# A Probabilistic Approach to the Hough Transform

**R S Stephens**
Computing Devices Eastbourne Ltd,
Kings Drive, Eastbourne,
East Sussex, BN21 2UE

*It is shown that there is a strong relationship between the Hough Transform and the Maximum Likelihood method. The Probabilistic Hough Transform (PHT), a mathematically "correct" form of the Hough Transform, is defined as the log of the probability density function of the output parameters.*

*A model of feature error characteristics is proposed, combining normally distributed measurement errors with uniformly distributed correspondence errors.*

*A PHT is illustrated using the familiar problem of finding straight lines from oriented edgels, and it is shown that the conventional Hough method gives a good approximation to the PHT. In situations where there are many unknown parameters, however, conventional methods do not perform well, and here the PHT does provide an effective alternative.*

In this paper a mathematical formulation of the Hough Transform [1] is derived from a treatment of the measurement of input features as a stochastic process. This theoretical approach was motivated by a desire to understand better the processes involved in the Hough Transform, and to improve the performance of the Hough Transform in solving problems with many unknown parameters.

The mathematical principles come from the "Maximum Likelihood Method" [2, 3], used in probability theory for the determination of distribution parameters from experimental data. The Maximum Likelihood analysis leads to the definition of the Probabilistic Hough Transform, which is a likelihood function. If certain assumptions are made about the error characteristics, the PHT is very close to conventional Hough Transforms. If, in a particular application, these assumptions are a reasonable approximation, good results are usually obtained using standard Hough methods. However, where these assumptions are far from the truth, the Hough Transform will not work well, and steps should be taken to improve the model of input feature errors, such as filtering the Hough space, or incrementing an extended region instead of just the voting space. As a last resort, the full PHT can be computed, but this is much more computationally expensive than conventional Hough methods.

In contrast to the conventional Hough Transform, which is usually defined algorithmically and is intrinsically discrete, the PHT is defined as a continuous mathematical function. This allows any of the standard techniques for finding maxima to be employed.

## Maximum Likelihood Parameter Estimation

The general problem of parameter estimation is reviewed, putting it into the context of the Hough Transform. Whilst equation (5) is a standard result, its derivation is straightforward, and may be useful for those readers more familiar with the Hough Transform than with probability theory.

The PDF of pattern parameters gives a measure of the relative likelihood of the presence of each possible instance of the pattern in the image. As more image features are taken into account, this PDF will change from a uniform distribution, in which all possible patterns are equally likely, to a distribution with a maximum at the most likely set of pattern parameters.

The notation used is as follows: PDFs are denoted by $\mathbf{f}()$, $\vec{X}$ is a random variable representing an image measurement, and $\vec{x}$ is a specific image measurement. Similarly $\vec{Y}$ is a random variable representing a set of pattern parameters, and $\vec{y}$ is a specific point in Hough Space. The conditional PDF $\mathbf{f}(\vec{X} = \vec{x} \,|\, \vec{Y} = \vec{y})$, or $\mathbf{f}(\vec{x}\,|\vec{y})$ for short, is the PDF of $\vec{X}$ *given* the value of $\vec{Y}$.

The PDF $\mathbf{f}(\vec{x}\,|\vec{y})$ represents the error characteristics of the feature measurement — given the exact value of something you are trying to measure, it tells you the probability of each of the possible values of the measurement. A model of the feature measurement process is described below, but for now, it is assumed that $\mathbf{f}(\vec{x}\,|\vec{y})$ is a known function.

