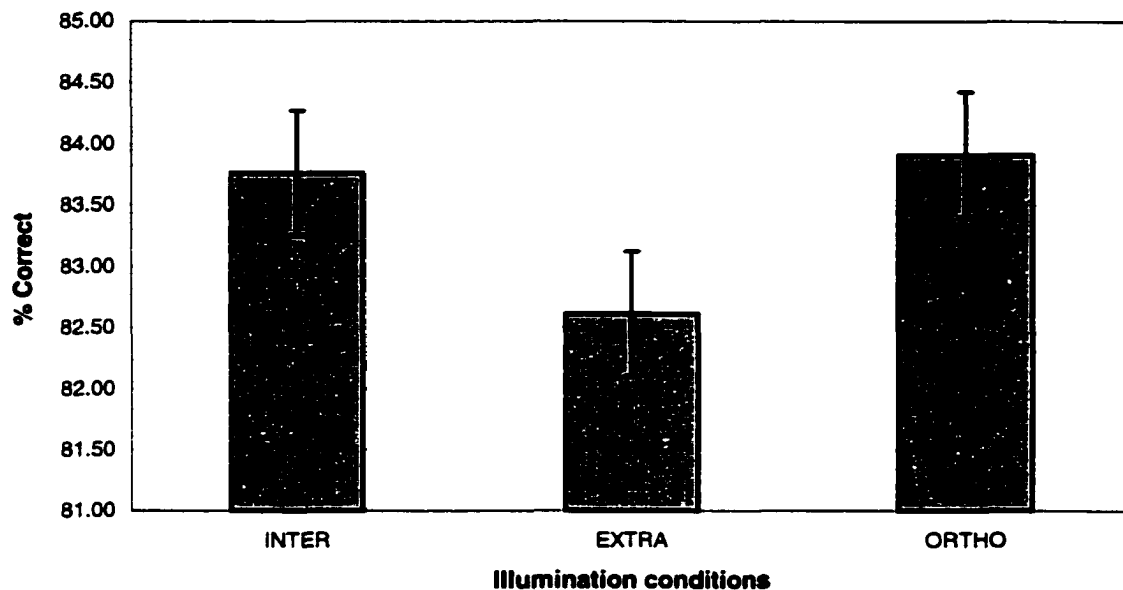


Figure 21. The effect of illumination condition on recognition performance in Experiment 2. As observed in Experiment 1, the EXTRA condition is significantly different than the other two illumination conditions ($F(2, 46) = 3.88, p < 0.05, \epsilon = 0.74$). Error bars represent the within-subject standard error of the mean.

Subsequent to the results of previous studies (Johnston et al., 1992; Enns & Shore, 1997), differences within the ORTHO condition (i.e., between faces that were chin-lit and faces that were brow-lit) were expected. Since the lighting followed the inversion of the faces, images containing lighting from below in Experiment 1 contained chin-lit lighting in this experiment; likewise, lighting from above in Experiment 1 is referred to as brow-lit in this experiment.

While separating lighting above and below the face did not result in any significant differences in recognition performance in Experiment 1, differences between these groups were expected in this experiment given previous results in other studies. The differences in performance were analyzed between brow-lit and chin-lit faces using a one-way within-subjects factorial design, however, no differences in recognition performance were found between the two orientations of lighting ($F = 0.42$, *ns*). While no differences were found between chin-lit and brow-lit faces, there was still a trend within each condition for decreasing performance with increasing eccentricity from the near-frontal training point (see Figure 22).

The results of this experiment are not consistent with previous studies that looked at the effect of lighting on inverted faces; the results of this experiment do not show differential effects for faces that are chin-lit versus faces that are brow-lit. There are several possible reasons for the differences between this experiment and the studies that have found differences between both upright and inverted brow-lit and chin-lit faces. One argument is that the previous studies relied too heavily on the observers' memories (Johnston et al., 1992) of previously seen individuals, which were not controlled for by the experimenters. Also, while class-level knowledge for all faces has a bias for lighting from above (the reason you might get differences between lighting from above and below), training on specific illumination conditions might remove or override this class-based lighting bias. This influence of specific trained lighting directions suggests that the internal representations may be less categorical and more

specified for certain lighting directions. The bias for class-level knowledge also may be diminished through training on specific *faces*. The Illumination Cone (IC) model constructs a representation for each face on which the model is trained. By training on specific faces, the model removes any class knowledge bias that might interfere with the recognition of the desired faces, but this strength of the model prevents any generalization to other instances within the class, i.e., faces that were not seen during training. Humans do not suffer from this weakness, while they still might use mechanisms that are similar to the IC model.

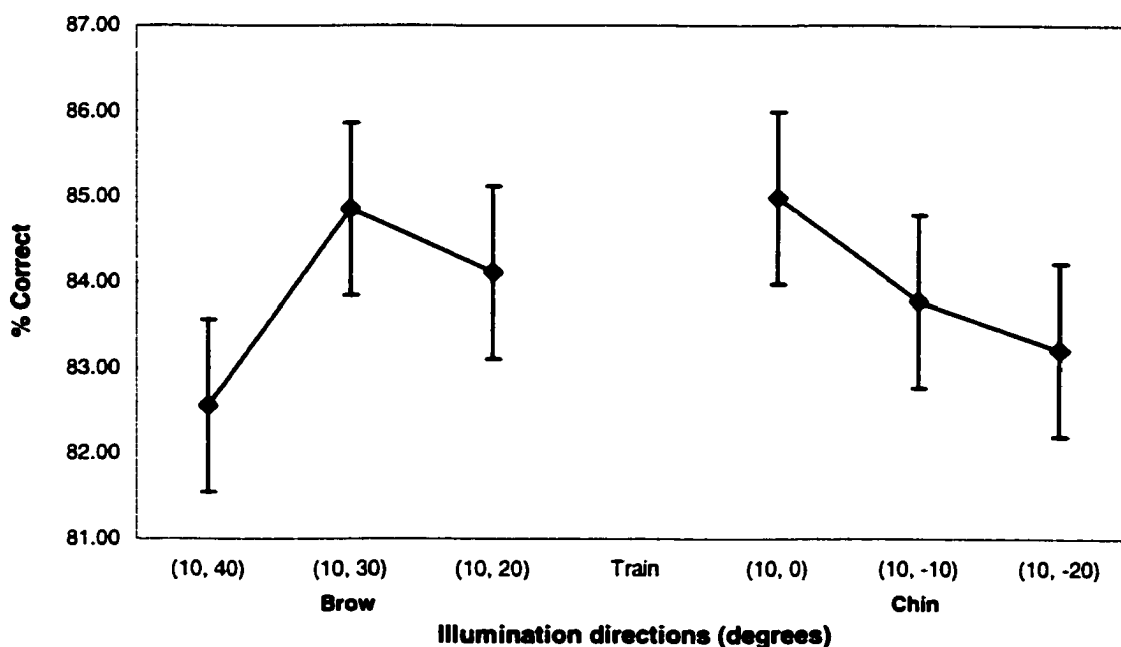


Figure 22. Recognition performance as a function of the lighting directions in the ORTHO condition in Experiment 2. Lighting coordinates to the right of the graph represent lighting directions that light the brow of the face while those to the left represent lighting directions that light the chin of the face. The error bars represent the within-subject standard error of the mean.

