Guaranteed Convergence of the Hough Transform

Menashe Soffer and Nahum Kiryati

Department of Electrical Engineering

Technion - Israel Institute of Technology

Haifa 32000, Israel

Abstract

The straight-line Hough Transform using normal parameterization with a continuous voting kernel is considered. It transforms the collinearity detection problem to a problem of finding the global maximum of a two dimensional function above a domain in the parameter space. The principle is similar to robust regression using fixed scale M-estimation. Unlike standard M-estimation procedures the Hough Transform does not rely on a good initial estimate of the line parameters: The global optimization problem is approached by exhaustive search on a grid that is usually as fine as computationally feasible.

The global maximum of a general function above a bounded domain cannot be found by a finite number of function evaluations. Only if sufficient a-priori knowledge about the smoothness of the objective function is available, convergence to the global maximum can be guaranteed. The extraction of a-priori information and its efficient use are the main challenges in real global optimization problems.

Convergence in the Hough Transform is the ability to ensure that the global maximum is in the immediate neighborhood of the maximal grid point. More than thirty years after Hough patented the basic algorithm, it is still not clear how fine should the parameter space quantization be in order not to miss the true maximum.

In this paper conditions for the convergence of the Hough Transform to the global maximum are derived. The necessary constraints on the variability of the objective (Hough) function are obtained by using the saturated parabolic voting kernel and by defining an image model with several application dependent parameters. Random errors in the location of edge points and background noise are allowed in the model

and lead to statistical convergence guarantees. Significant intermediate results are obtained on the structure of the peak region and on the spatial statistics of noise voting in the continuous kernel Hough Transform. Convergence strategies are studied and the necessary parameter space quantization intervals are derived. Guaranteed focusing policies for multi-resolution Hough algorithms are developed. The application of the theoretic results to images that deviate from the image model is considered and exemplified.

Keywords: computer vision, covering methods, global optimization, Hough transform, M–estimation, pattern recognition, robust regression

1 Introduction

The Hough Transform [7] is a well known technique for detecting collinearities or other predefined patterns [1] in edge images. In this paper the straight-line Hough Transform using the normal line parameterization as suggested by Duda and Hart [3] is considered. The strength of the Hough Transform is in fitting straight lines to a collinear subset of a set of points that can include very many outliers, often an order of magnitude more than good data points. On the other hand, pure Hough algorithms rely on an assumption that the good data points are error-free. This is the complete opposite to standard total least squares line fitting that handles data point location errors very well but collapses in the presence of outliers.

In the Duda and Hart [3] formulation of the Hough algorithm, each of the N data points (x_i, y_i) is transformed to a sinusoidal voting pattern

$$\rho = x_i \cos\theta + y_i \sin\theta \tag{1}$$

in the (θ, ρ) normal parameter plane. Ideally, the sinusoids that correspond to a collinear subset of data points intersect at a single point in the parameter space. In practice, due to edge point location errors, this is not the case. Thrift and Dunn [25] suggested to alleviate the problem by transforming each data

point into a sinusoidal *band* that can be regarded as a smoothed sinusoid. This leads to a continuous function

$$h(\theta, \rho) = \sum_{i=1}^{N} c(\rho - x_i \cos\theta - y_i \sin\theta) \tag{2}$$

defined on a domain in the parameter plane. Parallel to the ρ axis, the profile of each sinusoidal band does not depend on θ, and is defined by the function $c(r)$ which is symmetric, continuous, non-increasing as a function of $|r|$, positive for $|r| < R$ and zero for $|r| \geq R$, where R is a positive width parameter. At any given values of θ and ρ, $h(\theta, \rho)$ represents the contributions of data points in a *voting strip* of area $V \approx 2RD$ where D is determined by the image boundaries and for a square image is at most the length of the diagonal. Thrift and Dunn named the function c *influence function*, a somewhat unlucky name since their procedure is known [11, 12, 19] to be similar to robust M-estimation [6, 16], with c playing a role similar to the ρ-function of the M-estimator and not to its influence function that has a different meaning. In recent publications the function c is referred to as the *voting kernel*.

The original collinearity detection problem is thus transformed to the problem of finding the global maximum of the function $h(\theta, \rho)$, which is referred to in the sequel as the Hough function or the objective function, in the domain

$$P = \{(\theta, \rho) \;\; | \;\; |\rho| \leq A \;,\; 0 \leq \theta < \pi\}, \tag{3}$$

where A is the Euclidean radius of the set of edge points. The Hough approach to solving this global optimization problem is, by efficient voting, to evaluate $h(\theta, \rho)$ on a rectangular grid of sampling points that is represented by an accumulator array, and to take the location of the sampling point in which $h(\theta, \rho)$ is largest as a sufficiently good estimate of the location of the true maximum. The accuracy of the approximation depends on the density of the sampling grid, i.e., on the size of the accumulator array, and is limited by the available memory and computing time. The estimate can in principle be refined by interpolation around the maximal sampling point [10] or by a focusing multi-resolution approach in which a finer sampling grid is placed around the maximal sampling point and another iteration of voting

and maximum detection takes place [8].

The basic assumption behind the Hough approach is that the sampling grid is sufficiently fine such that the largest sample of $h(\theta, \rho)$ is in the immediate neighborhood of the true global maximum. $h(\theta, \rho)$ may have very many local maxima: If R, the width parameter of the voting kernel, is small enough, each of the $O(N^2)$ pairs of edge points in the data set has, in principle, a corresponding local maximum. It is well known [26] that the global maximum of a general function cannot be found using a finite number of function evaluations, so there is no obvious guarantee that a given sampling grid represents $h(\theta, \rho)$ sufficiently well such that the true maximum is not missed.

