Edited by Christopher Brown

ADVANCES IN COMPUTER VISION

VOLUME 2

ADVANCES IN COMPUTER VISION

VOLUME 2

ADVANCES IN COMPUTER VISION

VOLUME 2

EDITED BY
CHRISTOPHER BROWN
University of Rochester

LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1988 Hillsdale, New Jersey Hove and London

Lawrence Erlbaum Associates, Inc., Publishers
365 Broadway
Hillsdale, New Jersey 07642

Library of Congress Cataloging-in-Publication Data

Advances in computer vision/edited by Christopher Brown.
p. cm.
Bibliography: p.
Includes indexes.
ISBN 0-89859-648-3 (v. 1). ISBN 0-8058-0092-1 (v. 2)
1. Computer vision. 2. Brown, Christopher M.
TA1632.A265 1988 87-27282
006.3'7——dc19 CIP

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

CONTENTS

VOLUME 1

VOLUME 2

CONTRIBUTORS

John Aloimonos Department of Computer Science, University of Maryland, College Park, MD 20742

Dana Ballard Department of Computer Science, University of Rochester, Rochester, NY 14627

Christopher Brown Department of Computer Science, University of Rochester, Rochester, NY 14627

Allen Hanson Computer and Information Sciences Department, University of Massachusetts, Amherst, MA 01002

Robert Haralick Department of Electrical Engineering, University of Washington, Seattle, WA 98195

Alex Pentland Vision Sciences Group, Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139

Tomaso Poggio Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139

Ting-Chuen Pong Department of Computer Science, University of Minnesota, Minneapolis, MN

Edward Riseman Computer and Information Sciences Department, University of Massachusetts, Amherst, MA 01002

Linda Shapiro Department of Electrical Engineering, University of Washington, Seattle, WA 98195

Allen Waxman Electrical, Computer and Systems Engineering, College of Engineering, Boston University, Boston, MA 02215

Kwangyoen Wohn GRASPLAB, Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104

Allan Yuille Division of Applied Sciences, Harvard University, Pierce Hall, Cambridge, MA 02139

INTRODUCTION

VISION AND COMPUTER VISION

The series *Advances in Computer Vision* has the goal of presenting current approaches to basic problems that arise in the construction of a computer vision system, written by leading researchers and practitioners in the field. The first two volumes in the series comprise seven chapters, which together cover much of the scope of computer vision. Chapter 1 of Volume 1 is referred to as 1.1, and so forth.

In these volumes, computer vision means computer programs analyzing visual input (like a television image of a three-dimensional scene) and deriving from the image some description of the scene that is helpful to further reasoning or action concerning the scene (Pentland, 1986). The technical aspects of creating images and transferring them to the computer's memory are not addressed here: We are concerned with the techniques of computerized image analysis. Computer graphics, or the computerized production of images, is the inverse of computer vision: Graphics starts with world descriptions and produces images. Since graphics is so closely related to vision, some graphics techniques are indeed interesting to vision researchers (see chapter 2.4).

Vision is our most powerful sense. If we could endow machines with something like the power of sight, their actions and decision-making could be much more effective and efficient. Automatic navigation systems, industrial inspection, biomedical applications, consumer products, interactive systems, and a host of other areas are potential beneficiaries of breakthroughs in computer vision. Of course in a real system vision is purposive, working toward some particular goal. That means that vision is only a component of a larger system for reasoning and action. There are many hard problems yet to be solved in creating intelligent systems, even if the vision problem were overcome. Still, vision has always been a key problem in artificial intelligence (AI), and it is quite possible to make progress in vision without solving all other problems as well.

Despite our intimate association with them, our brains and minds still provide many of science's deepest and most elusive questions. Each generation applies the most powerful mechanisms it can conceive to try to describe and explain the mental functions of memory, cognition, and perception. Thus one of the currently most productive approaches to understanding mental activity and its underlying hardware is to make a computational model of the processing and then explore both its logical behavior and its implications for physical implementations. A computer program is an experimental artifact that can be used to explore theories of brain function, and the methods and representations it employs may have implications for our understanding of biological vision. The neurosciences are making great strides in describing brain structure, and there is increasing interest in relating structure to function. Thus the natural sciences and the sciences of the artificial are working together to understand general vision systems. Much has been learned, but we are a long way from knowing how to do general vision. The chapters in this volume detail aspects of the vision problem and indicate the range of issues that must be addressed.