Let $X_i$ be the set of $i$ input features: $\{\vec{x}_1, \vec{x}_2, ...\vec{x}_i\}$, and let $\mathbf{f}_n = \mathbf{f}(\vec{y}\,|X_n)$ (i.e. the PDF in Hough space given $n$ features). Let the *a priori* PDF, $\mathbf{f}(\vec{y})$ be denoted by $\mathbf{f}_0$. Now, by Bayes theorem,

$$\mathbf{f}_n = \mathbf{f}(\vec{y}\,|X_n) = \frac{\mathbf{f}(X_n|\vec{y})\,\mathbf{f}(\vec{y})}{\mathbf{f}(X_n)} \quad (1)$$

$$= \mathbf{f}(\vec{x}_1|\vec{y})\,\mathbf{f}(\vec{x}_2|\vec{y},\vec{x}_1)\,\mathbf{f}(\vec{x}_3|\vec{y},\vec{x}_1,\vec{x}_2)\cdots\frac{\mathbf{f}(\vec{y})}{\mathbf{f}(X_n)} \quad (2)$$

Note that $\vec{x}_{i+1}$ is not independent of $X_i$, but that

$$\mathbf{f}(\vec{x}_{i+1}|\vec{y}, X_i) = \mathbf{f}(\vec{x}_{i+1}|\vec{y}) \quad (3)$$

In other words, previous features do affect the likely values of the next feature, but if the actual value of $\vec{y}$ is



BMVC 1990 doi:10.5244/C.4.12

given, then previous features do not add any additional information. Hence,

$$\mathbf{f}_n = \prod_{i=1}^{n} [\mathbf{f}(\vec{x}_i|\vec{y})] \frac{\mathbf{f}(\vec{y})}{\mathbf{f}(X_n)} \tag{4}$$

Since $\mathbf{f}(X_n)$ is constant with respect to $\vec{y}$, it may be absorbed into an arbitrary constant $C$, giving:

$$\mathbf{f}_n = C\mathbf{f}_0 \prod_{i=1}^{n} \mathbf{f}(\vec{x}_i|\vec{y}) \tag{5}$$

If $\mathbf{f}_0$ is uniform, which is often a reasonable assumption, then it too may be absorbed into the arbitrary constant.

Thus, the combined probability density function should be formed by taking the *product* of the PDFs from the individual features. Naturally, this may be computed by summing the logs of the PDFs. Now the log of the PDF is a function with a high value where the pattern parameters are consistent with the input feature, whilst the Hough Transform is the sum of a set of binary (0 or 1) functions where the voting space (in which the function has the value 1) corresponds directly to the regions of highest probability in the PDF. Clearly there is a strong relationship between the maximum likelihood method just described and the Hough Transform.

# Definition of the PHT

**Definition 1** *The Probabilistic Hough Transform $H(\vec{y})$ is defined as the log of the PDF of the output parameters, given all available input features:*

$$H(\vec{y}) = \ln [\mathbf{f}(\vec{y}|\vec{x_1}, \vec{x_2}, ...\vec{x_n})] \tag{6}$$

*Where n is the number of input features.*

From equation (5),

$$H(\vec{y}) = \sum_{i=1}^{n} \ln [\mathbf{f}(\vec{x}_i|\vec{y})] + \ln [\mathbf{f}_0] + C \tag{7}$$

where C is an arbitrary constant.

Note that this is a definition of the Hough Transform as a continuous scalar function in parameter space, and all the mathematical tools available for manipulating functions become available. The Hough Transform is normally defined algorithmically, and results in a discrete array of integers, which is not nearly so convenient for this kind of analysis.

Suppose $\mathbf{f}(\vec{x}|\vec{y})$ is taken to be function having a uniform non-zero background level, with a higher uniform level in the cells corresponding to the voting space (the log of this function is also binary). By taking out a constant term and scaling, this is equivalent to the zero-or-one increment that is used in the conventional Hough Transform. Whenever a conventional HT is performed, there is an implicit assumption that the PDF has this binary form, and that the *a priori* PDF is uniform (i.e. all output patterns are equally likely). Note that this implicit model of feature errors includes a uniformly distributed background region, which allows features to be subject to large errors (such as errors of correspondence). This is an important characteristic of the Hough Transform, and is what gives the HT its legendary robustness.

