Bias and Noise in the
Hough Transform I: Theory

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

TR 105
June 1982

The preparation of this paper was supported in part by the Office of
Naval Research under Grants N00014-80-C-0197 and N00014-82-K-0193,
and in part by the Defense Advanced Research Projects Agency under
Grant N00014-78-C-0164.
Abstract

The main results of this work are the following.

1. An approach to Hough Transforms (HT) based on a linear imaging model. The HT produces a peak in accumulator (parameter) space corresponding to likely parameters of an interesting phenomenon in the image. It also produces background off-peak sidelobes, which are important because they can add significant variance and background bias to the accumulator space and make peak-finding a local and difficult process.

2. Definitions of bias and noise in the HT.

3. Analytical techniques for reasoning about sidelobe shapes, and some descriptions of sidelobes in practically important parameter spaces under continuous, noise-free conditions.

4. The CHough method, which can eliminate sidelobe bias and decrease sidelobe variance. The peak height is then an unbiased estimator for the amount of evidence consistent with the peak parameter vector, and simple global techniques (such as global thresholding) will find peaks. Compared to traditional HT, CHough seems to have much better sidelobe bias properties, significantly better sidelobe variance properties, equal or worse quantization noise properties, and equal or better resistance to "noise" features in the image.

Index terms: Hough transform, shape detection, pattern recognition, histograms, noise suppression, parallel algorithms.
0. Motivation and Overview

Imagine a noiseless image of a single shape, perhaps a circle or straight line. The shape is to be detected by a Hough transform (HT) (Section 1.1). In this case, the HT results in an accumulator array with a single peak—an array element that has collected the most votes. The coordinates of the peak element are the parameters of the shape (typically the coefficients of its equation, or its location, scale, and orientation). Also, the accumulator array contains off-peak votes, making up a sort of background to the peak. Call this background for a single perfect shape the contribution from sidelobes (this term is from the domain of imaging (Section 1.2)).

For an ideal image of a single shape, finding the peak is easy despite the sidelobe contribution. With multiple shapes, partial shapes, wrong shapes, and "noise" features, the sidelobes become important and troublesome. They add into and thus distort the heights of peaks in the accumulator, so peaks do not reflect accurately the amount of evidence pointing to a single shape instance. The sidelobe background means that local, not global, peaks must be detected in the accumulator—the sidelobe background may vary considerably over the accumulator, and may even create false peaks that are higher than peaks arising from shape instances.

The approach followed here is to model the HT as an imaging process and explore its inherent deficiencies, apart from its behavior with statistical and quantization noise and incomplete information. The hope is that the approach will separate the issues clearly so that real Hough applications may be studied with a clearer perception of possible factors contributing to their performance. Section 2 outlines the issues and proposes some definitions of signal and noise for application to the HT.

Sections 3, 4, and 5 present an analytical, continuous approach to the sidelobe behavior of circle-location (generalized to hypersphere location), line-location (generalized to hyperplane location), and general shape location (specialized to GHough techniques).

Section 6 presents CHough, a simple sidelobe-reduction trick. The trick is a general one that applies both to parametric HT and nonparametric GHough. In CHough a feature point produces "nay" votes as well as "yea" votes in the accumulator. As usual, the yea votes accumulate at the peak. The new wrinkle is that nay votes cancel off-peak yeas to reduce sidelobes. Only the germ of the idea appears here. A description of implementation and experiments appears in [Brown and Curtiss 1982]. One of the side effects of CHough is that with zero-mean sidelobe contributions, peak-finding reduces to simple thresholding. This is because the peak heights become unbiased estimators of the amount of evidence for the peak.
(Systematically positive sidelobe contributions bias the estimator.) Rather surprisingly, CHough seems to reduce sidelobe variance as well as bias.

1. The Hough Transform


1.1 The Hough Transform as Mapping

For basic shape detection, the HT

$$T_{S} : (x, I(x)) \rightarrow A(y)$$

is a function tuned to a shape S of n parameters.

The (generalized) image $I(x), x = (x_1, x_2, ..., x_m)$ is a (scalar or vector) function on an m-dimensional image space. The accumulator space $A(y), y = (y_1, y_2, ..., y_n)$ is a (scalar) function on an n-dimensional parameter space P. It is often simplest to consider the accumulator as an n-dimensional array whose elements contain accumulator space values. $T_{S}$ maps each vector $(x, I(x))$ into a set of values (votes) for a set of parameter vectors in $P$. Each $y$ in the set is a vector of parameters (e.g., values for location, scale, and orientation) that describes an instance of the shape S consistent with the image evidence $(x, I(x))$. The vote for each $y$ may be unity, or be weighted to represent a confidence value.

**Hough Transformation** is then:

1. Initialize the accumulator space to zero value.
2. Process image to find a set D of image feature points contributing evidence for the shape.
3. The value of $A(y)$ (i.e., the Hough Transform) is the sum of the $T_{S}(x, I(x))$ for each point $x$ in $D$.

The process of accumulating the $T_{S}(x, I(x))$ into $A(y)$ is sometimes known as **voting**. After all the votes from the image have been cast, then the transform must be interpreted.

**Hough Transform Interpretation** is simply to find maxima in $A(y)$. 

The idea is that maxima in accumulator space indicate parameter vectors that explain relatively much of the image data. That is, instances of S in the image with these parameters are likely, since there is relatively much evidence for them.