Selecting the "right" size for the accumulator array is a major open problem in Hough Transform theory. Van Veen and Groen [27] studied it and provided useful guidelines in the context of the original Duda and Hart formulation of the algorithm. Valuable experimental results have been provided by Niblack and Petkovic [17]. In the continuous kernel Hough Transform there should be a close relation between the smoothness of $h(\theta, \rho)$ and the necessary density of the sampling grid. Kiryati and Bruckstein [10] obtained the 2-D effective band-region of $h(\theta, \rho)$ as a function of the effective bandwidth of the voting kernel $c(r)$, and suggested to use a voting kernel that is compact in both the space and frequency domains as allowed by the signal uncertainty principle. They determined the sampling intervals $\Delta\theta$ and $\Delta\rho$ that satisfy the Nyquist conditions and expressed the required number of accumulators in terms of the localization accuracy of the algorithm. However, even though the samples carry all the needed information, complicated interpolation or significant oversampling are needed in order to determine the location of the true maximum. Thus, so far a rigorous theoretical guarantee that a given sampling density is sufficient to ensure that the global maximum is in the immediate neighborhood of the maximal sampling point has not been available.

The objectives of the continuous kernel Hough algorithm and of robust regression using M-estimators are essentially similar [11, 12, 19, 20], except that in M-estimation the ρ-function is zero at $r = 0$ and

nondecreasing as a function of $|r|$, thus the problem becomes a global *minimization* problem. Also, in the continuous kernel Hough Transform fitting errors are usually measured perpendicular to the line in a total least squares fashion (for a treatment of the general anisotropic case see Kiryati and Bruckstein [12]), while in M-estimation errors are measured in the dependent variable. Furthermore, in M-estimation it is usually assumed that the scale parameter, i.e., the width of the ρ-function, is unknown, while in image analysis it can often be determined by the known error characteristics of the edge detector. M-estimation algorithms often solve the global minimization problem by iterative search that begins in a good initial approximation, thus convergence to the global minimum is obtained. In typical regression problems in the natural and social sciences the number of outliers is significantly smaller than the number of relevant data points, thus intelligent initial guesses can be made. This is usually not the case in image analysis, where the number of outliers due to noise and other objects in the image can be an order of magnitude larger than the number of points in a line that should be detected.

In this paper the continuous kernel straight line Hough Transform is studied in order to obtain explicit conditions for convergence to the global maximum, i.e., for association of the global maximum of the Hough function with the parameter space sampling point that is represented by the accumulator with the highest value. Several related problems concerning kernel selection and focusing policies in multi-resolution Hough algorithms are considered. Additional details can be found in reference [23].

2 Definitions and models

The main purpose of this work is to obtain conditions for convergence of the Hough Transform. This means that the the global maxima of $h(\theta, \rho)$ should be closely related to the accumulators with the largest values, i.e., the highest sampled values of the Hough function $h(\theta, \rho)$. The continuous voting kernel

assumed in the analysis is a saturated parabola

$$c(r) = \max(0, 1 - (r/R)^2) \tag{4}$$

that achieves a maximum of 1 at $r = 0$, is positive for $|r| < R$ and is otherwise zero. As demonstrated in reference [11], with a parabolic kernel the Hough transform approximates total least squares line fitting and is thus very sensitive to outliers. The saturated parabolic kernel provides robustness by hard-limiting the penalty due to outliers, but retains the total least squares behavior with respect to inliers. Due to the discontinuity of its first derivative, the saturated parabola is not a convenient ρ-function for use in M-estimation algorithms that are based on steepest-descent optimization. This difficulty does not arise in the Hough approach that does not rely on derivatives in finding the global maximum. See also reference [19]. For practical image analysis purposes the precise shape of the kernel has relatively little importance as long as it remains symmetric, continuous, non-increasing as a function of $|r|$, positive for $|r| < R$ and zero for $|r| \geq R$. The saturated parabola satisfies these requirements on one hand and makes analysis possible on the other.

Any convergence guarantee must rely on some a-priori information on the variability of $h(\theta, \rho)$. For example, given that $c(r)$ is a saturated parabola it can easily be shown that $h(\theta, \rho)$ is a Lipschitz function. The Lipschitz constants can be used to calculate sampling grid intervals that ensure convergence. That approach is valid and might even be useful in certain line fitting applications in which a large width parameter R is appropriate. In image analysis applications, where R is in the order of pixel size, the Lipschitz condition is a very weak a-priori constraint on the variability of the objective function $h(\theta, \rho)$ and would lead to a very fine sampling grid that is not useful in most applications due to computational constraints.

In order to obtain tighter constraints on the Hough function $h(\theta, \rho)$, certain assumptions on the data set, i.e. the edge image, must be made. In defining an image model, the familiar trade-off of generality

against complexity of the model and subsequent analysis is encountered. The edge image model defined and used in this research consists of a model of the linear edge which is a set of nearly collinear points, and a model of the background noise, i.e. the set of outliers.

2.1 Linear edge model

The linear edge model consists of geometric and statistical characterizations. Consider the set of N_l data points $\{(x_n, y_n)\}_{n=1}^{N_l}$ that belong to a linear edge. It has a centroid

$$(x_c, y_c) = \frac{1}{N_l} \sum_{n=1}^{N_l} (x_n, y_n)$$

and a principal axis that passes through the centroid and minimizes the second order moment of the set.