Clearly, however, this model of feature errors could be improved upon, and this is discussed in the next section.

# A Model of Feature Error Distributions

It is proposed that the PDF of image features commonly takes the form:

$$\mathbf{f}(\vec{x}|\vec{y}) = A + B \exp\left(-\frac{f(\vec{x},\vec{y})^2}{2\sigma^2}\right) \tag{8}$$

where the mapping between feature space and Hough space is expressed as $f(\vec{x},\vec{y}) = 0$, and $A$ and $B$ are weighting factors. In other words, the PDF is a weighted sum of a uniform distribution and a normal distribution. For example, given that a certain line is present in an image, an edgel is taken at random. There is a certain probability that the edgel will have nothing at all to do with the given line, so its PDF is uniform, but there is some probability that the edgel does belong to the given line, in which case it is reasonable to assume that its errors are normally distributed. Weiss [4] proposed this model for the distribution of errors in edge points in a line-fitting application, but it can be applied to almost any kind of image feature.

The importance of the uniform component of the PDF cannot be over-emphasized. If it is left out, the resulting PDF of output parameters will simply be a normal distribution, and the result of the Hough Transform will be the same as the least squares solution. Consider the simple case of estimation a single parameter $y$, having made a number of direct measurements $x$ of that parameter. If it is assumed that each measurement suffers only a normally distributed error, with standard deviation $\sigma$, then the maximum likelihood estimator of $y$ is the mean of the measurements $x$. This is due to the fact that the product of any two Gaussians is simply another Gaussian, and is the well known result that justifies the use of a least squares solution where errors can be assumed to be normally distributed.

If, on the other hand, it is accepted that the measurements may contain errors of correspondence, uniformly distributed within some range, the PDF (within the allowed range of $y$) is:

$$\mathbf{f}(x|y) = \frac{1-p}{y_{\max} - y_{\min}} + \frac{p}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-y)^2}{2\sigma^2}\right) \tag{9}$$

and zero elsewhere. The product of a number of terms of this form is not a simple expression, as it is in the case of the normal distribution. In fact, if you go to the trouble of multiplying out some of the terms, it is apparent that the combined PDF is the sum over all possible combinations of any number of input features, of the normal distribution representing the mean of that combination, weighted by the probability that the combination contains correctly corresponding features.

## Example of the PHT

The familiar case of finding straight lines from oriented edgels is used as an example of the computation of the PHT. It is emphasized that this is not proposed as an efficient method for finding lines, but as an exercise that gives some insight into the PHT. The probability density function of errors present in the edgels is assumed to be normally distributed in both lateral and orientation error, together with a uniform distribution of correspondence error. It is also assumed that all possible lines are equally likely, so $\mathbf{f}_0$ is uniform.

Let the edgel, $\vec{x}$, have position $[X, Y]$ and orientation $\alpha$, and let the line, $\vec{y}$, have intercept $a$ and orientation $\theta$. The Mapping from feature space to Hough space is given by:

$$f(\vec{x}, \vec{y}) = (\alpha - \theta)^2 + (X \sin\theta + Y \cos\theta - a)^2 = 0 \quad (10)$$

The PDF $\mathbf{f}(\vec{x}\,|\vec{y})$ is split into two components:

$$\mathbf{f}(\vec{x}\,|\vec{y}) = A + B\mathbf{f}'(\vec{x}\,|\vec{y}) \quad (11)$$

within the range $\pi > \theta \geq 0$, $|a| < a_{\max}$, and zero otherwise. The normal component, $\mathbf{f}'(\vec{x}\,|\vec{y})$, is given by:

$$\mathbf{f}'(\vec{x}\,|\vec{y}) \propto \exp\left(-\frac{\epsilon^2}{2\sigma^2} - \frac{\phi^2}{2\rho^2}\right) \quad (12)$$

where $\epsilon = X \sin\theta + Y \cos\theta - a$ (the lateral error), $\phi = \theta - \alpha$ (the orientation error), and $\sigma$ and $\rho$ are the standard deviations of lateral and orientation errors respectively.