1.2 The Hough Transform as Imaging

The image of a point source in an imaging device such as a camera is called the point spread function of the device. In linear systems theory, the same concept is called the impulse response of the system. Real point spread functions have a central peak and non-zero sidelobes. In cameras the sidelobes arise from diffraction, aberration, etc. A very useful abstraction in geometrical optics is that the imaging system is linear. The main effect of interest is the superposition principle, which states that the image of a complex (e.g., continuous) spatial distribution of sources is the superposition of the response to each component source. Mathematically, the image is the convolution of the point spread function with the source distribution function, as we shall see. In a linear system, the ideal point spread function to reproduce the source is a delta function. The image of a point has energy proportional to that of the source point, concentrated in an infinitesimal space, with no spread of energy outside the peak.

We shall call a phenomenon of interest in the image an instance, and the evidence for an instance is a set of image features. In the HT, the feature point spread function (feature psf) is $T_S(x, I(x))$, the set of votes in parameter space that results from an image feature. The accumulator is a quintessential linear system, and by the superposition principle we define the parameter point spread function (parameter psf), which describes the result in parameter space of all the evidence for an instance. It is a superposition of feature psfs, one for each feature that voted. Likewise, if there are several instances in the image, the parameter space contains the superposition of their parameter psfs.

For definiteness, we shall often refer to shape detection here, but the issues are of course not restricted to that. In all that follows we shall loosely speak of shape when we mean the boundary of a shape, since often it is boundary features which provide the evidence for a shape. HT feature psfs are usually n-1 dimensional surfaces in n-space (lines or circles in the plane, for instance). Although nothing basic prevents a feature psf from being an n-dimensional subset of n-space, such psfs will not be investigated here.

Line finding provides a basic example of feature psfs. A point (not an edge element) on a line casts a sinusoid of votes in a rectangular ($r, \theta$) space of line parameters (or a circle of votes in a polar ($r, \theta$) space) [Sklansky 1978, Shapiro and Iannino 1979]. This should be clear from the line equation
\[ x \sin \theta + y \cos \theta + r = 0. \]

Circle location gives another example. Again, a point (not an edge element) on a circle of known radius \( R \) casts a circle of votes (of radius \( R \)) in the \((x,y)\) parameter space of circle centers. This is clear, too--any point \( R \) away from an edge point could be the center. Figure 1.1 shows a circle in \( x-y \) image space and four ideal circular feature psfs in \( x-y \) location space arising from the indicated feature points. The peak is at the intersection of the feature psfs.

<<Figure 1.1>>

Let us pursue the imaging analogy for 2-D shapes and location parameter space, because this case is basic and important. First, image space and parameter space are identical--they are \( x-y \) space (this was the case in the circle-location example but not the line-finding example). Note that every point on the shape in \( x-y \) image space adds a feature psf somewhere into accumulator (parameter) space.

Pick any point as the origin of a coordinate system for the shape \( S \) (this is like a master coordinate system in computer graphics [Newman and Sproull 1979]). Let \( u \) be the location of a feature point somewhere on the boundary of a shape instance \( S' \). If \( u \) actually corresponds to the point \( x \) on the boundary of the model shape, then the distinguished point of the instance must be at \((u-x)\). This is true for all boundary points \( x \). If

\[
g(x) = \begin{cases} 
1, & x \text{ in (the boundary of) } S \\
0, & x \text{ not in (the boundary of) } S 
\end{cases}
\]

then

Feature PSF: \( g(-(x-u)) = g(u-x) \)

Finally, the parameter point spread function due to all \( u \) in the instance of \( S' \) is

Parameter PSF \( = \int_{S'} g(u-x) \, du \)

\( = \int_{-\infty}^{\infty} g(u)g(u-x) \, du, \)

which is the autoconvolution of \( g \).
In fact, the parameter psf, or location-space image of a perfect shape in image space, is an image that would appear in a perfect pinhole camera with linear film. The object being imaged is the (boundary of the) shape. A point pinhole projects this shape boundary upside down and backwards; the right "complex pinhole" to give the parameter psf is itself the (boundary of the) shape (Figure 1.3).

Call this imaging-like transform the autoconvolution HT, after its ideal parameter psf. Autoconvolution HT is the general shape-location technique treated here. The GHough technique [Ballard 1981a], considered as pure shape detection, is a specialization of autoconvolution HT. GHough is a generalization of autoconvolution HT in that it extends to non-location parameters such as scale and rotation (see Section 1.3). GHough uses gradient information, only deals with a subset of possible feature point values, and perhaps weights the votes. The gradient information helps reduce sidelobes by choosing a small feature psf which varies with the feature. The effect of this is to produce just those votes near the peak, constraining sidelobes to a smaller area. By the superposition principle, the use of gradient information just produces a subset of the parameter psf for autoconvolution Hough. Thus GHough parameter psfs are exactly the near-peak portions of autoconvolution Hough parameter psfs.

The imaging view of HT simply recasts certain well-known problems, provides a small but useful vocabulary, and possibly indicates a source of applicable literature on the behavior of HT in various forms of noise [Helstrom 1968, Brown 1974]. It concentrates attention on the parameter psf, whose behavior determines inherent performance characteristics of an HT system.

1.3 The Hough Transform in Many Dimensions

It is often useful to use the HT to determine more than two parameters at once. Sections 3-5 treat multi-dimensional location parameters. Parameters orthogonal to the n-dimensional location parameters of the autoconvolution HT may be analyzed to determine their effect on the parameter psf.

Consider determining the location, orientation, and scale of a 2-D shape. What does the parameter psf look like? Start with the 2-D plane parameter space of x-y locations for a shape of known size and orientation. Determine the psf in this plane by techniques of Sections 3-5. Now there is one such plane for every orientation (in a continuous space these parameter psfs are rotations of one another). Note that the
resulting 3-D parameter space may be visualized as a set of independent 2-D ones for location. Finally, the whole 3-D space is swept through the fourth dimension of scale, being duplicated once (scaled) for every relevant size. Again the independence of the parameters means that the 2-D analysis will hold for any \((x,y)\) slice of 4-space.