Construct the smallest rectangle that contains all the linear edge points, such that its geometric center coincides with the centroid of the set and its principal symmetry axis coincides with the principal axis of the point set. Let l and d denote the length and the width of the rectangle. $l_{\min}$, $l_{\max}$ and $d_{\max}$, respectively the lower and upper bounds on l and the upper bound on d that can be encountered in a certain application, are the geometric characterization of the linear edge. $d_{\min}$ is not used since a perfectly collinear set of edge points should always be acceptable.

The uniformity of the spatial distribution within the $l \times d$ bounding rectangle has a significant effect on the structure of $h(\theta, \rho)$. Let I_d and I_l denote the central second order moments of the linear edge points in the directions of the principal axes, i.e.,

$$I_d = \sum_{n=1}^{N_l} \epsilon_n^2$$

$$I_l = \sum_{n=1}^{N_l} \chi_n^2$$

where ϵ_n and χ_n are respectively the distances between (x_n, y_n) and the two principal axes. Define now $\sigma_d^2 = I_d/N_l$ and $\sigma_l^2 = I_l/N_l$. Then $\sigma_{l\,\mathbf{min}}^2$, $\sigma_{l\,\mathbf{max}}^2$ and $\sigma_{d\mathbf{max}}^2$, respectively the lower and upper bounds on σ_l^2 and the upper bound on σ_d^2 that can be encountered in a given application, are the statistical characterization of the linear edge.

Scatter variances are often used in spatial data analysis. Uniformly populated rectangles have also been used as linear edge models, see [13]. The unified model suggested here is much more relevant to the structure of noisy linear edges in actual images and also sufficiently powerful to enable useful analysis.

2.2 Background noise model

The background noise model states that independently of the linear edge there are points in the data sets that appear at random locations. In particular, the background noise is modeled as a Poisson point process. The probability of a single background noise point in an infinitesimal area dS is λdS, where λ is a parameter. The probability of more than one background noise point in an infinitesimal area is zero, and the numbers of background noise points in disjoint areas are independent random variables. Thus, the probability of N background points in an area S is

$$Prob(N) = \frac{(\lambda S)^N}{N!} e^{-\lambda S}. \tag{5}$$

The parameter λ controls the density of the outliers in the background model. Let $\overline{Q} \equiv \lambda V \approx \lambda 2RD$ be the expected number of background points in a voting strip. Assuming that V is constant, i.e., neglecting image edge effects, $\overline{Q}$ is used to characterize the background noise density in the image.

It must be admitted that the Poisson point process is not a perfect model for false edges produced by actual edge detectors. Due to filtering or clustering, a spurious edge element seldom comes alone. Spurious edge segments tend to cluster and create spurious edge segments that could be more harmful

than isolated noise points. This phenomenon can be seen in the edge image shown in Fig. 12 in the sequel, that contains real edge detector noise. However, even with the simplified Poisson point process model, the analysis of the Hough Transform is evidently non-trivial. We fear that a more complicated, realistic background noise model might lead to intractable mathematical analysis and to loss of intuition. This could be an interesting topic for future research.

3 Linear edge peak analysis

In this section the contribution of the linear edge data points to $h(\theta, \rho)$ is studied. This contribution is denoted $h_l(\theta, \rho)$. Assume that $R > d/2$, i.e., that the voting range exceeds the maximal distance of the edge point from the ideal line. Then there exists a domain $S_{\#}$ in the parameter plane in which all N_l linear edge data points contribute to $h_l(\theta, \rho)$. We first determine its boundary.

Let $\tilde{\rho}(\theta)$ denote the sinusoid that corresponds to the centroid of the linear edge data points in the Duda & Hart formulation of the algorithm:

$$\tilde{\rho}(\theta) \stackrel{\triangle}{=} x_c \cdot \cos\theta + y_c \sin\theta \tag{6}$$

Let $(\tilde{\theta}, \tilde{\rho}(\tilde{\theta}))$, or simply $(\tilde{\theta}, \tilde{\rho})$, denote the parameters of the principal axis of the linear edge points. Define relative parameters as

$$\delta\theta \equiv \theta - \tilde{\theta} \tag{7}$$

and

$$\delta\rho \equiv \rho - \tilde{\rho}(\tilde{\theta} + \delta\theta). \tag{8}$$

See Fig. 1.

Let $\Delta\rho_i(\theta)$ denote the difference, in terms of ρ, between a sinusoid $\rho_i(\theta)$ that corresponds to any of

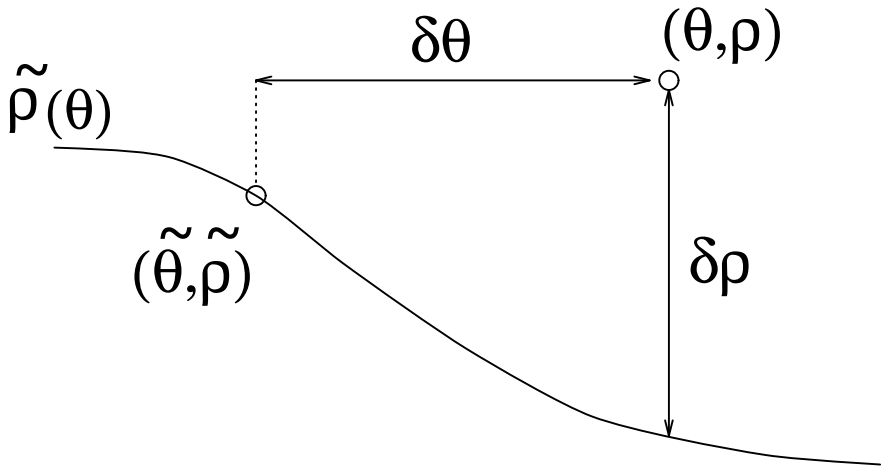

Figure 1: Relative parameters: geometric interpretation

the linear edge points and the sinusoid that corresponds to the centroid:

$$\Delta\rho_i(\tilde{\theta}+\delta\theta) \equiv \rho_i(\tilde{\theta}+\delta\theta) - \tilde{\rho}(\tilde{\theta}+\delta\theta). \tag{9}$$