As stated previously, the minimum distance of each test lighting direction from the near-frontal (10°, 10°) training point was calculated. This distance was

used to determine mean percent correct recognition performance. Using the near-frontal training point as the baseline for measuring lighting distance from training resulted in five distance “bins”: 10°, 20°, 30°, 50°, 60° (no 40° bin exists, as it would contain the other training point).

The simple effect of lighting distance on correct recognition is only significant between the five lighting directions associated with the EXTRA condition ($F(4, 92) = 4.717, p < 0.01, \epsilon = 0.626$). A significant linear trend also exists between these lighting directions ($F(1, 23) = 12.263, p < 0.01$). While there is a trend towards decreasing performance with increasing eccentricity from training in the lighting directions of the INTER and ORTHO illumination conditions, these trends do not reach statistical significance.

Averaging over the three test conditions (INTER, EXTRA, and ORTHO), five lighting directions emerge: 10°, 20°, 30°, 50°, and 60°. Figure 23 shows how these lighting deviations affect recognition of the faces. As the figure illustrates, increasing the distance of the test lighting direction from the (10°, 10°) trained lighting direction caused a significant decrease in observers' recognition performance ($F(4, 92) = 6.391, p < 0.01, \epsilon = 0.562$). Figure 23 also shows a significant linear trend associated with these lighting directions ($F(1, 23) = 11.943, p < 0.01$). Again, as was shown in Experiment 1, as the distance of the illumination in the test images increased from the near-frontal trained lighting direction, recognition performance for the faces suffered.

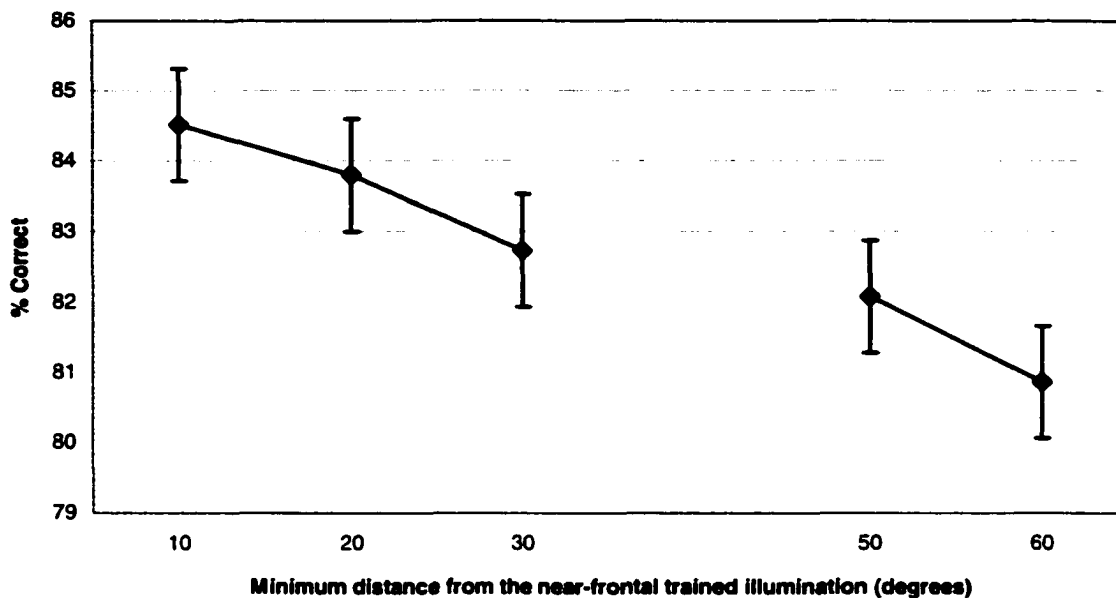


Figure 23. Recognition performance as lighting deviated from the near-frontal (10°, 10°) training point in Experiment 2. The distances are grouped over all three of the lighting conditions.

Experiment 3

In order to reduce the influence of any class-based knowledge on the part of the observers when viewing lighting directions in the ORTHO condition, the lighting axes were flipped. That is, instead of training on the horizontal axis of the lighting sphere and testing the ORTHO condition on the vertical axis, these conditions were switched. Since the results in the previous two experiments suggested that the good recognition performance in the ORTHO condition was probably due to the influence of class-level knowledge of lighting on the vertical axis of the face (as typically seen in the real world since lighting is mostly above the head), reduced performance in the ORTHO condition was expected when switching the axis of training to the vertical axis; the ORTHO condition would

coincide with the horizontal axis across the face. Since the previous experiments also suggested that extrapolating to lighting directions outside of the trained illumination directions was difficult, poor recognition performance in the EXTRA lighting condition was again expected.

Methods

In this experiment, the orientation of the lighting was switched. This meant that the trained lighting directions were oriented vertically with respect to the lighting sphere, where in the past two experiments, the axis of training was horizontal relative to the lighting sphere. In this case, the two training illumination points were (10° , 30°) and (10° , -10°). Figure 24 illustrates the positions of the training points and the three lighting conditions.

As shown in Figure 24, training along a vertical axis flips the three groups of lighting conditions. For this experiment, the lighting directions in the INTER condition still were positioned between the two training points, but now along the 10° vertical axis on the lighting sphere. As well, the illuminations in the EXTRA condition were oriented vertically. Lighting directions in the ORTHO condition, corresponding to lighting directions on an axis orthogonal to the axis of training, were moved to the 10° horizontal axis of the lighting sphere.