2. Peaks, Sidelobes, and Noise

This section proposes definitions and treatments for phenomena of bias and noise that can affect HT performance. In most imaging applications, the image should be a faithful representation of the object: a point source should result in a point in the image, and the intensity of the image point should be predictably related to the intensity of the source. The HT parallel is that a single instance should result in a single point in parameter space corresponding to the correct parameters.

However, imaging systems do not have ideal point spread functions—hence finite resolving power, for instance. The HT parameter psf is likewise not ideal, having significant contributions off the peak from votes that did not contribute to the peak. Figure 2.1 shows a discrete version of the parameter psf for a circle. It is the superposition of several feature psfs, each one itself a circle. The feature psfs overlap maximally at the central peak and less elsewhere. Notice the external ring of relatively high values. This parameter psf is explored in Section 3.

\[ \text{Figure 2.1} \]

2.1 Sidelobes and Bias

If instances occur closely enough that sidelobes overlap peaks, the peaks start receding into the background bias. Say the peak is of unit height, the mean sidelobe height is \(p < 1\), and the number of instances involved is \(s\). Then the expected ratio of \((\text{total votes at a peak}) / (\text{total votes in a neighboring off-peak element})\) is

\[
\frac{\text{Peak} + \text{Bias}}{\text{Bias}} = 1 + \frac{1-p}{sp}.
\]

(2.1)

The most reasonable measure of signal is thus the \textit{height of the peak above local bias}. In the above case all the signals remain at unit height. However, from a practical standpoint it is worth noting that the "visibility" of the signal is often related to the measure given by eq. (2.1), which can deteriorate rapidly with non-zero sidelobes and many instances. Local peaks are harder to find than global ones. Sidelobe bias can accumulate to produce hills greater than local peaks. Further problems arise because of variation in the peak height and background due to statistical or inherent fluctuations (Figure 2.2)
2.2 Sidelobes and Inherent Noise

Sidelobes contribute bias, which is usually thought of as a low spatial frequency phenomenon. They also may have higher spatial frequency fluctuations. Without apriori knowledge of the source distribution, it is reasonable to model the contribution of sidelobes to peak height as a *noise process inherent in the imaging*. Then the bias component is reflected in the *mean* of this inherent noise process, and statistical fluctuations arise from its variability, measured by its *variance*. If the sidelobe heights were independent within a parameter psf, such a statistical view of the situation would clearly be called for. Such is not the case here, of course; however, multiple instances at random unknown locations effectively deposit random samples from the sidelobe distribution on other peaks. It is not unreasonable to apply a statistical model and speak of sidelobe variance as a form of (albeit inherent) noise in the imaging system.

A reasonable measure of signal to such fluctuation (a *signal to noise ratio* or SNR) is the ratio of peak height above background to the standard deviation of background fluctuation. For this calculation we just suppose the sidelobes have mean height zero while keeping their variance (alternatively, suppose the bias is magically subtracted out). Then with several overlapping source point spread functions, each with sidelobe variance $V_i$, we have:

$$\text{Peak}/\text{(Inherent Noise)} = \text{Peak}/\sqrt{(\Sigma V_i)}$$

(2.2)

2.3 Quantization and Location Noise

Section 3 uses continuous mathematical models of parameter psfs. With discrete spaces, the effects of *quantization* become significant. Quantization affects the shape of the feature psfs and the location (in parameter space) of the feature psfs. Another quantization effect is caused by *phasing*, or the choice of location, not frequency, of samples.

A rich tradition of quantization and location noise analysis exists [Sklansky 1978, Shapiro and Iannino 1979] It has a geometric flavor, and is based on the same sort of geometry that is analyzed in Section 4. Consider a point $x = (x, y)$. The measured position of a point is $x' = (x', y')$. Location noise is an isotropic uncertainty in the location of $x$: $|x-x'|<d$. Quantization noise is an uncertainty, not necessarily isotropic, in the measured components of $x$: $|x'-x<d_1$, $|y'-y<d_2$. When the feature psf is a single point the HT is a 1:1 mapping. This is a limiting case opposite in spirit to that of this paper (representing for example maximal rather than
minimal use of gradient information). The performance of 1:1 HT schemes in noise and many related issues are thoroughly treated in [Shapiro 1975, 1978].

A usable model of location, phasing, and quantization noise is (arbitrary) blur functions to be applied after the accumulator space image is calculated. They are shift-independent peak-broadening effects, which by the linearity of the system may be treated as a process applied to the final accumulator space.

Conjecture: anti-aliasing techniques applied to voting patterns can reduce quantization noise. Performance may be improved by the familiar computer graphics "de-jagging" technique that (in this context) votes into a quantized volume of parameter space with a strength proportional to the amount of feature point psf that contributes to it [Brown and Curtiss 1982]. This sort of weighting is not at all related to (or in conflict with) other vote-weighting schemes based on some measurement of feature strength.

2.4 Statistical Noise

Image features not associated with instances contribute background votes. Such votes from other instances are not independently distributed throughout accumulator space and may create significant accumulator space structure that interferes with detection of an instance. Other votes can arise from features that behave more like statistical noise: edge elements from irrelevant texture, short edge segments from unimportant image structures, and so forth.

A workable model for all such noise is that it arises from independently distributed features occurring with a certain frequency. Thus their contribution to the parameter space image is modelled as simply a level of background variance [Helstrom 1968].