To find the weighting factor between the uniform and the normal components of the PHT, it will be necessary to know the constant of proportionality in (12). Since the integral under a PDF is, by definition, unity, it is necessary to determine the volume under the function on right hand side of (12). It may be shown that, to a first order approximation,

$$\int_0^\pi \int_{-\infty}^{+\infty} \exp\left(-\frac{\epsilon^2}{2\sigma^2} - \frac{\phi^2}{2\rho^2}\right) da\, d\theta \approx 2\pi\sigma\rho \quad (13)$$

and the volume under the uniform component of (11) is $2\pi a_{\max}$. If $p$ is the probability that an edgel chosen at random corresponds to a particular line, then the combined PDF, with appropriate weighting, is:

$$\mathbf{f}(\vec{x}\,|\vec{y}) = \frac{(1-p)}{2\pi a_{\max}} + \frac{p}{2\pi\sigma\rho} \exp\left(-\frac{\epsilon^2}{2\sigma^2} - \frac{\phi^2}{2\rho^2}\right) \quad (14)$$

Note that any inaccuracy in the volume under $\mathbf{f}'(\vec{x}\,|\vec{y})$ due to the first order approximation will have two effects: the volume under the final PHT will not be exactly one. This does not matter as it is only relative probabilities that we are interested in. Secondly, there will be a slight inaccuracy in the balance between the uniform and the normal error distributions, which has a marginal effect on the result.

The PHT can now be computed according to (7), using the following algorithm:

PHT for Lines:

In: A set of oriented edgels
Out: An array of samples of the PHT

Initialize all elements of the Hough space, $H(\vec{y})$ to zero

**for** each edgel $\vec{x}$:

**for** each element $\vec{y}$ of Hough space:

compute $\ln[\mathbf{f}(\vec{x}\,|\vec{y})]$ using equation (14) and add to $H(\vec{y})$

Note that this algorithm is much more computationally expensive than the conventional Hough Transform. For each edgel, a substantial floating point expression has to be evaluated at each point in Hough Space, whereas the conventional H.T. increments only one Hough Space cell for each edgel. Note also that the Hough Space has to be floating point, whereas in conventional Hough Transforms it is an integer array.