Unlike the ORTHO condition in Experiments 1 and 2, the lighting directions in the ORTHO condition in this experiment were *not* on the same axis as one of the trained lighting directions. Instead, due to the limitations of the stimulus set, the ORTHO condition in this experiment contained lighting directions in the INTER condition of Experiments 1 and 2. In other words, the illumination angles

in the ORTHO condition laid on the horizontal meridian 10° above the equator of the lighting sphere. This meant that the axis that the lighting directions in the ORTHO condition were on was 20° (minimally) from either of the training points.

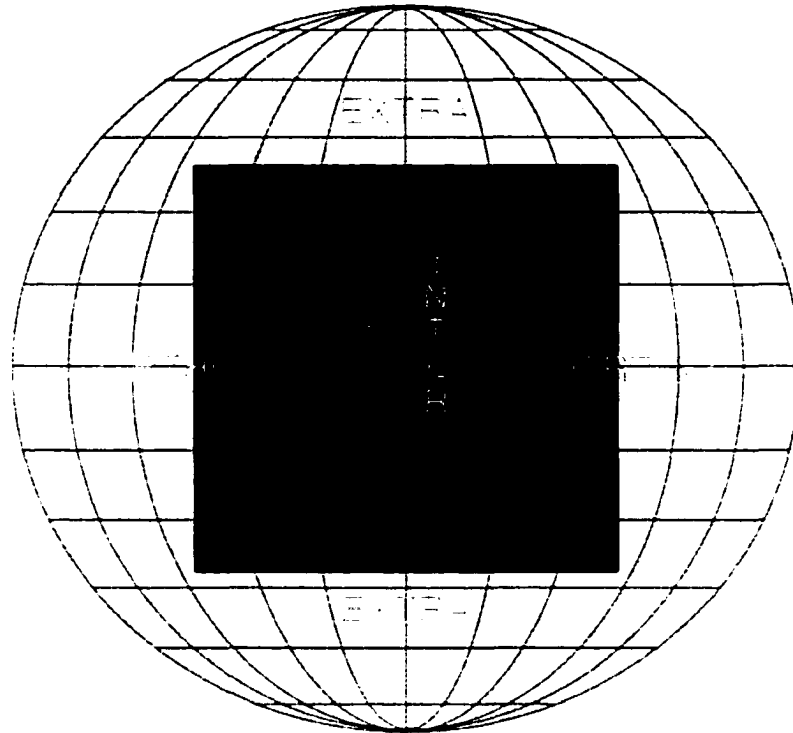


Figure 24. The lighting sphere used in Experiment 3 showing the relative positions of the testing conditions with respect to the trained illumination directions. The dots indicate the trained lighting directions. The conditions in which the test illumination directions were grouped are shown: 1) lighting between the two training conditions (INTER); 2) lighting outside of the trained lighting directions (EXTRA); 3) lighting directions orthogonal to the axis defined by the training illuminations (ORTHO).

Results and Discussion

Again, the lighting coordinates for each image were recorded and the minimum Euclidean distance from the near-frontal trained illumination condition to that coordinate was computed. In this experiment, the near-frontal training point was located at $(10^\circ, -10^\circ)$ on the lighting sphere. The dependent variable

was again mean percent correct recognition calculated as the ratio of correct identifications to the total number of test trials for a particular individual trained face. Although test trials were timed-out after waiting for a response for three seconds, subjects responded to more than 96.8% of the total trials.

As in the last two experiments, several one-way within-subject ANOVAs were analyzed, instead of the two-way analysis of variance, due to an unequal number of levels in the conditions. Where necessary in analyses with more than one degree-of-freedom in the numerator, the Greenhouse-Geisser (1959) correction was used and is denoted by reporting the epsilon (ϵ) correction value with the F statistic.

As Figure 25 illustrates, recognition performance between the three lighting conditions showed significant differences ($F(2, 36) = 4.474, p < 0.05, \epsilon = 0.906$). However, the pattern of results was unlike the patterns seen in the two previous experiments. The observers' recognition performance in each of the last two experiments was similar between the INTER and ORTHO lighting conditions. Also, as previously seen in the other experiments, there was a significant difference between the INTER and EXTRA lighting conditions, which was expected, given the ease with which observers had previously used lighting interpolation ($F(1, 18) = 7.453, p < 0.05$). As shown in Figure 25, recognition performance in this experiment was different between the INTER and ORTHO conditions ($F(1, 18) = 6.336, p < 0.05$). The results of the last two experiments also displayed differences between the EXTRA and ORTHO lighting conditions. With a decrease in recognition performance, in this experiment, with lighting

directions in the ORTHO condition, observers' performance in recognizing faces with lighting conditions from these two lighting conditions was the same ($F(1, 18) = 0.577, ns$).