If the density of noise features in the image is D, and the variance induced in parameter space by a point feature spread function is V, then

\[
\text{Peak}/(\text{Statistical Noise}) = \text{Peak}/\sqrt{(DV)}
\]  

(2.3)
3. The Hypersphere Shape PSF

Here we investigate parameter psfs that are ideal in the sense that they are continuous versions of the usual discrete situation and they arise from complete shape instances. The simplest case is that of circle location, which leads to a radially symmetric parameter psf. The n-dimensional hypersphere generalization is interesting not just for n-1 sphere detection (perhaps not a common case) but because it illustrates a limiting case of nonplanar isotropically distributed surfaces intersecting in n-space.

3.1 Two Dimensions and General Intuition

The continuous parameter psf $P(x)$ of a circle $C$ of radius $R$ is the sum of an infinite number of circles of radius $R$ with centers on points of $C$'s circumference. We model the height of the psf as the limit of the density of feature psf contributions in a small area. Figure 3.1 shows that $P(x)$ is radially symmetric. At the peak, all the feature psfs contribute; the density seems to fall off with increasing distance from the peak and then to increase again near the edge.

<<Figure 3.1>>

Consider $Q(r)$, the density of the parameter psf as a function of radial distance from the central peak. Thus $Q(r)$ is related to the amount of arc length $dA$ cut off $C$, a single circle of radius $R$, by an annulus of radius $r$ and thickness $dr$, centered on the circumference of $C$ (Figure 3.2).

<<Figure 3.2>>

In the limit, $Q(r)$ is proportional to $dA/(2\pi r dr)$, or to the amount of cut-off arc length distributed through the area of the annulus. Anticipating Section 3.3, the n-dimensional situation is that the height of the sidelobes is proportional to $dA/dr$ (the rate at which n-dimensional surface area is being cut off with increasing $r$) divided by the surface area of the hypersphere of radius $r$.

Back to two dimensions, refer to Figure 3.3. Here $A$ is the arc length being cut off circle $C$ of radius $R$ by circle $C'$ of radius $r$.

<<Figure 3.3>>
From the diagram:

\[ A(r) = 2s, \]
\[ s = R\theta \]
\[ \theta = 2 \arcsin(r/(2R)). \]

So

\[ A(r) = 4R \arcsin(r/(2R)), \]

and

\[ \frac{dA}{dr} = \frac{2}{\sqrt{1-(r/(2R))^2}}. \]

Hereafter, we shall always assume \( R = 1. \) Then

\[ Q(r) = \frac{1}{\pi \sqrt{1-(r^2/4)}} \tag{3.1} \]

Equation (3.1) shows that \( Q(r) \) goes to infinity at \( r = 0, \) which accords with intuition, and at \( r = 2, \) which may be surprising.

3.2 Three Dimensions

The analogous problem in 3-D is a textbook problem in computing a surface of revolution (Figure 3.3). From the diagram (here let \( A \) be the surface area of the cap of the sphere cut off by the sphere of radius \( r):\)

\[ y = \cos \theta \]
\[ x = \sin \theta \]
\[ \cos \Phi = \sqrt{1-(r/2)^2} \]
\[ \cos \theta = \cos 2\Phi = 2 \cos^2 \Phi - 1 = 1-(r^2)/2. \]

The standard formula for the area of the surface of revolution is

\[ A(r) = 2\pi \int y(t) \sqrt{(dx/dt)^2 + (dy/dt)^2} \, dt \]

Substituting, we find the radical goes to unity and

\[ A(r) = 2\pi (1-\cos \theta) = \pi r^2. \tag{3.2} \]

Equation (3.2) again may be a surprise: it says that the expanding sphere cuts area out of the fixed sphere at the same rate it would cut area out of the plane. Finally,

\[ Q(r) = (dA/dr) / 4\pi r^2 = 1/(2r). \tag{3.3} \]
3.3 N Dimensions

Mark Hopkins [Hopkins 1982] furnished the general solution given in the appendix. We write \( A_n \) for the n-dimensional version of \( A \), the area cut off by a hypersphere on the circumference of the fixed hypersphere. We follow the practice of calling the sphere in \( n \)-space the \( n-1 \) sphere.

\[
dA_n/dr = 2\pi r (rq)^{-3} a^{-b} / ((n-3)(n-5))...(3.4)
\]

where

\[
q = \sqrt{1-(r^2/4)} \\
a = \text{floor}((n-2)/2) \\
b = \text{floor}((n-3)/2)
\]

This formula may be integrated to derive the same results about the surface area \( T_n \) of the \( n-1 \)-sphere as those provided by the recurrence [Coxeter 1963]

\[
Z(1) = 2 \\
Z(2) = 2\pi \\
Z(n+2) = 2\pi Z(n-2)/n \\
T_n = Z(n) r^{n-1}.
\] (3.5)

Table 1 shows some values of \( T_n \), \( dA_n/dr \), and \( Q(r) \) for dimensions 2 to 6. \( q \) is as in eq. (3.4).