The condition for a point in the parameter space whose relative coordinates are $(\delta\theta, \delta\rho)$ to be in $S_{\#}$, i.e. to receive contributions from all N_l linear edge points, is

$$\max_i |\Delta\rho_i(\tilde{\theta}+\delta\theta)| + \delta\rho \leq R \tag{10}$$

where the maximization is over all the linear edge points. But from geometric considerations,

$$|\Delta\rho_i(\tilde{\theta}+\delta\theta)| \leq \frac{d}{2}\cos|\delta\theta| + \frac{l}{2}\sin|\delta\theta|. \tag{11}$$

Thus, in terms of the relative parameters,

$$S_{\#} = \{(\delta\theta, \delta\rho) \;\; | \;\; |\delta\theta| < \delta\theta^* \;\; , \;\; |\delta\rho| < \delta\rho_{\#}(\delta\theta)\} \tag{12}$$

where $\delta\theta^* \geq 0$ is determined by

$$R = \frac{d}{2}\cos\delta\theta^* + \frac{l}{2}\sin\delta\theta^* \ , \tag{13}$$

and

$$\delta\rho_{\#}(\delta\theta) = R - \frac{d}{2}\cos|\delta\theta| - \frac{l}{2}\sin|\delta\theta|. \tag{14}$$

3.1 Inside $S_{\#}$

Let the contribution of the linear edge data points to the Hough function expressed in relative parameters be denoted by

$$h_{lr}(\delta\theta, \delta\rho) \equiv h_l(\tilde{\theta} + \delta\theta \ , \ \tilde{\rho}(\tilde{\theta} + \delta\theta) + \delta\rho). \tag{15}$$

Inside $S_{\#}$ all linear edge points provide positive contribution to $h_{lr}(\delta\theta, \delta\rho)$ according to the saturated parabolic voting kernel (Eq. 4), i.e.,

$$h_{lr}(\delta\theta, \delta\rho) = \sum_{i=1}^{N_l}\left[1 - \left(\frac{\epsilon_i}{R}\right)^2\right] \tag{16}$$

where ϵ_i is the distance between the linear edge point (x_i, y_i) and the line defined by the relative normal parameters $(\delta\theta, \delta\rho)$. Alternatively

$$h_{lr}(\delta\theta, \delta\rho) = N_l - \frac{I(\delta\theta, \delta\rho)}{R^2} \tag{17}$$

where $I(\delta\theta, \delta\rho)$ is the second order moment of the linear edge points with respect to that line. It is known that

$$I(\delta\theta, \delta\rho) = I_d\cos^2\delta\theta + I_l\sin^2\delta\theta + N_l\delta\rho^2, \tag{18}$$

so for $(\delta\theta, \delta\rho) \in S_{\#}$

$$h_{lr}(\delta\theta, \delta\rho) = N_l\left(1 - \frac{\sigma_d^2\cos^2\delta\theta + \sigma_l^2\sin^2\delta\theta + \delta\rho^2}{R^2}\right). \tag{19}$$

Thus, without outliers, close to the parameters of the best-fit line the shape of the peak of the Hough Transform is completely determined.

3.2 Outside $S_{\#}$ - upper bound

Within $S_{\#}$ the contribution of the linear edge to the shape of the peak in the parameter space is fully determined by the linear edge model parameters. Outside $S_{\#}$ only bounds on $h_{lr}(\delta\theta, \delta\rho)$ can be derived. An upper bound is derived first.

At a given location $(\delta\theta, \delta\rho)$, $h_{lr}(\delta\theta, \delta\rho)$ is determined by the arrangement $\{\Delta\rho_i(\tilde{\theta}+\delta\theta\}_{i=1}^{N_l}$ of distances from that location to the sinusoids corresponding to data set points. The linear edge model parameters impose the following three constraints on the arrangement:

1. The second order moment of the linear edge points with respect to the line whose relative normal parameters are $(\delta\theta, \delta\rho)$ is given by Eq. (18).

2. From the definition of the centroid of the linear edge pixels it follows that

$$\frac{1}{N_l}\sum_{i=1}^{N_l}\rho_i(\tilde{\theta}+\delta\theta) = \tilde{\rho}(\tilde{\theta}+\delta\theta). \tag{20}$$

3. As stated above, from geometric considerations

$$|\Delta\rho_i(\tilde{\theta}+\delta\theta)| \leq M_{\Delta\rho}(\delta\theta), \tag{21}$$

where $M_{\Delta\rho}(\delta\theta) \equiv \frac{d}{2}\cos|\delta\theta| + \frac{l}{2}\sin|\delta\theta|$.