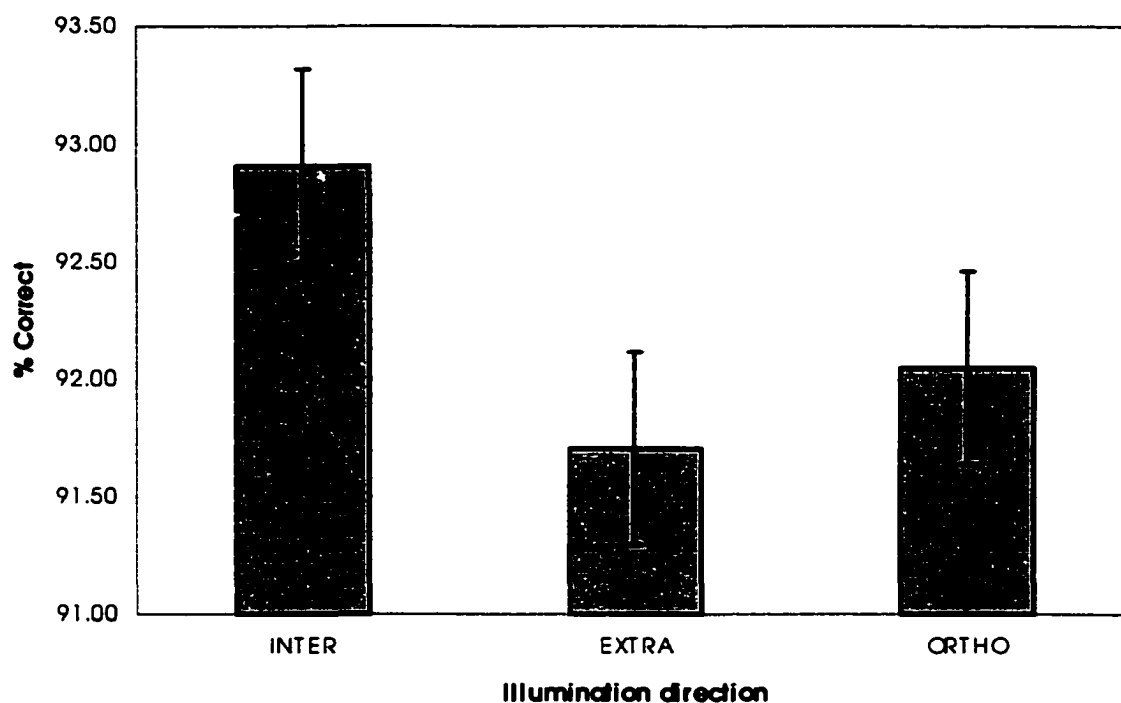


Figure 25. The effect of illumination condition on recognition performance in Experiment 3. Recognition performance in the INTER condition was significantly different from performance in the other two illumination conditions. Error bars represent the within-subject standard error of the mean.

There are two potential reasons for a decrease in performance in the ORTHO condition in this experiment, although performance was equal to the INTER condition in the last two experiments: 1) it was harder to extrapolate the effects of lighting on the geometry of the faces when the lighting directions were not coincidental to one of the trained illumination points; 2) the representation of lighting on the faces was not as well defined with training in the vertical

dimension as it was with training in the horizontal dimension, i.e., the lighting representation built using the trained vertical lighting directions was poor.

Since the EXTRA condition in this experiment was the ORTHO condition in the previous two experiments, one might expect that recognition performance in this condition might be pretty good, considering that performance in this condition was good in the last two experiments. However, that level of recognition performance was seen with trained lighting directions on the horizontal axis. If observers were merely using class-level knowledge about the role of lighting on faces in this experiment, then one might not expect a drop-off in recognition performance. The decrease in performance here suggests that observers used the information provided by the trained lighting directions to construct a representation of the faces, instead of using any class-level knowledge about faces. The fact that the recognition performance was still high overall suggests that observers were able to fall back on some class-level knowledge of lighting on faces when the representation provided by the training conditions failed to provide adequate information for recognition. However, while this class-level knowledge seemed to help subjects in the last two experiments, it seems to be failing them in this context.

There are several possibilities for this performance decrease in the ORTHO condition in this experiment. It might be that recognition using lighting outside of the training points is always bad because observers have a difficult time extrapolating from one lighting condition to others, but an easy time interpolating between two illumination directions. This would account for the

poor recognition performance in the EXTRA and ORTHO conditions. However, the EXTRA condition contained illuminations that were oriented along the vertical axis of the face. By previous accounts concerning the use of class-level information, and the typicality of vertical illumination conditions on the human face, these lighting directions should not have caused as much difficulty with the subsequent recognition task. In fact, one would think that training with lighting on the vertical axis of the faces would have enhanced the information already present concerning vertically oriented lighting; however, it did not. Why? In actuality, the performance for the two lighting directions 10° from either training point improved from the performance for those same lighting directions in Experiment 1. The recognition performance for lighting at $(10^\circ, 40^\circ)$ went from 91.9% in Experiment 1 to 92.16% in this experiment. Likewise, performance for the $(10^\circ, -20^\circ)$ illumination point improved from 92.77% in Experiment 1 to 93.2% in this experiment. The worst performance in the EXTRA condition comes from images with lighting at $(10^\circ, 50^\circ)$ and $(10^\circ, -30^\circ)$. These extreme lighting directions create severe shadows on the faces in the images, require extrapolation from only one reference point (one of the two training illuminations), and are quite distant from both of the training points. All of these factors combined most likely prevented observers from correctly recognizing the faces under those lighting directions.

Even when the EXTRA condition was broken into two groups: lighting above the face (brow-lit) and lighting below the face (chin-lit), there was a difference between the two groups ($F = 6.488, p < 0.05$). In fact, the recognition

performance was *better* for the group with lighting directions below the face than for lighting directions above the face (see Figure 26). This is totally counter to any previous results that have found differences between the two groups (e.g., Johnston et al., 1992; Enns & Shore, 1997). The reason probably lies in the specificity of the representation of the faces created by the observers using the lighting conditions viewed during training, paired with the prior belief in lighting on faces being primarily frontal. The (10° , -10°) training lighting direction was only 10° below the point with frontal lighting relative to the faces. As the previous experiments in this study have shown, a 20° deviation from a near-frontal trained lighting direction usually was not sufficient to produce significant errors in recognition performance. This suggests that the chin-lit lighting directions in the EXTRA condition in this experiment should not have impaired recognition performance to a high degree. However, the brow-lit lighting directions in the EXTRA condition in this experiment were at least 50° away from the near-frontal trained illumination point, and therefore we expected more errors when trying to identify faces with these illumination conditions present. Indeed, these lighting directions did hamper correct recognition of the faces and, as stated, a significant difference between these lighting directions and the EXTRA "chin-lit" lighting directions was found.