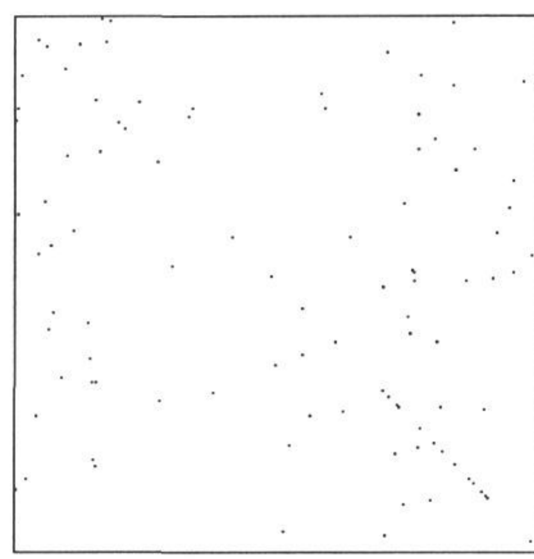
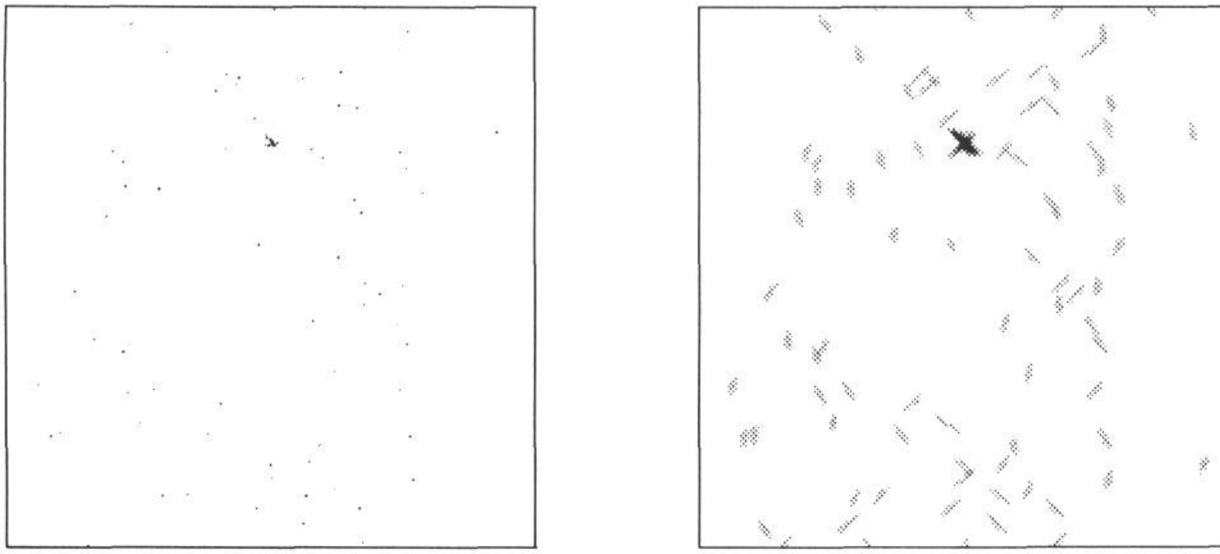


Figure 1: Synthetic edgels. 20% of the edgels lie on the leading diagonal

Figure 1 shows 100 randomly generated edgels. Each edgel has a probability $p = 0.2$ of belonging to a line on the leading diagonal. If it belongs to the line, its orientation error is normally distributed with $\sigma = 0.02$rad and its lateral position error is normally distributed with $\rho = 1.0$pixel. Otherwise the position and orientation are uniformly distributed. Figures 2(a) and 2(b) show the conventional and probabilistic HTs for this data. The values of $p$, $\sigma$ and $\rho$ used in the PHT are the exact values given above. Figures 3, 4(a) and 4(b) show the corresponding results for a real image. In the real data set, the conventional HT gives a better approximation to the PHT than in the synthetic data set, as can be seen in the topographic views in figures 5 and 6. This is be-



(a) Normal Hough Transform.

(b) Probabilistic Hough Transform.

Figure 2: Hough Transforms of synthetic data.

cause of the much larger number of edgels present in the real data set, the random scatter of the votes forming a better approximation to the PDF.

This shows that in this case, the conventional method for forming the Hough Transform is a good approximation to the "correct" method, provided there is a large number of input features, and the accumulator cell size is of the same order of magnitude as the measurement errors. However, in situations where the input features are very sparse, or where the accuracy of measurements is very low, the implicit assumptions made about the error distributions are not acceptable, and a technique based on better assumptions is required, which will give results closer to the PHT.

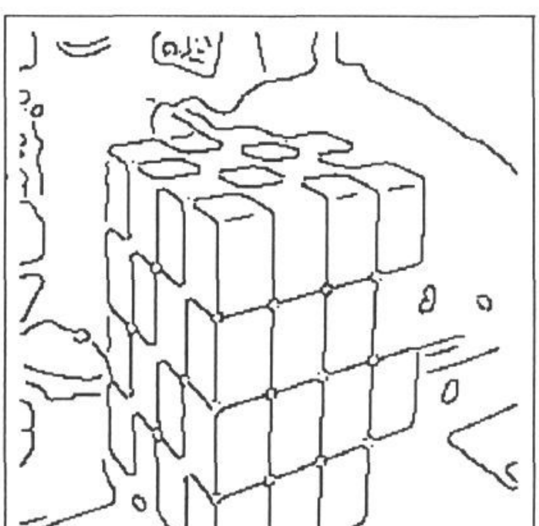


Figure 3: Real data: edgels from image of camera calibration cube.