<table>
<thead>
<tr>
<th>Dim</th>
<th>( dA_n/dr )</th>
<th>( T_n )</th>
<th>( Q(r) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>( q^{-1} )</td>
<td>( 2\pi r )</td>
<td>( 1/\pi rq )</td>
</tr>
<tr>
<td>3</td>
<td>( 2\pi r )</td>
<td>( 4\pi r^2 )</td>
<td>( 1/2r )</td>
</tr>
<tr>
<td>4</td>
<td>( 4\pi qr^2 )</td>
<td>( 2\pi^2 r^3 )</td>
<td>( 2q/\pi r )</td>
</tr>
<tr>
<td>5</td>
<td>( 2\pi^2 q^2 r^3 )</td>
<td>( (8/3)\pi^2 r^4 )</td>
<td>( 3q^2/4r )</td>
</tr>
<tr>
<td>6</td>
<td>( (8/3)\pi^2 q^3 r^4 )</td>
<td>( \pi^3 r^5 )</td>
<td>( 8q^3/3\pi r )</td>
</tr>
</tbody>
</table>

The height of the parameter psf as a function of radius \( r \) is generally \( (kq^{(n-3)})/r \), for various constants \( k \). In two dimensions, the height increases for \( r > \sqrt{2} \), and approaches the limit of infinity again at \( r = 2 \) (Figure 2.1 showed a digital approximation). After the two surprises in two and three dimensions, the situation stabilizes, and the parameter psfs (with \( k = 1 \) for each) take on the form indicated by Figure 3.4. They approach weighted versions of \( 1/r \) with increasing dimensionality.
If gradient information is used to restrict the feature point psf to a small segment of arc, then only the sidelobes near the central peak are developed. The parameter psf will then be the same as given here out to some $r < 2$. For any significant reduction in $r$ by this method, the behavior of the parameter psfs in all dimensions is essentially identical, since they are similar near the peak.

<<Figure 3.4>>

4. The Hyperplane Shape PSF

Line (and ultimately hyper-plane) detection is a more complex case. It is mildly interesting that the feature point spread functions for hyperspheres and hyperplanes can be made identical—they can both be spheres.

In n-dimensional Cartesian coordinates, take the feature psf of a point $x$ on a linear n-1-space surface $L$ to be an n-1 sphere (in the same space) whose diameter is $||x||$. The sphere goes through $x$ and the origin. Then [Sklansky 1978, Shapiro and Iannino 1979] all such feature psfs from $L$ intersect at $(p,a_1,a_2,...,a_n)$ where

$$a_1x_1 + a_2x_2 + ... a_nx_n + p = 0 \quad (4.1)$$

is the equation for $L$. This version of the accumulator array is a polar version of the perhaps more usual Cartesian non-periodic one in which the feature psfs are sinusoids. A two-dimensional example is shown in Figure 4.1.

<<Figure 4.1>>

The parameter psf for a linear surface is thus the superposition of an infinite number of n-1 spheres of diameter varying from $p$ to infinity whose centers are all points of a linear surface parallel to the one being detected. This parameter psf is manifestly not radially symmetric. Its density goes toward infinity in the direction normal to the surface as increasing numbers of spheres of large radii cut through the peak.

4.1 Near-Peak Sidelobe Behavior

Going straight to N dimensions, we characterize the density of the parameter psf near the peak by making the peak the origin of the n-dimensional variant of the spherical coordinate system shown in Figure 4.2.

<<Figure 4.2>>
In this coordinate system with origin at the peak, the hyperspheres contributing surfaces that are normal to the angles in a small solid angle in hyperspace have their centers in the intersection of that solid angle and the linear hypersurface C. Let the linear hypersurface C containing the sphere centers be unit distance out along the $x_3$ axis (Figure 4.3).

Our task is then to characterize a volume element $dx_1dx_2...dx_{n-1}$ in the n-dimensional hyperplane at unit distance as a function of its angular direction $(\varphi_1, \varphi_2,...,\varphi_{n-1})$. From the diagram

$$\rho_1 = \cot\varphi_2$$
$$d\rho_1 = \csc^2\varphi_2d\varphi_2.$$  
$$\rho_1 = \rho_2\cos\varphi_2$$

In fact, in general

$$\rho_i = \rho_{i+1}\cos\varphi_{i+1}.$$  

Now we are essentially done; following the same development as in the Appendix,

$$dx_1dx_2...dx_{n-1} = d\rho_{n-1}\rho_1d\varphi_1\rho_2d\varphi_2...\rho_{n-1}d\varphi_{n-1}$$  

$$d\rho_{n-1} = \cot\varphi_n$$
$$d\rho_{n-1} = \csc^2\varphi_n.$$  

Thus the hypervolume in the direction $(\varphi_1, \varphi_2,...,\varphi_{n-1})$ is

$$A_n = \cot^{n-1}\varphi_n\csc^2\varphi_n\cos^{n-2}\varphi_{n-1}\cos^{n-3}\varphi_{n-2}...\cos\varphi_2d\varphi_1d\varphi_2...d\varphi_{n-1}.$$  

Specializing this to $n = 2$ (Figure 4.4) shows that the amount of line cut off by a $d\varphi_2$ is $\csc^2\varphi_2d\varphi_2$. (See also Figure 4.6.)

The result $A$ is independent of $\varphi_1$, as we would expect; the parameter psf is rotationally symmetric about this axis.
4.2 Off-Peak Sidelobe Behavior

The peak height of an instance's HT should be proportional to the amount of evidence contributing to it. For hyperspheres, the evidence was the bounded surface of the hypersphere. Hyperplanes are unbounded, and thus the peak height is unbounded as well. Section 4.1 showed that for hyperplanes the rate at which the peak falls off depends on the direction taken away from the peak except for one axis of rotational symmetry.

Figure 4.5 shows the geometry of the (2-D) situation. The question of off-peak height is (as usual) the question of how fast an expanding hypersphere cuts off hyper surface area from a collection of hyperspheres. In this case the latter hyperspheres are of varying radius.

It is easy to see what happens with finite hyperplane segments (line segments, say). Only those points on the segment contribute to the parameter psf, and their distribution on the segment determines the shape of the peak. One prediction is thus that segments with their coordinates large relative to \( p \) have ridgelike parameter psfs with no well-defined peak. A similar problem plagues the case of large \( p \). These problems are alleviated to an extent by peak-sharpening techniques such as restricting the range of \( \theta \) with gradient information. The help is perhaps not very satisfying—the "peak" is still a ridge, just a shorter one.