The upper bound on $h_{lr}(\delta\theta, \delta\rho)$ is obtained by constructing an arrangement of distances $\{\Delta\rho_i(\tilde{\theta}+\delta\theta)\}_{i=1}^{N_l}$ that leads to the largest possible $h_{lr}(\delta\theta, \delta\rho)$ subject to these constraints. Let N_l' denote the cardinality of the subset of linear edge points that are within a distance R (voting range) from the line

whose relative normal parameters are $(\delta\theta, \delta\rho)$, and let $I'(\delta\theta, \delta\rho)$ denote the second order moment of the subset with respect to that line. The contribution of the $N_l - N_l^{'}$ remaining linear edge points to $I(\delta\theta, \delta\rho)$ is obviously $I(\delta\theta, \delta\rho) - I'(\delta\theta, \delta\rho)$. From Eq. (17) (with $I'(\delta\theta, \delta\rho)$ substituted for $I(\delta\theta, \delta\rho)$) it follows that the sought arrangement minimizes $I'(\delta\theta, \delta\rho)$.

Let Γ denote the set of arrangements in which all N_l' sinusoids that correspond to pixels within voting range pass through $(\tilde{\theta} + \delta\theta, \tilde{\rho}(\tilde{\theta} + \delta\theta) + \delta\rho)$ itself (i.e., contribute 1 to $h_{lr}(\delta\theta, \delta\rho)$), and all other $N_l - N_l^{'}$ sinusoids are at the maximal (absolute) distance $M_{\Delta\rho}(\delta\theta)$ from $\tilde{\rho}(\tilde{\theta} + \delta\theta)$. Thus, for an arrangement in Γ, at $\theta = \tilde{\theta} + \delta\theta$ sinusoids that correspond to the linear edge points have three possible ρ values:

- $\tilde{\rho}(\tilde{\theta} + \delta\theta) + \delta\rho$.
- $\tilde{\rho}(\tilde{\theta} + \delta\theta) + M_{\Delta\rho}(\delta\theta)$,
- $\tilde{\rho}(\tilde{\theta} + \delta\theta) - M_{\Delta\rho}(\delta\theta)$,

It can be shown that if there exists an arrangement in Γ that that satisfies the constraints, then it minimizes $I'(\delta\theta, \delta\rho)$ and is thus the sought arrangement. Indeed, for $|\delta\rho| < I(\delta\theta, 0)/|N_l \cdot M_{\Delta\rho}(\delta\rho)|$ an arrangement in Γ that satisfies the three constraints exists. In particular, satisfying the second order moment constraint dictates that the fraction of sinusoids in the first group must be

$$\alpha_I = \frac{|M_{\Delta\rho}(\delta\theta)|^2 - \frac{1}{N_l} I(\delta\theta, 0)}{|M_{\Delta\rho}(\delta\theta)|^2 - |\delta\rho|^2}. \tag{22}$$

Then the upper bound is $h_{lr}(\delta\theta, \delta\rho) \leq N_l \alpha_I$.

For $|\delta\rho| = I(\delta\theta, 0)/|N_l \cdot M_{\Delta\rho}(\delta\rho)|$, the arrangement in Γ that satisfies the constraints degenerates to just two groups, the first and one of the other two. As $|\delta\rho|$ further increases, no arrangement in Γ can satisfy the three constraints, and at least one of the two remaining groups must diverge. Then, using

Eq. (20) it can be shown that $h_{lr}(\delta\theta, \delta\rho) \leq N_l \alpha_{II}$, where

$$\alpha_{II} = \frac{|M_{\Delta\rho}(\delta\theta)|}{|M_{\Delta\rho}(\delta\theta)| + |\delta\rho|}. \tag{23}$$

To summarize,

$$h_{lr}(\delta\theta, \delta\rho) \leq N_l \begin{cases} \frac{|M_{\Delta\rho}(\delta\theta)|^2 - \frac{1}{N_l} I(\delta\theta, 0)}{|M_{\Delta\rho}(\delta\theta)|^2 - |\delta\rho|^2} & : \quad \delta\rho_{\#}(\delta\theta) < |\delta\rho| < \frac{\frac{1}{N_l} I(\delta\theta, 0)}{|M_{\Delta\rho}(\delta\rho)|} \\ \\ \frac{|M_{\Delta\rho}(\delta\theta)|}{|M_{\Delta\rho}(\delta\theta)| + |\delta\rho|} & : \quad |\delta\rho| > \frac{\frac{1}{N_l} I(\delta\theta, 0)}{|M_{\Delta\rho}(\delta\theta)|} \end{cases} \tag{24}$$

3.3 Outside $S_{\#}$ - lower bound

A rough lower bound for $h_{lr}(\delta\theta, \delta\rho)$ outside $S_{\#}$ can be obtained by replacing the saturated parabolic voting kernel (Eq. 4) by an unsaturated parabola $c'(r) = 1 - (r/R)^2$ and using Eq. (19). Thus

$$h_{lr}(\delta\theta, \delta\rho) > N_l \left(1 - \frac{\sigma_d^2 \cos^2 \delta\theta + \sigma_l^2 \sin^2 \delta\theta + |\delta\rho|^2}{R^2} \right) \tag{25}$$

4 Background noise analysis

In this section results concerning the contribution of the background noise data points to $h(\theta, \rho)$ are derived. This contribution is denoted $h_n(\theta, \rho)$. First, $h_n(\theta, \rho)$ at any given point (θ, ρ) is studied. Then, results on the difference $\Delta h_n \equiv h_n(\theta_1, \rho_1) - h_n(\theta_2, \rho_2)$ between the background noise contributions at different points in the parameter space are obtained. This is necessary in order to evaluate the effect of the background noise on the correct identification and exact location of the peak. Results are easier to obtain if (θ_1, ρ_1) is far away from (θ_2, ρ_2) since then the contributions at the two points are essentially statistically independent.