A HISTORICAL PERSPECTIVE

The First Vision System. Computer vision began in the mid 1960s with what today would be called a vision system that operated in the polyhedral or "blocksworld" domain (Roberts, 1965). The system started with a photographic image of a blocks scene of planar polyhedral shapes built of a small set of primitives that could be scaled, rotated, and combined by gluing. A scanning densitometer converted the photograph to an array of numbers corresponding to brightnesses. The numerical array was analyzed in several stages: Small "edge elements" were detected where the image changed brightness. These small edges were linked together into longer line segments that ideally corresponded with the images of straight edges in the polyhedral scene. The line segments were linked together at vertices that should correspond to the image of polyhedral corners. Rings of connected line segments formed polygons that delineated the blocks' faces in the image.

Topological analysis of the resulting line segments, vertices, and polygons allowed matching between the two-dimensional image and a computer-held data structure representing the three-dimensional polyhedral primitive building blocks. After matching, the program could infer how the scene was constructed from transformed and combined primitives. The final output was a line drawing of the scene from any desired viewpoint.

Besides its many technical contributions, Roberts's work provided a paradigm that in many ways still holds in computer vision: that of using a careful combination of local image evidence (e.g., edge location) and relational image evidence (e.g., how line segments touch) to make a progressive abstraction from the image signal into symbolic representations that can be used in practical jobs such as recognition or navigation.

Vision as Cognition. After Roberts, vision systems continued to be built, but their performance seemed weak on an absolute scale, or even weak compared to the amount of work that went into them. Something was missing. One influential idea, favored also by cognitive psychologists, was that "high-level" (cognitive) processes were at the heart of vision, and that vision should be approached as some form of problem-solving (Ballard & Brown, 1982). This view has its points and also happened to fit in neatly with the economics of computing at the time—low-level vision processing was computationally very expensive. However, the cognitive vision approach faltered because automated symbolic reasoning proved to be very difficult, and because in vision, the input often does not correspond to expectation. Thus, new research directions arose.

Perception of Three-Dimensions from a Two-Dimensional Image. The next major idea was to use physics and applied mathematics to determine the sorts of information available in an image and how it can be extracted (Horn, 1986). An important first question is how to retain and represent image information. In Volume 1, image information is usually represented as the gray-level in the image. Volume 2 presents more ideas for intermediate data representations. Chapter 2.1 discusses the representation of an image as a faceted surface, and chapter 2.2 considers the representation of an image as zero-crossings of an operator (which is something like a representation of the edges in an image).

The next step is computing "intrinsic images," or invariant physical scene parameters, from a potentially varying image signal. With general viewing conditions, the image of a scene can vary widely, while the physical scene itself, including the distance of its objects or their surface reflectance, remains the same. Understanding the physical parameters of the scene is a big step toward being able to make symbolic descriptions that can be used for tasks. By what knowledge and algorithms can the imaging process be inverted to yield a representation of the physical properties of the scene that produced the image? Particular forms of this question have been the most popular topics for computer vision research for the last decade, and have led to work on "vision modules" that might be incorporated into a complete image-understanding system. Usually these modules operate with minimal assumptions about the domain of the scene, and thus fit into a "low level" or "preattentive" role in a system. Usually research on low-level modules avoids the engineering and systems issues of an integrated system. Chapters 1.2 and 1.3 describe state-of-the-art work in extracting three-dimensional parameters from shading, stereo, and motion.

Biological and Computer Vision. Another important trend that started in the mid 1970s was to try to learn technical things about low-level computer vision from biological systems and, conversely, to try to make computational models to explain biological systems (Marr, 1982; Levine, 1985). Biological vision systems work very well compared to computer vision systems. This cross-fertilization between the neurosciences and computer sciences has been increasingly productive and promises to be a major force in the future of the field. Technology is advancing toward the goal of powerful parallel computers that can comfortably accommodate more brain-like models of computation. Chapters 2.2 and 2.3 are indicative of work at the interface between biological and computer systems.

Modern Vision Systems. Recently, in the past few years, another reaction has begun. The work on vision modules has been so productive of results that the community has turned a substantial part of its attention back to the problem of building integrated systems. This activity is resulting in autonomous land vehicles, robotic systems, and laboratory image-understanding systems. The work described in chapter 1.1 is an example of a large-scale integrated system. A system for recognizing objects must have a way to represent the objects, and one of the greatest challenges in "high-level" vision is to represent symbolically the everyday real-world objects of our visual world—objects with biological or manufactured shapes and with complex surface properties. Chapter 2.4 examines techniques for representing texture and shape that could be useful in an integrated system. Chapter 2.3 shows how a hierarchy of abstraction for polyhedral vision can be implemented with nets of neural-like simple computing units.