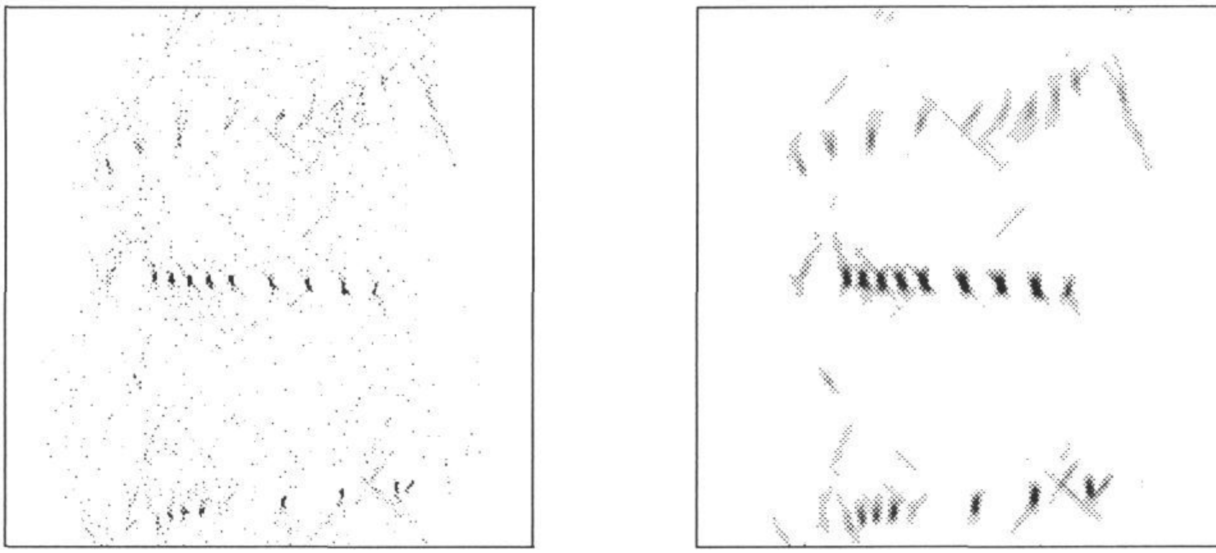
# Use of the PHT in High Dimensional Hough Spaces

Experience has shown that the normal methods of handling high dimensional Hough spaces suffer a degradation of robustness when the dimensionality exceeds four or five. Although the PHT is not a viable alternative to conventional methods in problems involving few unknown parameters, as the dimensionality increases, the PHT does have some definite advantages: In contrast to conventional HTs, the PHT is independent of the size, shape, and arrangement of accumulator cells. In dynamic Hough Transforms, in which the cells have a large coverage, these factors have a strong influence on the number of votes accumulated. In conventional Hough Transforms, there is a sharp cut-off between a cell receiving a vote or not, which generates something a bit like aliasing distortion. The smooth Gaussian roll-off of the PDF means that the PHT has a lower bandwidth and generates less noise.

Because the PHT is a mathematically defined continuous function, any of the standard techniques for locating maxima may be brought into use. Many of these require evaluation of the gradient, and this can be determined by algebraically differentiating the individual PDFs, and summing the resultant gradient vector. The need to sample the function in the neighbourhood of the point where the gradient is required is thus avoided.

One technique that has been used is Cauchy's Method of Steepest Ascent [5]. Although this method has the problem that it may converge to a local maximum rather than the global maximum, it has successfully been applied in a 6D Hough space for tracking the unconstrained motion of a rigid object [6, 7].

One difficulty encountered in using this algorithm is that the PHT tends to contain narrow ridges leading up to the



(a) Normal Hough Transform.