4.1 Preliminaries

Let $g(\theta, \rho)$ denote a certain transformation of a point process from an (x, y) image space to the (θ, ρ) normal parameter space. Generally,

$$g(\theta, \rho) = \sum_i k_g(\theta, \rho, x_i, y_i) \tag{26}$$

where $k_g(\theta, \rho, x_i, y_i)$ is the kernel of the transformation. For example, when $g(\theta, \rho) = h_n(\theta, \rho)$, i.e. the Hough transform of the noise points,

$$k_g(\theta, \rho, x_i, y_i) = c(\rho - x_i \cos\theta - y_i \sin\theta). \tag{27}$$

The Moment Generating Function (MGF) of the transformed point process is defined as

$$\phi_{g(\theta,\rho)}(s) = E[\exp\{g(\theta, \rho) \cdot s\}]. \tag{28}$$

The MGF of a transformed Poisson process can be obtained in two steps:

Step 1 Assume that just a single noise point exists. Then

$$\phi_{g(\theta,\rho)|1}(s) = \int_{(x,y)} e^{s \cdot k_g(\theta,\rho)} f(x, y|1) dx dy \tag{29}$$

where $f(x, y|1)$ denotes the probability density of the location of that noise point.

Step 2 The MGF is now given by:

$$\phi_{g(\theta,\rho)}(s) = e^{\overline{Q}[\phi_{g(\theta,\rho)|1}(s)-1]} \tag{30}$$

where $\overline{Q}$ is the expected number of noise points that contribute to $g(\theta, \rho)$. (If $g(\theta, \rho)$ is the Hough transform of noise, $\overline{Q}$ is the expected number of noise points in the voting strip).

For a proof see, e.g., [24].

The mean and the variance of $g(\theta,\rho)$ can be extracted from the MGF by:

$$\overline{g(\theta,\rho)} = \left.\frac{\partial \phi_{g(\theta,\rho)}(s)}{\partial s}\right|_{s\to 0} \tag{31}$$

$$\sigma^2_{g(\theta,\rho)} = \left.\frac{\partial^2 \phi_{g(\theta,\rho)}(s)}{\partial s^2}\right|_{s\to 0} - \overline{g(\theta,\rho)}^2 \tag{32}$$

If the noise level is significant, the distribution of $g(\theta,\rho)$ can be approximated to be Gaussian with the above mean and variance.

4.2 Noise contribution at a given point

Explicit expressions for the expectation and variance of the noise contribution at a given point in the parameter space can be obtained using the MGF approach as follows. The first step is to consider $h_{n|1}$, the contribution of a single noise point within the voting strip to $h_n(\theta,\rho)$. According to the voting kernel, $h_{n|1} = 1-(r/R)^2$ where r is a uniformly distributed random variable: $r \sim U[-R,R]$. Using Eq. (29), the MGF of $h_{n|1}$ is

$$\begin{aligned}\phi_{h_n|1}(s) &= E\left\{\exp\left[s(1-(r/R)^2)\right]\right\} \\ &= \frac{1}{2R}\int_{-R}^{R}\exp\left[s(1-(r/R)^2)\right]dr = \frac{e^s}{\sqrt{s}}\int_0^{\sqrt{s}} e^{-y^2}dy \qquad (s>0)\end{aligned}$$

The second step is to apply Eq. (30):

$$\phi_{h_n}(s) = \exp\left\{\overline{Q}\left[\frac{e^s}{\sqrt{s}}\int_0^{\sqrt{s}} e^{-y^2}dy - 1\right]\right\} \qquad (s>0) \tag{33}$$

The expected value and variance of $h_n(\theta,\rho)$ were obtained from Eqs. (31) and (32) using symbolic mathematics software. In terms of $\overline{Q}$, the expected number of background noise points in the voting strip,

the expected value of the noise contribution is

$$\overline{h_n} = \frac{2}{3}\overline{Q} \tag{34}$$

and the variance is

$$\overline{\sigma^2_{h_n}} = \frac{8}{15}\overline{Q}. \tag{35}$$

In principle, using the MGF, bounds on the probability that the noise contribution exceeds a given value can be derived. Significant effort is however needed in order to obtain tight bounds. By simulations it has been shown that for most practical purposes the distribution can be approximated by a Gaussian distribution with the above mean and variance. This approach has been followed in the sequel.

4.3 Noise difference: statistical independence

Let $\Delta h_n(\theta, \rho, \delta\theta, \delta\rho)$ denote the difference between the noise contribution in two points in the parameter space, defined as follows:

$$\Delta h_n \stackrel{\triangle}{=} h_n(\theta + \delta\theta, \rho + \delta\rho) - h_n(\theta, \rho). \tag{36}$$

Suppose that $\bar{Q}$ is similar for the two voting strips that correspond to (θ, ρ) and to $(\theta + \delta\theta, \rho + \delta\rho)$, and that the overlap between the voting strips is zero or small enough to justify a statistical independence assumption. The mean and variance of the noise difference can then be obtained either by considering it as the difference between two statistically independent and identically distributed noise contributions, or by applying the MGF approach to Δh_n itself. Note that in the latter approach, the kernel of the transformation $k_g(\theta, \rho, x_i, y_i)$ is

$$c(\rho + \delta\rho - x_i\cos(\theta + \delta\theta) - y_i\sin(\theta + \delta\theta)) - c(\rho - x_i\cos\theta - y_i\sin\theta).$$