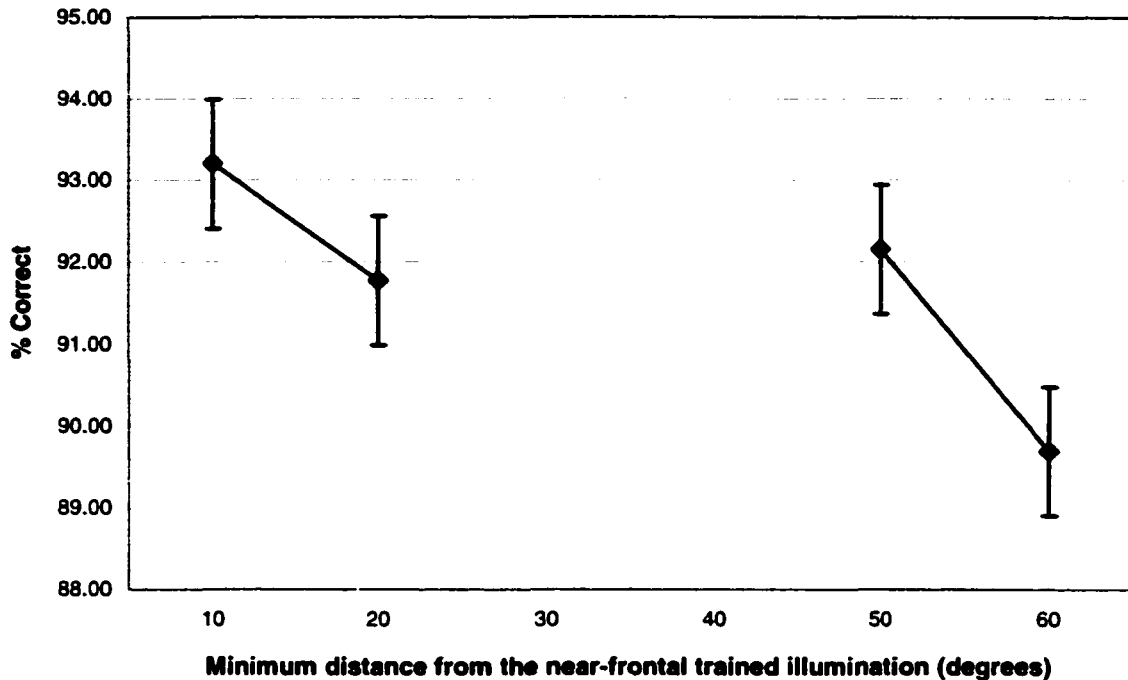


Figure 26. Recognition performance for the lighting directions in the EXTRA condition in Experiment 3. The distances were measured from the near-frontal (10°, -10°) training point. The images represented by the 10° and 20° distances contained faces with chin-lit illumination and the images represented by the 50° and 60° distances contained faces with brow-lit lighting.

When the EXTRA condition was analyzed purely in terms of deviations from the near-frontal training point, a significant decrease in recognition performance with increasing eccentricity from the training point was found ($F = 6.942$, $p < 0.01$, $\epsilon = 0.843$). The other two lighting conditions did not show any significant reductions in recognition performance with increasing distance from the near-frontal training point, as Figure 27 illustrates.

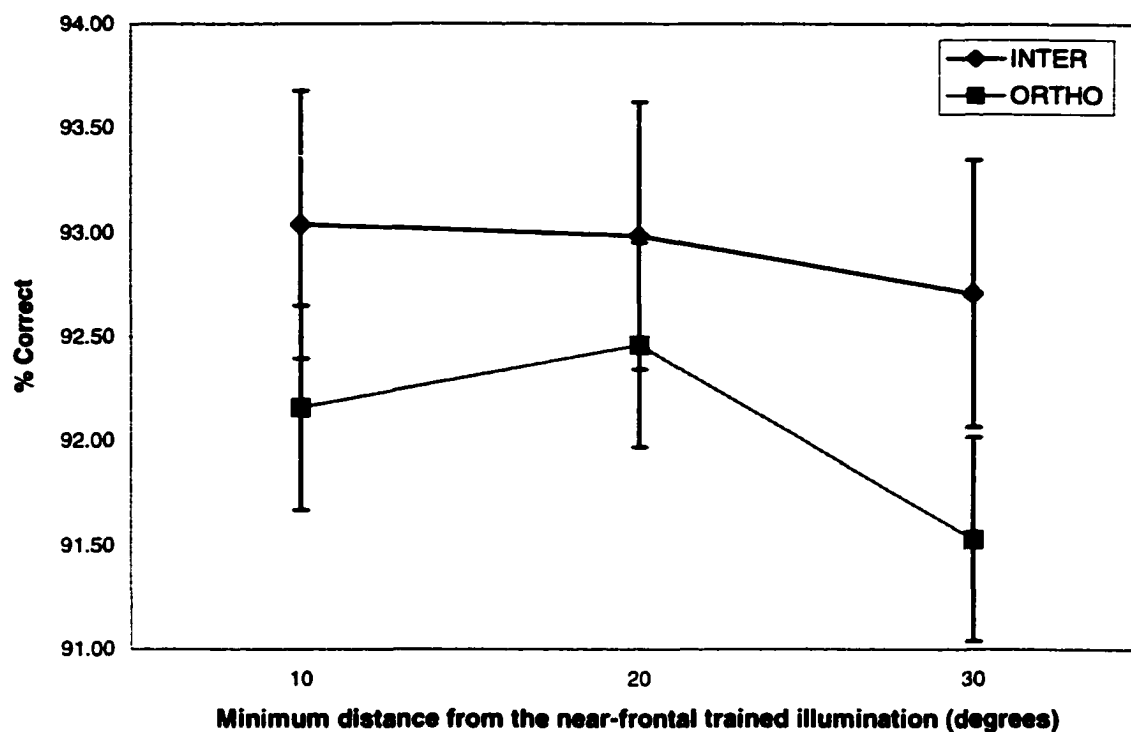


Figure 27. Recognition performance for the lighting directions in the INTER and ORTHO conditions in Experiment 3. The distances are measured from the near-frontal (10° , -10°) training point. The error bars represent the within-subject standard error of the mean.

The distance measurements were also collapsed over the three illumination conditions from the near-frontal (10° , -10°) training point to get a sense of the overall effect of eccentricity from frontal illumination (see Figure 28). When the results were grouped in this way, an effect of distance from training on recognition performance was seen ($F = 7.612$, $p < 0.01$, $\epsilon = 0.645$). However, as the figure also shows, all of the effect was contained in the 60° distance from training. This distance included only one lighting direction, (10° , 50°), which was an extreme illumination in the ORTHO lighting condition (illumination from above the face). Observers exhibited the worst recognition performance while viewing images with this particular illumination direction, with a mean correct

recognition rate of 89.69%. In fact, this mean recognition performance was 1.1% worse than the performance in the next worst lighting direction, (-20° , 10°), which was an extreme illumination in the EXTRA condition.