The asymmetric nature of the parameter psf might be alleviated by quantizing the parameter space anisotropically. Dynamic quantization techniques [Sloan 1981, O'Rourke 1981] might find application here. It is hard to see how they would interact with CHough techniques (Section 6).

The reader may wish to use the result of the Appendix to solve the problem of average sidelobe height as a function of radius. First, allow the stationary (non-expanding) hypersphere to have non-unit radius \( R \) in the expression for \( dA/dr \), which now becomes a function of \( r \) and \( R \). Then integrate over \( R \) for any \( r \) to get the total sidelobe contribution to be distributed over the hypersphere of radius \( r \). It is easy to get a qualitative feel for the solution. Clearly the parameter psf is the superposition of a number of functions, each of which looks like the ideal hypersphere parameter psf of Section 3.3 shown in Figure 3.4. These functions are just versions of one another scaled in the \( r \) axis by amounts proportional to the \( R \) of the hypersphere being considered. Thus the "peakiest" one is that with smallest \( R \), and the worst case is that with largest \( R \). The limiting case of infinite \( R \) is easy to calculate. It is just the ratio of surface area of the \( n-1 \) dimensional hypersphere to
the n-dimensional one of the same radius, and goes as some constant times $1/r$ (see eq. 3.5). Thus the worst-case parameter psf has an average cross section height that falls off in a hyperbola from an asymptotic infinity to an asymptotic zero.

Of course the extreme directional dependence (ridginess) of the parameter psf means that the average sidelobe height at a radius is probably not a very useful way to characterize the uncertainty contributed by such psfs. Better might be inherent noise modeled as line segments (corresponding to the ridges) in accumulator space.

### 4.3 Other Parameter Spaces for Lines

The polar version of $(\rho, \theta)$ space for lines is not the usual one, but the reader may verify that using a Cartesian representation of $(\rho, \theta)$ space does not change the ridge-like parameter psf for hypersurfaces.

The formula $y = mx + b$ yields the parameter space $(m, b)$ of slope and intercept. This space is little used because $m$ goes to infinity for vertical lines. Figure 4.6 shows that (apart from a scale factor) all ideal parameter psfs look alike in $(m, b)$ space. They are closely related to those in $(\rho, \theta)$ space except that the ridges are always parallel to the $y$ axis. From Figure 4.6, $b = mtan\theta$, so the parameter psf density in direction $\theta$ is $db = msec^2\theta d\theta$ (the situation is familiar from Figure 4.4).

The CHough technique (Section 6) tries to cancel out off-peak positive votes with negative votes. CHough will not work well with line detection if the evidence only contributes nearly overlapping line segment-like voting patterns. CHough works best with radially symmetric parameter psfs. An open research issue is to find one such for lines (or show why it cannot exist). One candidate is the $(x, y, \theta)$ space, a redundant parameter space in which the feature point psf for a point $(x, y)$ is a "spiral staircase" centered on the line $(x, y)$ whose "steps" (actually lines) are inclined at angle $\theta$ (Figure 4.7). From Figure 4.7, a line or line segment at $\theta = \theta_0$ produces a parameter psf whose height (really density) falls off as $csc(\theta - \theta_0)$. The density fall-off through $\theta$ is the same for infinite lines and line segments. Figure 4.7 shows that density is constant near the peak in any plane of constant $\theta$. Since the "peak" in this new space is a line, it requires nontraditional peak-finding techniques (one comes to mind immediately). It is not clear how much better the performance of this scheme is than the two others. It is interesting to extend all these analyses to higher-dimensional surfaces in $(x, \xi)$ space, with $\xi$ the surface normal vector.

<<Figure 4.6>>

<<Figure 4.7>>
5. Autoconvolution Shape PSF

The parameter psf for any boundary is its autoconvolution. Each point of the boundary contributes to the peak, and likewise each neighborhood of a boundary point contributes to the sidelobe behavior near the peak. Thus the behavior of sidelobes near the peak depends on the slope density function (SDF) of the boundary [Ballard and Brown 1982]. The SDF is just the histogram of slopes (tangent vectors) in the shape. (Figure 5.1). The SDF extends easily to n-space if surface normal vectors, not slopes, are histogrammed. In this guise it becomes an n-dimensional version of the familiar direction histogram or extended Gaussian image [Brown 1979, Ikeuchi 1981]. Parameter psfs for hypersphere location are radially symmetric because of the hypersphere's flat SDF.

Autoconvolution parameter psfs are thus not in general radially symmetric. However, for any direction, the falloff of sidelobe height with increasing distance from the peak shares the general character of the parameter psf for sphere detection.

Finally, the CHough technique proper [Ballard 1981] is an implementation of autoconvolution HT using gradient information. We expect it to approximate the near-peak behavior described above. The finite number of feature point angles means a discrete version of the SDF applies. The use of gradient information means there will be no sidelobes far from the peak.

6. CHough

Sections 3 and 4 showed that sidelobes are a source of considerable bias and quasi-statistical noise. How might they be reduced? The problems are similar to well-known ones in so-called "correlation detection," which is used to detect time-varying signals of an expected waveform. This matched-filtering (template matching) technique merely correlates the incoming waveform with the expected signal. When the signal comes through, the correlator produces the autocorrelation of the signal to indicate detection. Again the ideal such autocorrelation is a delta function, indicating with infinite peak/sidelobe ratio a precise moment of detection. Multi-valued signals may be designed (say with values of 1 and -1) such that some of the products in the autocorrelation function cancel [e.g., Barker 1953, Boehmer 1967, Brown 1974] (Table 2). In fact, n-ary codes have been designed for the same purpose [Welti 1960].
Table 2: The output of a correlation detector (i.e., the autocorrelation) of a 
(random) signal of thirteen 1's and 0's and a (clever) signal of thirteen 1's 
and -1's. Note that the sidelobe mean and variance of the latter are much 
reduced.