As expected, the mean of the difference is zero

$$\overline{\Delta h_n} = 0 \tag{37}$$

and its variance is

$$\overline{\sigma^2_{\Delta h_n}} = \frac{16}{15}\overline{Q}. \tag{38}$$

4.4 Noise difference: small shift in ρ

Consider the noise difference $\Delta h_n = h_n(\theta, \rho + \delta\rho) - h_n(\theta, \rho)$, where $\delta\rho < 2R$. The two voting strips are in this case parallel with partial overlap. It is again assumed that $\bar{Q}$ is similar in the two voting strips, and also that the voting strips can be approximated by rectangles. The union of the voting strips is thus a rectangle of width $2R + \delta\rho$ and their intersection is a narrower rectangle $2R - \delta\rho$ wide. The relative location of a noise point that is in the union of the two voting strips can be characterized by a variable $r \in [-R, R + \delta\rho]$ as follows:

$r \in [-R, -R + \delta\rho]$ if the point contributes only to $h_n(\theta, \rho)$

$r \in [-R + \delta\rho, R]$ if it contributes to both

$r \in [R, R + \delta\rho]$ if it contributes only to $h_n(\theta, \rho + \delta\rho)$

The contribution of that point to the noise difference is:

$$\Delta h_{n|1} = \begin{cases} -c(r) = (r/R)^2 - 1 & : \quad -R < r < -R + \delta\rho \\ c(r - \delta\rho) - c(r) = (2r - \delta\rho)\delta\rho/R^2 & : \quad -R + \delta\rho < r < R \\ c(r - \delta\rho) = 1 - (r - \delta\rho)/R^2 & : \quad R < r < R + \delta\rho \end{cases}$$

Using Eq. (29), the MGF of $\Delta h_{n|1}$ is

$$\phi_{\Delta h_n|1}(s) = \frac{2R}{2R+\delta\rho}\int_{1-\delta\rho/R}^{1}\cosh[s(1-y^2)]dy + \frac{R^2}{\delta\rho\, s\,(2R+\delta\rho)}\sinh\frac{s\,\delta\rho\,(2R-\delta\rho)}{R^2} \tag{39}$$

The MGF of the noise difference Δh_n can now be obtained using Eq. (30). Note that the expected number of noise data points in the union of the voting strips is $\overline{Q}\cdot(2R+\delta\rho/2R)$.

$$\phi_{\Delta h_n}(s) = \exp\left\{\overline{Q}\left[\int_{1-\delta\rho/R}^{1}\cosh[s(1-y^2)]dy + \frac{R^2}{2s\,\delta\rho\,R}\sinh\frac{s\,\delta\rho\,(2R-\delta\rho)}{R^2} - \frac{2R+\delta\rho}{2R}\right]\right\} \tag{40}$$

Using Eqs. (31) and (32) with symbolic mathematics software, the expected value of the background noise difference is found to be zero

$$\overline{\Delta h_n} = 0, \tag{41}$$

and its variance is

$$\overline{\sigma^2_{\Delta h_n}}(\delta\rho) = \overline{Q}\left[\frac{4}{3}\left(\frac{\delta\rho}{R}\right)^2 - \frac{2}{3}\left(\frac{\delta\rho}{R}\right)^3 + \frac{1}{30}\left(\frac{\delta\rho}{R}\right)^5\right]. \tag{42}$$

4.5 Noise difference: small shifts in θ and ρ

Consider the noise difference $\Delta h_n = h_n(\theta+\delta\theta,\rho+\delta\rho) - h_n(\theta,\rho)$. Suppose that the two voting strips have a similar length D and that $\delta\theta$ is sufficiently small to yield significant overlap between the voting strips. We proceed to determine the expected value and variance of the noise difference.

Disregarding small edge effects, the voting strips can be considered rectangular. Parameterize each of the principal axes of the voting rectangles by a parameter t that assumes values in the range

$$-\alpha D < t < (1-\alpha)D, \quad 0 \le \alpha \le 1$$

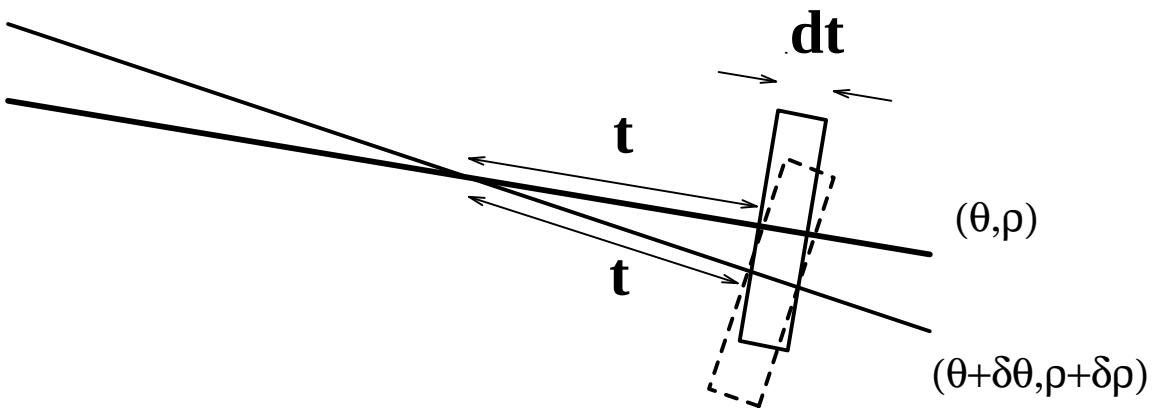

Figure 2: Dividing the voting rectangles to infinitesimal segments for noise difference analysis.

where D is the length of the voting rectangle and α is chosen so that the intersection of the two principal axes corresponds to $\alpha = 0$.