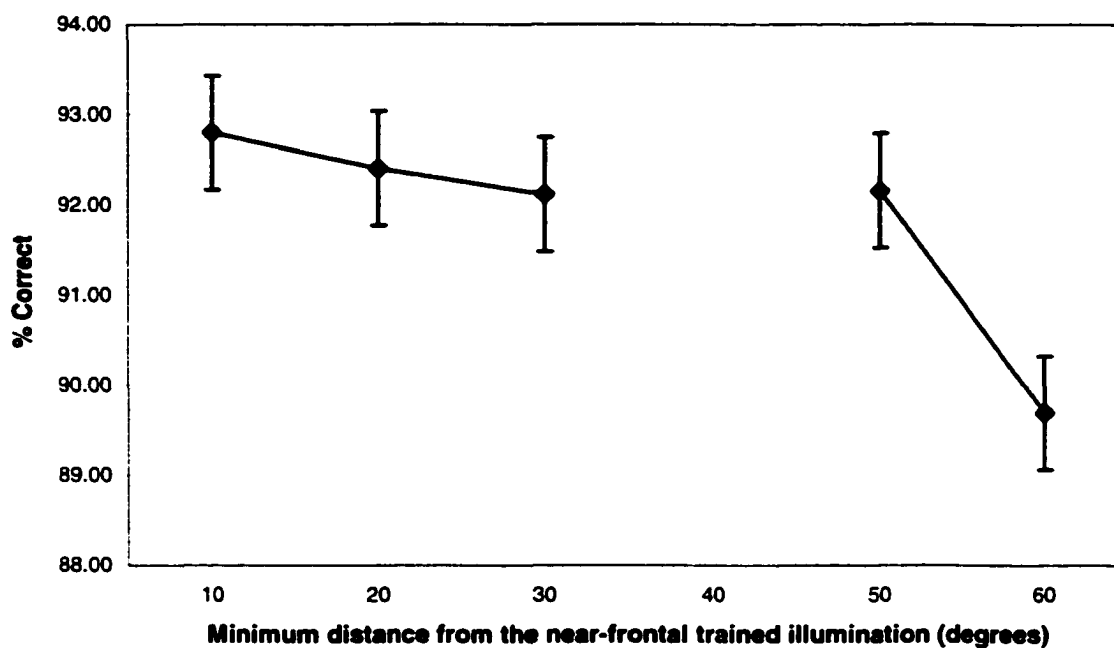


Figure 28. Recognition performance for the lighting directions in the INTER and ORTHO conditions in Experiment 3. The distances are measured from the near-frontal (10° , -10°) training point. The error bars represent the within-subject standard error of the mean.

While the analyses using the near-frontal (10° , -10°) training point as a baseline was considered more informative as to the nature of observers' performance with respect to lighting in the faces, overall performance as measured from *either* training point, (10° , -10°) and (10° , 30°), was also analyzed. As Figure 29 shows, there was no significant difference in recognition performance with increasing eccentricity from the trained lighting directions. However, a linear trend did exist for the decrease in recognition performance as

the distance from either training point increased. This result provides further evidence for the use of a specific representation of the lighting on the faces based on the training observers received during the experiment. If a lighting model was not represented as part of the knowledge base of the faces, or if observers merely were utilizing a class-based lighting model based on previous experience recognizing faces, we would not expect to see differential performance as the lighting varied over the faces. The fact that a pattern exists for decreasing performance that is consistent with the previous studies in Chapter 2 is sufficient.

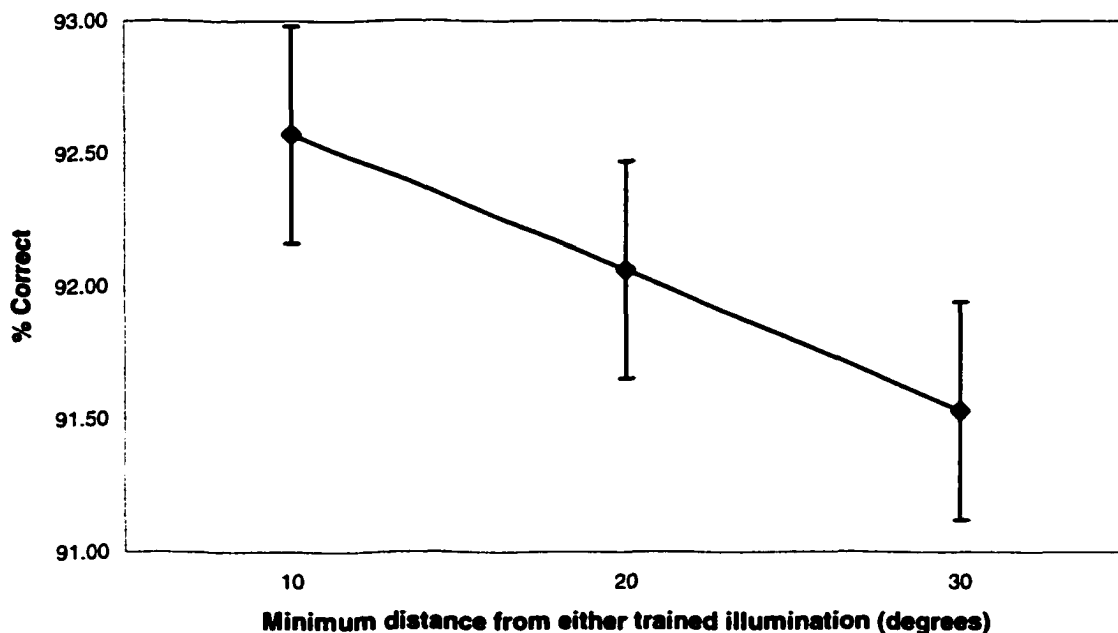


Figure 29. Recognition performance as a function of distance from either of the trained lighting directions in Experiment 3. A significant linear trend exists between the distances ($F = 5.231$, $p < 0.05$). The error bars represent the within-subject standard error of the mean.

Discussion

The results from the experiments presented in Chapter 2 indicate that the human visual system has some sort of image-based representation for modeling lighting on faces. Furthermore, by using an identical methodology in testing human observers and the Illumination Cones (IC) computer vision model (Belhumeur & Kriegman, 1998), the results point to a possible algorithm for this lighting representation. Also, since the IC model predicts better recognition performance for unknown lighting directions surrounded by known lighting directions, the model exhibits behavior that is similar to models that use two-dimensional view interpolation to compensate for changes in object orientation (e.g., Poggio and Edelman, 1990). Given the similarity between the interpolation models for object viewpoint and the behavior of the Illumination Cones model, with respect to lighting directions near known illumination conditions, we expected that interpolation between known lighting conditions would hold as it does for viewpoint.