<table>
<thead>
<tr>
<th>Signal</th>
<th>Autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0 1 1 0 0 1 0 1 0 1 1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Signal</th>
<th>Autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 1 1 1 -1 -1 1 1 -1 1 -1 1</td>
<td></td>
</tr>
</tbody>
</table>

Usual autoconvolution HT parameter psfs are like autocorrelations of non-negative functions. The essential idea behind CHough is that sidelobe mean and variance can be reduced by casting negative votes as well as positive ones. We desire the positive votes all to come together at the peak. Thus they should be in the same locations as in autoconvolution HT. However, the feature point psf should also have negative entries which will overlap and cancel off-peak votes (Figure 6.1).

There are many "classical" ways to sharpen peaks in accumulator space. Some, such as use of gradient information to reduce feature psf size, maintain the linearity of the system; others (such as only voting where two or more feature psfs overlap) are nonlinear. In general, nonlinear schemes that reward consistency (or punish inconsistency) in evidence for a single instance are not appropriate for detection of multiple instances. CHough is linear, and is thus most potentially valuable when many instances may appear "close together" in parameter space. CHough is not a panacea—it is especially useful with the radially symmetric, or at least not violently directional, parameter psfs, for here maximum off-peak cancellation occurs. CHough will probably not be effective against the problems of ridge-like peaks in the \((\rho, \theta)\) and \((m, b)\) parameter spaces for line-finding. It may perform better in the \((x, y, \theta)\) space (Section 4.3). For \((x, y, \theta)\) space a feature psf of a line of positive votes with an extended surround of negative ones should serve to cancel most off-peak votes.

Not only does zero mean sidelobe height and reduced variance aid in peak detectability, it means that the peak is an unbiased estimator of the amount of evidence contributing to the peak. Zero-mean sidelobes mean that simple thresholding techniques can detect peaks. Preliminary experiments show that CHough parameter psfs have lower variance than regular autoconvolution Hough parameter psfs.
The continuous abstraction of CHough is not particularly interesting. The cancellation may be made as exact as is wished, since the continuous nature of the abstraction allows the cancelling negative votes to be arbitrarily near the positive ones. The more interesting cases are the discrete digital parameter psfs and their behavior on real data (see [Brown and Curtiss 1982]). The form of effective feature psfs (voting patterns) is also a topic of considerable interest and current research.
APPENDIX: Deriving the Surface Area of N-sphere Caps

Set up an n-dimensional coordinate system \((\rho_{n-1}, \varphi_1, \varphi_2, \ldots, \varphi_{n-1})\) (Figure 4.2). Notice

\[
\begin{align*}
  x_1 &= \rho_1 \cos \varphi_1 \\
  x_2 &= \rho_1 \sin \varphi_1 \\
  \varphi_{i-1} &= \rho_i \cos \varphi_i \\
  x_{i+1} &= \rho_i \sin \varphi_i, \quad 1 < i < n, \ -\pi/2 \leq \varphi_i \leq \pi/2, \ \rho_i \geq 0
\end{align*}
\]

Let \(S_1\) be the unit n-sphere \(\rho_{n-1} = 1\). Let \(S_2\) be the sphere of radius \(r\) centered at \(\varphi_{n-1} = \pi/2\) (at \((0,0,\ldots,0,1)\)). The spheres intersect where

\[
x_1^2 + x_2^2 + \ldots + x_n^2 = 1
\]

and

\[
x_1^2 + x_2^2 + \ldots + (x_{n-1})^2 = r^2,
\]

or at

\[
x_n = (2-r^2)/2 = \rho_{n-1} \sin \varphi_{n-1} = \sin \varphi_{n-1}.
\]

Then the area of the cap cut off is that with

\[
\pi/2 \leq \theta \leq \theta_0 = \arcsin(\varphi_{n-1}).
\]

Now to compute the area of the cap. A small patch of area bounded by "longitude" and "latitude" lines.

\[
dA_n = \rho_1 d\varphi_1 \rho_2 d\varphi_2 \ldots \rho_{n-1} d\varphi_{n-1}.
\]

Expanding the Jacobean,

\[
dA_n = (\rho_{n-1} \cos \varphi_{n-1} \cos \varphi_{n-2} \ldots \cos \varphi_2) d\varphi_1 \ldots
\]

\[
(\rho_{n-1} \cos \varphi_{n-1} \cos \varphi_{n-2} \ldots \cos \varphi_3) d\varphi_2 \ldots
\]

\[
\ldots
\]

\[
\rho_{n-1} d\varphi_{n-1}.
\]

\[
dA_n = \rho_{n-1}^{-n-1} \cos^{n-2} \varphi_{n-1} \cos^{n-3} \varphi_{n-2} \ldots \cos \varphi_2 d\varphi_1 d\varphi_2 \ldots d\varphi_{n-1}.
\]
Since $\rho_{n-1} = 1$,

$$A_n = \int_{\theta_1}^{\theta_2} \cdots \int_{\varphi_{n-2}}^{\varphi_{n-1}} dA$$

$$A_n = \int_{\theta_1}^{\theta_2} \cos^{n-2} \varphi_{n-1} d\varphi_{n-1} \int_{\varphi_{n-2}}^{\varphi_{n-1}} \cos^{n-3} \varphi_{n-2} d\varphi_{n-2} \cdots \int_{\varphi_2}^{\varphi_3} \cos \varphi_2 d\varphi_2$$