A FUNCTIONAL PERSPECTIVE

Another way to think about the chapters in this series is in terms of the functional components of a vision system. The organization of chapters is top-down, then bottom-up. Chapter 1.1 starts with a description of a modern vision system that is not only of considerable technical interest, but also provides an example vision system that may help the reader structure the content of the remaining chapters in the two volumes. Volume 1 then goes on in chapters 2 and 3 to concentrate on the extraction of three-dimensional physical parameters from two-dimensional images as part of early vision. Volume 2 moves from representations of the input image that are useful for further processing (chapters 1 and 2) to higher-level vision processes and three-dimensional representations (chapters 3 and 4).

Low-level Input Analysis and Image Representation. The idea that the image can be used to suggest a few clues and then be discarded, leaving perception to cognitive levels, seems indefensible (Fodor, 1983). Chapters 2.1 and 2.2 deal explicitly with image representations. Chapter 2.2 deals with the vital question of the "natural scale" of a vision calculation: How can a system discover the proper spatial resolution in which to perform image analysis? Where is the meaningful signal in the image? Chapter 2.1 provides a scheme in which the original image information is preserved but also put into a form more accessible for later processing. Chapter 1.1 addresses these issues in the context of an integrated system.

Segmentation. The problem of segmentation is related to the psychological phenomenon of figure-ground perception, or the perception of objecthood. Our natural and unconscious ability to see objects as unities and to delineate their boundaries has no easy implementation in a computer vision system. It is likely that biological systems use multiple sources of image information for figure-ground discrimination (motion, color, texture, and so on—see chapter 1.2). Segmentation is a hard problem that practical systems must solve. Most intrinsic image computations assume that segmentation of the scene into objects and background has been done. Certainly high-level vision (say recognition) usually assumes segmentation. Chapters 1.1, 1.3, 2.3, and 2.4 directly address the topic of segmentation.

Intrinsic Image Computations. The purpose of these computations is to extract physical information about the scene from the image. The motivation is that the physical information will be less sensitive to irrelevant variations to which the image is prey (such as lighting effects), and thus will be a more reliable basis for practical decisions such as manipulation, navigation, or recognition. Chapters 1.2 and 1.3 are mainly concerned with intrinsic images calculation, using clues of shading, multiple cameras, and motion.

System Architecture. System architectures for vision are now being designed as part of several major projects in research centers around the country. Often the system is built around a "blackboard" that serves as communications medium, central representation, and sometimes as a form of autonomous inference or geometrical computing engine. In this volume, chapter 1 gives an explicit description of the detailed makeup of such a vision system. The neural architecture that could underlie vision and thinking in general is of course an important topic in the cognitive sciences. Chapter 2.3 outlines a hierarchy of abstractions implemented in a neural-like net of small processing units.

Hardware Architecture. There is much activity today in designing and building special-purpose hardware architectures. Low-level vision (like graphics and image processing) has long been a candidate for such work because of the simplicity and physical locality of many of the operations. Recently high-level vision, or cognitive processes in general, has influenced the development of computational models (chapter 2.3; Feldman, 1985) and actual computers (Hillis, 1986). Advances in the power and packaging of microcomputers have led to new general-purpose architectures involving many computers working in parallel. These and more special-purpose hardware architectures are very promising tools for advancing all levels of the computer vision problem. In this volume the hardware architecture issue is addressed in chapter 1.

REFERENCES

Ballard, D., & Brown, C. (1982). *Computer vision.* Englewood Cliffs, NJ: Prentice-Hall.

Feldman, J. A. Editor. (January–March 1985). *Cognitive Science 9,* 1, Special Issue on Connectionist Models.

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Hillis, D. (1985). *The connection machine.* Cambridge, MA: MIT Press.

Horn, B. K. P. (1986). *Robot vision.* Cambridge, MA: MIT Press.

Levine, M. (1985). *Vision in man and machine.* New York: McGraw-Hill.

Marr, D. (1982). *Vision.* San Francisco, CA: Freeman.

Pentland, A. P. (1986). *From pixels to predicates.* Norwood, NJ: Ablex.