Using a methodology similar to that used by Bülthoff and Edelman (1992) in their study of viewpoint in object representations, we performed three experiments to investigate the role of interpolation in the representation of lighting direction in the human visual system. This set of experiments had two goals:

- **Examine the specifics of how a model of illumination might function given the need to use interpolation in order to achieve adequate object recognition;**

- Determine whether object geometry and its relation to lighting direction affected how the illumination model is computed.

All of the experiments presented in this study showed results similar to those presented in Chapter 2. As the direction of lighting in the test images of the faces became more disparate from the lighting directions used during training (and, presumably, used to build a lighting model for the faces), performance for recognizing the individual faces decreased. This decrease in performance was more dramatic in the experiments of Chapter 2, but this is most likely due to the degree of variability in the lighting directions present in the stimuli of that set of experiments.

The results of this chapter suggest that including lighting parameters in high-level representations of faces is necessary to derive shape information and to constraint otherwise ambiguous scene information, such as is present in scenes with extreme lighting conditions. There are more comparisons than just those made with the results of the experiments of the previous chapter. Grouping the results of the present experiments, as Bülthoff and Edelman (1992) did in their study of object viewpoint and interpolation models, provides clues as to the use of interpolation in the algorithms used by the visual system to deal with variations in illumination conditions.

The results from Experiment 1 suggest that the visual system does seem to use interpolation between known illumination conditions in order to provide adequate face recognition. However, these results also show that, unlike the results dealing with view interpolation, lighting directions orthogonal to the axis

of training provide enough information for good face recognition. This result is inconsistent with a system that only uses illumination interpolation for performing object recognition with varying lighting conditions. One suggestion for this result is that the familiarity with faces by the observers allowed them to use class-level knowledge concerning the interactions of vertical illuminants on faces to adequately perform the recognition tasks. This does suggest that a hierarchy of processing may exist with respect to lighting representations. When possible, the visual system will probably use specific knowledge of the faces to perform necessary recognition tasks. However, if the information derived from the image does not permit such specificity, then class-level knowledge will probably be used to try to compensate for the lack of information.

The results from Experiment 2 build on this idea of specificity first, class-level knowledge second, in terms of processing the image information through any high-level object representations. Inverting the images of the faces did not change the pattern of results from Experiment 1. The only difference between the two studies was due to the inverted-face effect, i.e., the overall recognition performance across all of the lighting conditions was diminished due to the inherent difficulty in identifying upside-down faces.

Changing the pattern of training on the lighting directions in Experiment 3 provided results different than those found in the other two experiments. The EXTRA lighting condition, as predicted, led to the worst recognition performance by the observers. However, instead of the observers showing

similar recognition performance for the INTER and ORTHO lighting conditions, the performance in the ORTHO condition was drastically reduced; in fact, the ORTHO and EXTRA conditions showed the same recognition performance.

The pattern of recognition performance shown in Experiment 3 is more similar to the pattern of results postulated by the view interpolation results of Bülthoff and Edelman (1992), and the results of Chapter 1, which suggest that the Illumination Cones (IC) model uses an algorithm similar to an interpolation mechanism. A possible reason for this pattern of results in Experiment 3, where the INTER condition exhibits better performance than the other two illumination conditions, is that the lighting directions used for training, on the vertical axis of the lighting sphere, did not allow observers to default to any expected class-level representation of lighting on the faces when they were unable to use the available information due to lighting.

CHAPTER 4

General Discussion

There is evidence for a model-based representation of illumination information in humans. Tarr, Kersten, and Bülthoff (1998) explored whether human object recognition was lighting invariant, in part, motivated by the result that cast shadows helped constrain the perceived three-dimensional layout of a scene (Kersten, Knill, Mamassian, & Bülthoff, 1996; Kersten, Mamassian, & Knill, 1997). They found that shadows are intrinsic to object representations – possibly because they provided information about the three-dimensional structure of an object or scene.

If object representations include the effects of lighting, then lighting context should influence recognition performance. For example, faces are a highly familiar object class typically seen with the lighting above and in front. Johnston, Hill, and Carman (1992) reported on the horror film effect that faces lit from below (unusual lighting) look very different than when lit from above. Braje et al. (1998) found that faces lit from one side were recognized more poorly when the light was moved to the opposite side, both with and without cast shadows; thus demonstrating a recognition advantage for a learned

lighting direction. This pattern of lighting dependence during recognition provides some constraints on the computational algorithms in human vision, and allows us to compare these results to those of current computer vision models.

Recent computer vision models rely on two-dimensional image information – across multiple images – to represent illumination variability. Of three versions of this general approach, only the Illumination Cones approach exhibits good recognition performance for faces over a variety of illumination conditions. Belhumeur and Kriegman (1998) proposed constructing a hypersurface, or “cone,” which represented the set of potential images for one face under all possible point light sources; thus representing the entire lighting space for that particular face. Under this model, multiple illumination cones are needed to represent many unique faces for correct identification; this is also the model’s weakness, in that generalization within a class of objects is impossible without some prior knowledge about the geometry of the object class. As shown in the results of the studies in Chapter 2, the IC model is robust to training on a variety of illumination directions, and exhibits behavior that suggests that part of the algorithm is interpolating between trained lighting directions to achieve good face recognition.

Several experiments investigated how humans performed under conditions of varying illumination conditions. Observers were trained to recognize faces with lighting configurations designed to elicit different mental representations of lighting in each instance. In the set of studies discussed in Chapter 2, observers

were trained on various lighting directions that were either singular, close in proximity, or regularly spaced (as in the training set with lighting directions along the horizontal axis of the lighting sphere). These sets of trained lighting conditions ranged from frontal, or near-frontal, to lighting at the extreme sides of the faces. Across these different training conditions, the following results were obtained:

- Although the IC model exhibited higher accuracy than humans 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, lighting generalization fell off rapidly with distance from training images for human observers.
- 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.

Some of the above differences are inherent in the comparison made between the full vision system of humans and the 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 to the expectation that human observers should display both better generalizations 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.