Now recall

$$\int \cos^n \varphi d\varphi = 1/n \sin \varphi \cos^{n-1} \varphi + (n-1)/n \cos^{n-2} \varphi d\varphi , \quad n \geq 2$$

$$\int \cos \varphi d\varphi = \sin \varphi + C$$

$$\int d\varphi = \varphi + C.$$ 

So

$$\int_{-\pi/2}^{\pi/2} \cos^n \varphi d\varphi = \left[ ((n-1)(n-3) \cdots 2)/(n(n-2) \cdots 1) \right] \cdot 2 \quad \text{n odd}$$

$$= \left[ ((n-1)(n-3) \cdots 1)/(n(n-2) \cdots 2) \right] \cdot \pi \quad \text{n even}$$

$$= 2 \quad n = 1$$

Thus if

$$B = \prod_{i=1}^{n-3} \int_{-\pi/2}^{\pi/2} \cos^i \varphi d\varphi = (2((n-2)/2!)(n-3)/2!)/((n-3)(n-5) \cdots )$$

Then

$$A_n = 2\pi B \int_{\theta_1}^{\theta_2} \cos^{n-2} \theta d\theta$$

Finally, to see how the area of the cap varies with increasing $r$, compute

$$dA_n/dr.$$
Notice

\[ y = \int_{\theta_0}^{\pi/2} \cos^{n-2} \theta d\theta \]

\[ \frac{dy}{dr} = -\cos^{n-2} \theta_0 \left( \frac{d\theta_0}{dr} \right) \]

\[ \sin \theta_0 = 1 - r^2/2 \]

\[ \cos \theta_0 d\theta_0 = -r \]

\[ \frac{dy}{dr} = r \cos^{n-3} \theta_0 \]

\[ \frac{dA_n}{dr} = 2\pi r \cos^{n-3} \theta_0 B. \]

Since

\[ \cos^2 \theta_0 = 1 - \sin^2 \theta_0 = r^2 - (r^4)/4, \]

\[ \cos \theta_0 = r \sqrt{1 - r^2/4}, \]

\[ \frac{dA_n}{dr} = 2\pi r B \left( r \sqrt{1 - r^2/4} \right)^{n-3}. \]
References


Welti, G.R., "Quarternary codes for pulsed radar," IRE Trans. IT-6, 3, 400-408, June 1960.
Figure Captions

Figure 1.1: (a) A circle in x-y image space with four feature points. (b) Four feature psfs in x-y center-location space.

Figure 1.2: (a) A general shape and a distinguished point O. A feature point x on the shape does not know where it is on the boundary, and must vote for one location of O for each possible boundary point it could be. (b) The feature psf of x shows the possible locations for O, and is the (boundary of the) shape with directions reversed.

Figure 1.3: Autoconvolution (shape-location) HT as pinhole camera imaging.

Figure 2.1: (a) A parameter psf for circle location, radius = 10. (b) Radius = 40, values linearly scaled with saturation. Values 0-24: intensities 0-255; values above 24: intensity 255.

Figure 2.2: The parameter space image of 6 identical image shapes. The sidelobes contribute both bias and variance (inherent noise) to the peak heights.

Figure 3.1: Some of the infinite number of circles superposed to form the parameter psf of the circle.

Figure 3.2: The parameter psf is related to the amount of circular arc cut off by an annulus.

Figure 3.3: Geometry for 2-D and 3-D parameter psf calculation.

Figure 3.4: Graphs showing the normalized height (as function of radius) for n-dimensional parameter psfs for unit hypersphere location.

Figure 4.1: Two feature psfs for points p1 and p2 on the line L. All such psfs are n-1 spheres in n-space, their centers on another linear subspace C parallel to L half as far from the origin.

Figure 4.2: A variant of the n-dimensional spherical coordinate system.

Figure 4.3: The geometry for hyperplane parameter psf calculation.

Figure 4.4: Behavior of the parameter psf for lines.

Figure 4.5: Feature point psfs from the infinite line contributing to the parameter psf for an ideal line (1-D hyperplane in 2-D.)
Figure 4.6: (a) Points on a line. (b) The points' contribution to the ideal parameter psf in (m, b) space, and an idea of how the psf looks. (c) The psf height varies as the \((\sec^2)\) of the direction.

Figure 4.7: (a) A redundant three-parameter space for line detection. (b) Sections through the space for several \(\theta\) showing detection of a vertical line. The "peak" in this redundant space is a line. (c) The density of the psf falls off with \(\theta\) as the cosecant of the angular difference.

Figure 5.1: (a) A Shape. (b) Its slope density function, which also gives the density of its autoconvolution parameter psf sidelobes in all directions near the peak.

Figure 6.1: CHough: (a) A discrete feature psf for circle location, radius = 10. (b) Resulting parameter psf. (c) Parameter psf, radius = 40 (feature psf is ring of 1's surrounded by rings of \(-\frac{1}{2}\)'s; intensity scaling as in 2.1c). The 1's in the feature psfs are the feature psfs for the usual autoconvolution HT. Compared to Figure 2.1, (b) and (c) show the sidelobe-diminishing properties of CHough.
SHAPE  "PINHOLE"  "IMAGE"

Figure 1.3
Figure 2.1
Figure 2.2
Figure 4.5

Figure 4.6
Figure 4.7
Figure 6.1
Figure 6.1 (continued).

...v

2