Roberts, L. G. (1965). Machine perception of three-dimensional solids. In J. P. Tippet et al. (Eds.), *Optical and electro-optical information processing* (pp. 159–197). Cambridge, MA: MIT Press.

1 THE USE OF THE FACET MODEL AND THE TOPOGRAPHIC PRIMAL SKETCH IN IMAGE ANALYSIS

LINDA G. SHAPIRO
ROBERT M. HARALICK
University of Washington

TING-CHUEN PONG
University of Minnesota

INTRODUCTION

The *facet model* states that all processing of digital-image data has its final authoritative interpretation relative to what the processing does to the underlying gray-tone intensity surface. The digital image's pixel values are noisy sampled observations of the underlying surface. Thus, in order to do any processing, we must estimate this underlying surface at each pixel position. This requires a model that describes what the general form of the surface would be in the neighborhood of any pixel if there were no noise. To estimate the surface from the neighborhood around a pixel, then, amounts to estimating the free parameters of the general form. The processing that takes place is then defined in terms of the estimated parameters.

The topographic primal sketch (Haralick, Watson, & Laffey, 1983) is one possible way of representing the fundamental structure of a digital image in a rich and robust way. The basis of the topographic primal sketch is the classification and grouping of the underlying image-intensity surface patches according to the categories defined by monotonic, gray-tone invariant functions of directional derivatives. Examples of such categories are peak, pit, ridge, ravine, saddle, flat, and hillside. From this initial classification, categories can be grouped to obtain a rich, hierarchical, and structurally complete representation of the fundamental image structure. By contrast, representations of the fundamental image structure only involving edges or the primal sketch as described by Marr (1976) are impoverished in the sense that they are insufficient for unambiguous matching. They also do not have the required invariance with respect to monotonically increasing gray-tone transformations.

The facet approach can also be used in classical gradient-based edge detection, in image segmentation, as well as in determining the topographic primal sketch of an image. The following sections discuss the facet model for image-data specialized to the sloped facet case and its direct application to gradient-edge detection; the facet model concepts as they can apply to image segmentation; the definition of the topographic primal sketch and how the information it requires can all come from the facet-model estimates; and three-dimensional object surface-shape estimation based on the patterns of the topographic primal sketch.

THE FACET MODEL FOR IMAGE DATA

The commonly used general forms for the facet model include piecewise constant (flat facet model), piecewise linear (sloped facet model), piecewise quadratic, and piecewise cubic. In the flat model, each ideal fitting neighborhood in the image is constant in gray tone. In the sloped model, each ideal fitting neighborhood has a gray tone surface that is a sloped plane. Similarly, in the quadratic and cubic models, regions have gray tone surfaces that are quadratic and cubic surfaces, respectively.

Given a noisy defocused image, and assuming one of these models, the problem is to estimate the parameters of the underlying surface for a given neighborhood and estimate the variance of the noise. These estimates can then be used in a variety of ways: edge detection, line detection, corner detection, and segmentation. In this section we review the parameter estimation problem for the sloped facet model and illustrate its use in the classic gradient edge detector application.

Sloped Facet Parameter and Error Estimation

In this discussion we employ a least-squares procedure to estimate the parameters of the sloped facet model for a given rectangular neighborhood whose row index set is R and whose column index set is C. The facet parameter estimates are obtained for the central neighborhood of each pixel on the image. We assume that for each $(r, c) \in R \times C$, the image function g is modeled by

$$g(r, c) = \alpha r + \beta c + \gamma + \eta(r, c)$$

where η is a random variable indexed on $R \times C$, which represents noise. We will assume that η is noise having mean 0 and variance σ^2 and that the noise for any two pixels is independent.

The least-squares procedure determines an $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$, which minimize the sum of the squared differences between the fitted surface and the observed one:

$$\epsilon^2 = \sum_{r \in R} \sum_{c \in C} [\hat{\alpha} r + \hat{\beta} c + \hat{\gamma} - g(r, c)]^2.$$

Taking the partial derivatives of ϵ^2 and setting them to zero results in

$$\begin{pmatrix} \frac{\partial \epsilon^2}{\partial \hat{\alpha}} \\ \frac{\partial \epsilon^2}{\partial \hat{\beta}} \\ \frac{\partial \epsilon^2}{\partial \hat{\gamma}} \end{pmatrix} = 2 \sum_{r \in R} \sum_{c \in C} (\hat{\alpha} r + \hat{\beta} c + \hat{\gamma} - g(r, c)) \begin{pmatrix} r \\ c \\ 1 \end{pmatrix} = 0. \tag{1}$$

Without loss of generality, we choose our coordinate system $R \times C$ so that the center of the neighborhood $R \times C$ has coordinates (0, 0). When the number of rows and columns is odd, the center pixel, therefore, has coordinates (0, 0). When the number of rows and columns is even, there is no pixel in the center but the point where the corners of the four central pixels meet has coordinates (0, 0). In this case, pixel centers will have coordinates of an integer plus a half.