The second set of experiments presented in Chapter 3 used two well-separated lighting directions on either side of, or above and below, the faces as training sets. The purpose of these experiments was to investigate whether human observers exhibited the same tendency to perform better face recognition when the viewing conditions included faces with lighting directions that were *between* the two trained illumination conditions, which would necessitate the interpolation of the known lighting directions. Since the IC model showed results that were consistent with a mechanism that might use

interpolation, the finding that human observers used lighting interpolation would provide further information concerning the mechanisms that humans use in a potential lighting representation. Across the three experiments, the following results were obtained:

- Human observers again exhibited decreased performance for lighting directions distant from the trained illumination conditions
- Lighting interpolation was exhibited by observers, but they also displayed the use of class-based knowledge for the effects of lighting on faces
- Observers seemed to use the specific learned lighting directions in their lighting representation, instead of a class-based representation
- When insufficient information was available in the image, observers fell back on their class-based lighting representations

The pattern of recognition performance shown in all three of the experiments suggests that the representation use by humans to compensate for changes in lighting across the scene might use an interpolation mechanism; one possible candidate for this model is the Illumination Cones (IC) model.

The results presented in Experiment 3 in Chapter 3 are the most similar to the pattern of results postulated by the view interpolation results of Bülthoff and Edelman (1992). A possible reason for this pattern, where the INTER condition exhibits better performance than the other two illumination conditions, is that the lighting directions used for training, in this case on the vertical axis of the

lighting sphere, did not allow observers to default to any expected class-level representation of lighting on the faces.

Of interest in this work is not simply the relative performance of human subjects and computational models, but what assumptions are made in order to make such comparisons and the evaluation of these comparisons. To provide the most useful comparisons between human subjects and the IC model, the experimental procedures used in the human psychophysical experiments closely mimicked in those used in the execution of the IC model. However, the IC model in no way implements the large bulk of what we think of as vision. Therefore its output is in many ways derived under entirely different conditions from the data obtained with humans, who necessarily bring their entire visual system into play in recognizing faces. Thus, the IC model may be at somewhat of a disadvantage, yet it performs as well as or better than the human observers under some conditions. In large part, this performance may be due to the fact that the IC model only “knows” about a small subset of all possible images. In contrast, humans are equipped with a lifetime of experience and knowledge of 100,000’s of objects. This apparent disadvantage also has positive implications for human observers. Specifically, most humans are face experts, and thus have class-level knowledge regarding the appearance of faces *in general*. This knowledge allows them to rapidly learn and recognize entirely novel faces, as well as generalize from a single view of a face to an entirely new lighting (or viewpoint) context – the idea being that other faces have been seen under the new conditions. In contrast, the IC model has no

knowledge of faces beyond the training it receives and, therefore, can never generalize between individual faces.

These same issues arise in nearly every extant computer vision model that is compared to human data. Nearly all address only a small part of the “vision problem”; in contrast, the human observer applies a complete vision system that includes filtering, sophisticated mid-level organization, and a rich representational space (and years of learning). It would be a mistake to claim that a given method does any more than model one specific mechanism of human vision. In the case of the IC model, that mechanism is generalizing 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 a larger vision system, but it may be the particular component that determines how performance modulates across lighting variation. Therefore, the *patterns* of generalization from known to unknown lighting conditions may be compared between the IC model and the human subjects. Similar comparisons are possible in many domains, so long as one is willing to make explicit the assumptions used in both the simulations with the computer vision model and the analogous psychophysical experiments. Indeed, it is argued that such comparisons ultimately improve both sides of the problem – refining the algorithms used in computer vision implementations and constraining the space of solutions for explaining elements of human vision.

The results of the experiments presented in Chapters 2 and 3 indicate that important future study should involve extending the present methods to entirely

novel object classes for which neither humans nor any computer vision model would have pre-existing knowledge since human subjects might have 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 a 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 computer 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., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, *20*, 38-64.
- 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.
- Enns, J. T., and Shore, D. I. (1997). Separate influences of orientation and lighting in the inverted-face effect. *Perception and Psychophysics*, *59*(1), 23-31.
- Etcoff, N. L., Freeman, R., and Cave, K. R. (1991). Can we lose memories of faces? Content specificity and awareness in a prosopagnosic. *Journal of Cognitive Neuroscience*, *3*(1), 25-41.
- 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.

- Georghiadis, 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.
- Greenhouse, S. W., and Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrics*, 24, 95-112.
- 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.
- Hietanen, J. K., Perrett, D. I., Oram, M. W., Benson, P. J., and Dittrich, W. H. (1992). The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex. *Experimental Brain Research*, 89, 157-171.
- Hochberg, J., and Galper, R. E. (1967). Recognition of faces: I. An exploratory study. *Psychonomic Science*, 9, 619-20.
- 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., & Bühlhoff, 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.
- Logothetis, N. K., and Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577-621.
- Lowe, D. G. (2000). Towards a computational model for object recognition in IT cortex. In S.-W. Lee, H. H. Bühlhoff & 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.
- Pentland, A. (1991). Photometric motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(9), 879-90.
- 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.

- Rock, I. and DiVita, J. (1987). A case of viewer-centered object recognition. *Cognitive Psychology*, 19, 280-293.
- Rolls, E. T. (1994). Brain mechanisms for invariant visual recognition and learning. *Behavioural Processes*, 33, 113-138.
- Rolls, E. T., Baylis, G. C., Hasselmo, M. E., and Nalwa, V. (1989). The effect of learning on the face selective responses of neurons in the cortex in the superior temporal sulcus of the monkey. *Experimental Brain Research*, 76, 153-164.
- 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.
- Shashua, A. (1997). On photometric issues in 3D visual recognition from a single 2D image. *International Journal of Computer Vision*, 21(1/2), 99-122.
- 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.
- Troje, N. F., & Siebeck, U. (1998). Illumination-induced apparent shift in orientation of human heads. *Perception*, 27(6), 671-80.
- 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. (1979). *The Interpretation of Visual Motion*. Cambridge, MA: MIT Press.
- Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, 32, 192-254.
- Ullman, S., and Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transcripts for Pattern and Analytical Machine Intelligence*, 13, 992-1005.
- 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.
- Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, 79, 471-91.

- Warrington, E. K. (1982). Neuropsychological studies of object recognition. *Philosophical Transcriptions of the Royal Society of London, B*, 298, 15-33.
- Warrington, E. K., and James, M. (1986). Visual object recognition in patients with right-hemisphere lesions: Axes or features. *Perception*, 15(3), 355-66.
- Williams, P. and Tarr, M. J. *RSVP: Experimental control software for MacOS* [Online]. Available: <http://www.tarrlab.org/RSVP/> [2001, September 4].
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81, 141-5.
- Yin, R. K. (1970). Face recognition: A dissociable ability? *Neuropsychologia*, 23, 395-402.