(b) Probabilistic Hough Transform.

Figure 4: Hough Transforms of real data.

maximum. Because each "leg" is at right angles to the previous one, this produces a large number of zig-zags in the climb. A small amount of random error (actually a quantization error) is added to each leg, giving a 60% reduction in the number of zig-zags. This is significant, as in a parallel implementation, there is a heavy overhead associated with the change of direction with each leg.

Any amount of Gaussian smoothing may be applied (with no overhead) by increasing the estimated standard deviation of measurement errors. Thus, it is easy to reduce the bandwidth of the function in the early stages of the search, which reduces the likelihood of converging to a local maximum, and allows the use of a larger step size. The actual standard deviation of the error specifies the minimum spatial scale that can be relied upon, and hence the minimum amount of smoothing that should be used in the final stage of the search.

The technique was tested using the experimental data described in [6], and showed much better robustness than the Fast Hough Transform of Li, Lavin and Le Master [8] which was previously used in this application. Whilst the method is more computationally expensive than that of Li, Lavin and Le Master, and the sequential climb is difficult to parallelize, it is not prohibitively expensive: the 3D tracking algorithm takes about one second per frame on a single T800-20 transputer.

## Conclusions

The relationship between the Hough Transform and Maximum Likelihood parameter estimation has been explored, and it has been shown that the Hough Transform *is* a maximum likelihood method. The Probabilistic Hough Transform is a Hough Transform based on a rigorous treatment of the parameter estimation problem.

The maximum likelihood method requires a model of feature error characteristics. There is a model that is implicit in conventional Hough Transform techniques. An important attribute of this model is that there is a non-zero "background" probability spread over the whole Hough space, which allows the presence of any number of arbitrarily large errors (such as errors of correspondence) with little or no effect on the solution.

Often the implicit model is a good approximation to the feature error characteristics, but in some cases, and in dynamic methods applied to high dimensional problems in particular, this model is to blame for the poor performance that has been experienced. A model consisting of a weighted sum of normal and uniform distributions has been found to give better results in these circumstances.

## References


[1] P V C Hough. *A Method and Means for Recognising Complex Patterns.* US Pattent 3069654, 1962.

[2] R A Fisher. The maximum likelihood method. *Messenger Math*, 41:155–160, 1912.

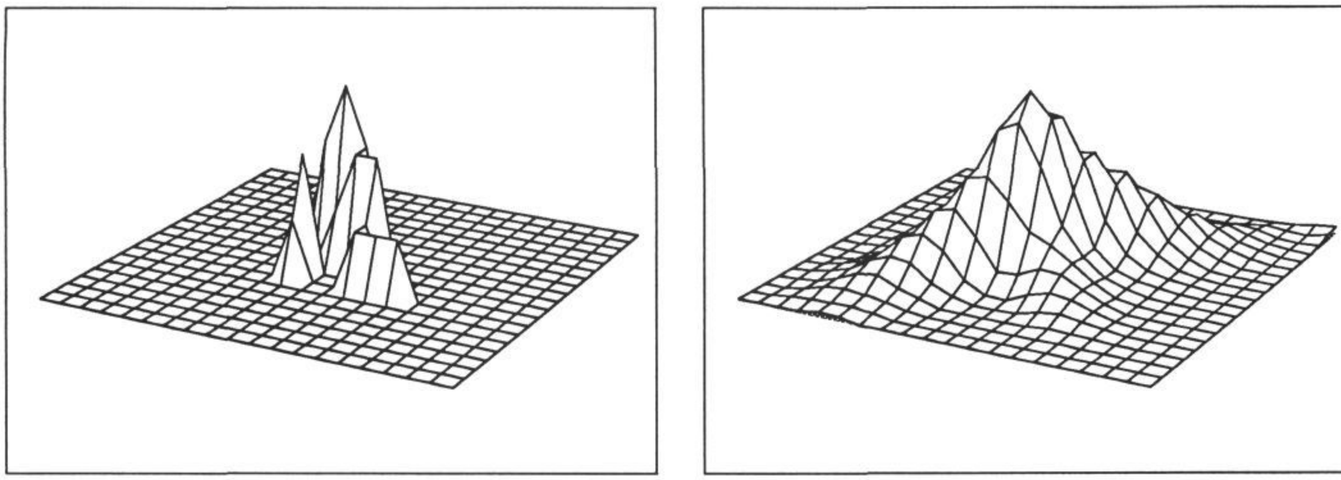
[3] Robert V Hogg and Allen T Craig. *Introduction to Mathematical Statistics*, pages 200–235. Collier Macmillan, 1978.