The symmetry in the chosen coordinate system leads to

$$\sum_{r \in R} r = 0 \text{ and } \sum_{c \in C} c = 0$$

Hence,

$$\sum_r \sum_c \hat{\alpha} r^2 = \sum_r \sum_c r g(r, c),$$

$$\sum_r \sum_c \hat{\beta} c^2 = \sum_r \sum_c c g(r, c),$$

$$\sum_r \sum_c \hat{\gamma} = \sum_r \sum_c g(r, c).$$

Solving for $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ we obtain

$$\begin{aligned} \hat{\alpha} &= \sum_r \sum_c r g(r, c) / \sum_r \sum_c r^2, \\ \hat{\beta} &= \sum_r \sum_c c g(r, c) / \sum_r \sum_c c^2, \\ \hat{\gamma} &= \sum_r \sum_c g(r, c) / \sum_r \sum_c 1. \end{aligned} \tag{2}$$

Replacing $g(r, c)$ by $\alpha r + \beta c + \gamma + \eta(r, c)$ and simplifying the equations will allow us to explicitly see the dependence of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ on the noise. We obtain

$$\hat{\alpha} = \alpha + \left(\sum_r \sum_c r\eta(r, c) \,/\, \sum_r \sum_c r^2 \right),$$

$$\hat{\beta} = \beta + \left(\sum_r \sum_c c\eta(r, c) \,/\, \sum_r \sum_c c^2 \right),$$

$$\hat{\gamma} = \gamma + \left(\sum_r \sum_c \eta(r, c) \,/\, \sum_r \sum_c 1 \right).$$

From this it is apparent that $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\eta}$ are unbiased estimators for α, β, and γ, respectively, and have variances

$$V[\hat{\alpha}] = \sigma^2 \,/\, \sum_r \sum_c r^2,$$

$$V[\hat{\beta}] = \sigma^2 \,/\, \sum_r \sum_c c^2,$$

$$V[\hat{\gamma}] = \sigma^2 \,/\, \sum_r \sum_c 1.$$

Normally, distributed noise implies that $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are normally distributed. The independence of the noise implies that $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are independent since they are normal and that

$$E[(\hat{\alpha} - \alpha)(\hat{\beta} - \beta)] = E[(\hat{\alpha} - \alpha)(\hat{\gamma} - \gamma)] = E[\hat{\beta} - \beta)(\hat{\gamma} - \gamma)] = 0$$

as a straightforward calculation shows.

Examining the squared error residual ϵ^2 we find that

$$\begin{aligned}
\epsilon^2 &= \sum_r \sum_c [(\hat{\alpha} r + \hat{\beta} c + \hat{\gamma}) - (\alpha r + \beta c + \gamma + \eta(r, c))]^2 \\
&= \sum_r \sum_c [(\hat{\alpha} - \alpha)^2 r^2 + (\hat{\beta} - \beta)^2 c^2 + (\hat{\gamma} - \gamma)^2 + \eta^2(r, c) \\
&\quad - 2(\hat{\alpha} - \alpha) r\eta(r, c) - 2(\hat{\beta} - \beta) c\eta(r, c) - 2(\hat{\gamma} - \gamma)\eta(r, c)].
\end{aligned}$$

Using the fact that

$$(\hat{\alpha} - \alpha) = \sum_r \sum_c r\eta(r, c) \,/\, \sum_r \sum_c r^2,$$

$$(\hat{\beta} - \beta) = \sum_r \sum_c c\eta(r, c) \,/\, \sum_r \sum_c c^2,$$

$$(\hat{\gamma} - \gamma) = \sum_r \epsilon R \sum_c \eta(r, c) \,/\, \sum_r \sum_c 1$$

we may substitute into the last three terms for ϵ^2 and obtain after simplification

$$\epsilon^2 = \sum_r \sum_c \eta^2(r, c) - (\hat{\alpha} - \alpha)^2 \sum_r \sum_c r^2 - (\hat{\beta} - \beta)^2 \sum_r \sum_c c^2 - (\hat{\gamma} - \gamma)^2 \sum_r \sum_c 1$$

Now notice that

$$\sum_r \sum_c \eta^2(r, c)$$

is the sum of the squares of

$$\sum_r \sum_c 1$$

independently distributed normal random variables. Hence,

$$\sum_r \sum_c \eta^2(r, c)/\sigma^2$$

is distributed as a chi-squared variate with

$$\sum_r \sum_c 1$$

degrees of freedom. Because $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ are independent normals,

$$((\hat{\alpha} - \alpha)^2 \sum_r \sum_c r^2 + (\hat{\beta} - \beta)^2 \sum_r \sum_c c^2 + (\hat{\gamma} - \gamma)^2 \sum_r \sum_c 1)/\sigma^2$$

is distributed as a chi-squared variate with 3 degrees of freedom. Therefore, ϵ^2/σ^2 is distributed as a chi-squared variate with

$$\sum_r \sum_c 1 - 3$$

degrees of freedom. This means that $\epsilon^2/(\sum_r \sum_c 1 - 3)$ can be used as an unbiased estimator for σ^2.

Gradient-Based Facet-Edge Detection

Suppose that our model of the ideal image is one where each object part is imaged as a region that is homogeneous in gray tone. In this case the boundary between object parts will manifest itself as jumps in gray level between successive pixels on the image. A small neighborhood on the image that can be divided into two parts by a line passing through the middle of the neighborhood and in which all the pixels on one side of the line have one gray level is a neighborhood in which the dividing line is indeed an edge line. When such a neighborhood is fitted with the sloped facet model, $\hat{\alpha}r + \hat{\beta}c + \hat{\gamma}$, a gradient magnitude of $\sqrt{\hat{\alpha}^2 + \hat{\beta}^2}$ will result. The gradient magnitude will be proportional to the gray-level jump. On the other hand if the region is entirely contained within a homogeneous area, then the true surface $\alpha r + \beta c + \gamma$ will have $\alpha = \beta = 0$ and the fitted, sloped facet model $\hat{\alpha}r + \hat{\beta}c + \hat{\gamma}$ will produce a value of $\sqrt{\hat{\alpha}^2 + \hat{\beta}^2}$ which is near zero. Hence, it is reasonable for edge detectors to use the estimated gradient magnitude $\sqrt{\hat{\alpha}^2 + \hat{\beta}^2}$ as the basis for edge detection. Such edge detectors are called "gradient based edge detectors." There are other kinds of edge detectors such as zero-crossing edge detectors. A discussion of how the facet model can be used to determine zero crossings of second directional derivatives as edges can be found in Haralick (1984).

The most interesting question in the use of the estimated gradient $\sqrt{\hat{\alpha}^2 + \hat{\beta}^2}$ as an edge detector is how large does the gradient have to be in order for it to be considered significantly different from 0. The discussion that answers this question begins by noting that $\hat{\alpha}$ has a normal distribution with mean α and variance $\sigma^2 / \sum_r \sum_c r^2$, that $\hat{\beta}$ has a normal distribution with mean β and variance $\sigma^2 / \sum_r \sum_c c^2$, and that $\hat{\alpha}$ and $\hat{\beta}$ are independent. Hence,

$$\frac{(\hat{\alpha} - \alpha)^2 \sum_r \sum_c r^2 + (\hat{\beta} - \beta)^2 \sum_r \sum_c c^2}{\sigma^2}$$

is distributed as a chi-square variate with 2 degrees of freedom. From this it follows that to test the hypothesis of no edge under the assumption that $\alpha = \beta = 0$, we use the statistic G

$$G = \frac{\hat{\alpha}^2 \sum_r \sum_c r^2 + \hat{\beta}^2 \sum_r \sum_c c^2}{\sigma^2}$$

which is distributed as a chi-squared variate with 2 degrees of freedom. If the statistic G has a high enough value, then we reject the hypothesis that there is no edge.

If the neighborhood used to estimate the facet is square, then $\sum_r \sum_c r^2 = \sum_r \sum_c c^2$ so that the test statistic is a multiple of the estimated squared gradient magnitude $\hat{\alpha}^2 + \hat{\beta}^2$. Such an edge operator is the well known Prewitt edge operator. However, by knowing the conditional distribution given no edge, it becomes easier to choose a threshold. For example, suppose we want the edge detector to work with a controlled