[4] Isaac Weiss. Line fitting in a noisy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(3):325–329, 1989.

[5] Erwin Kreyszig. *Advanced Engineering Mathematics*, pages 870–873. Wiley, 1983.

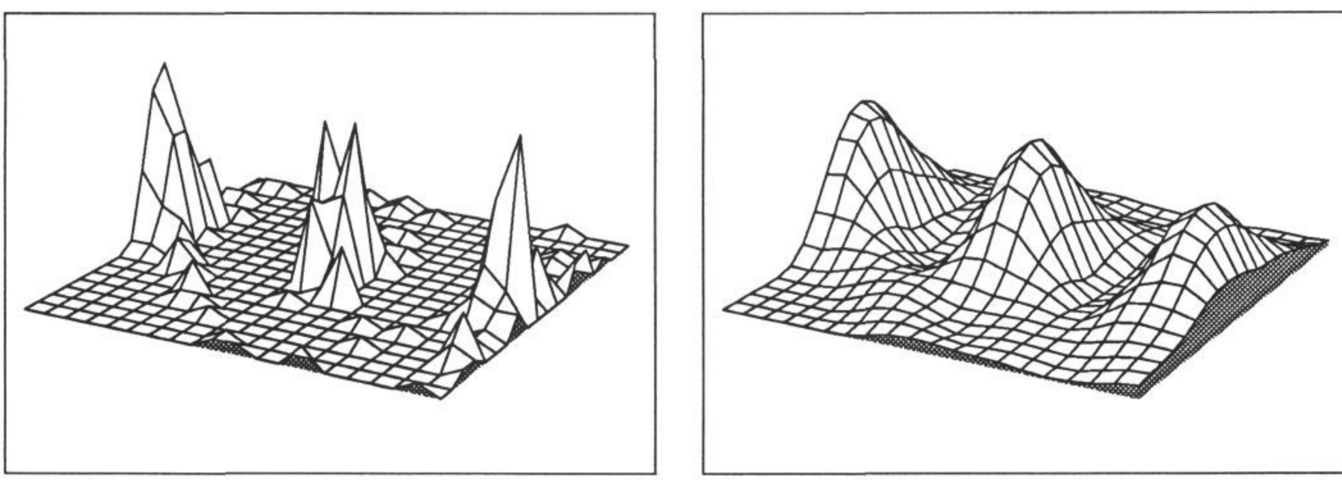
[6] R S Stephens. Real-time 3d object tracking. *Image and Vision Computing*, 8(1):91–96, February 1990.

[7] R S Stephens. *The Hough Transform — A Probabilistic Approach.* Forthcoming Ph.D. dissertation, Cambridge University, 1990.



(a) Normal Hough Transform. (b) Probabilistic Hough Transform.

Figure 5: Synthetic data Hough Transforms: topographic views.



(a) Normal Hough Transform. (b) Probabilistic Hough Transform.

Figure 6: Real data Hough Transforms: topographic views.

[8] Hungwen Li, Mark A Lavin, and Ronald J le Master. The Fast Hough Transform: A hierarchical approach. *Computer Vision, Graphics and Image Processing*, 36:139–161, 1986.