The voting rectangles can be divided into infinitesimal segments, see Fig. 2. Consider an infinitesimal segment centered at the point t on the principal axis of one of the voting rectangles, and the infinitesimal segment corresponding to the same parameter t in the other voting rectangle. The distance between the centers of the two infinitesimal segments is approximately $|\delta\theta \cdot t|$. If $\delta\theta$ is small enough, then the two segments can be approximated by two infinitesimal voting rectangles that correspond to two points in parameter plane that differ just in their ρ values, by $\delta\rho = |\delta\theta \cdot t|$.

The total noise difference is the sum of the noise difference contributions in all the pairs of infinitesimal voting rectangles. The expected value of each infinitesimal noise difference is zero, and its variance is given by Eq. (42) substituting $\overline{Q}_{\text{eff}} = \frac{\overline{Q}}{D}dt$ as the expected number of noise points in an infinitesimal segment and $\delta\rho = |t \cdot \delta\theta|$ as the ρ shift. The expected value of the total background noise difference is thus zero. The variance is

$$\overline{\sigma^2_{\Delta h_n}}(\delta\theta, \delta\rho) = \frac{\overline{Q}}{D} \int_{-\alpha D}^{(1-\alpha)D} \overline{\sigma^2_{\Delta h_n}}(|\delta\theta \cdot t|)dt \tag{43}$$

Where $\overline{\sigma^2_{\Delta h_n}}$ inside the integral is given by Eq. (42) and the dependence on $\delta\rho$ is implicit in the definition of α. For voting strips of similar length D that intersect at their middle, the variance of the noise difference

is

$$\overline{\sigma^2_{\Delta h_n}} = \begin{cases} \overline{Q}\left(\frac{x^2}{9} - \frac{x^3}{48} + \frac{x^5}{5760}\right) & \text{if } x < 4 \\ \overline{Q}\left(\frac{16}{15} - \frac{16}{9}\frac{1}{x}\right) & \text{if } x \geq 4 \end{cases}, \tag{44}$$

where $x \equiv \delta\theta D/R$. For other relative orientations of the voting strips, the variance of the noise difference can be similarly obtained.

4.6 Conclusions from the noise difference analysis

As can be expected from the properties of the Poisson noise process, the expected value of the noise difference between equally sized voting strips has been shown to be always zero. We explicitly derived the variance of noise difference for shifts in ρ and for shifts in θ. The dependence of the standard deviation of the noise difference on a shift in ρ (normalized by R) and on a shift in θ (normalized by R/D) is presented in Figs. 3 and 4 respectively. As the shifts increase, both graphs converge to $\sqrt{\frac{16}{15}}$ that corresponds to statistical independence (for $\overline{Q} = 1$).

In all cases it has been found out by simulations [23] that whenever the noise difference is large enough to become a significant interference, the distribution can be approximated the Gaussian distribution with the respective expected value and variance.

5 Continuous parameter plane analysis

The results on the shape of the linear edge peak and on the background noise have been used to solve the following problems:

- Disregarding noise, how can it be guaranteed that random arrangements of the linear edge data points within the voting strip will not lead to peaks that are higher than the "true" peak?
- To what extent can the random background noise move the peak from its true position?

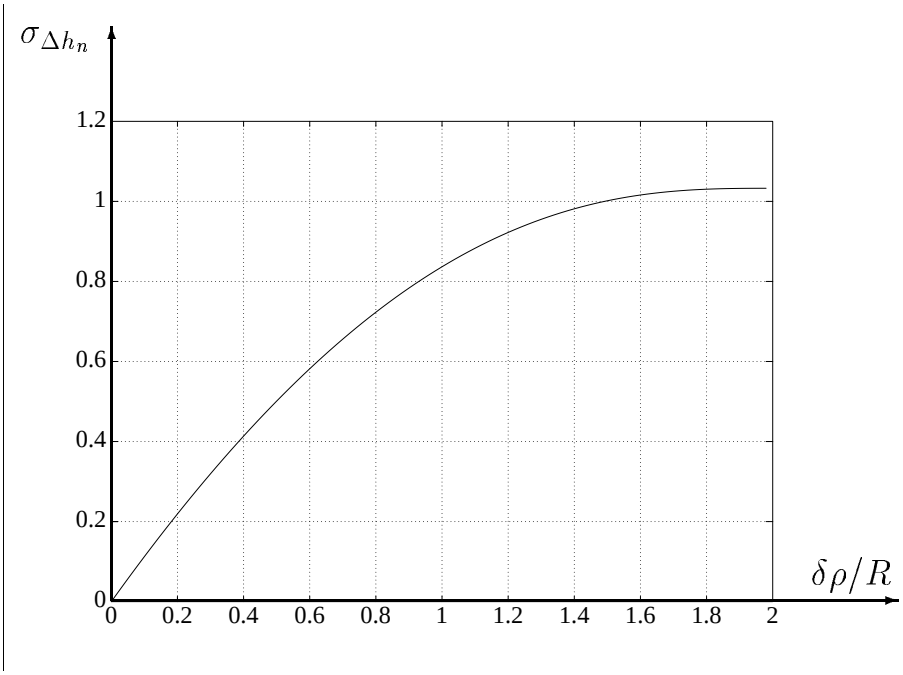

Figure 3: The standard deviation of the difference between the noise contributions at two points in the parameter space with similar θ coordinates and different ρ coordinates as a function of $\delta\rho/R$, for $\overline{Q} = 1$.

The answer to both questions depends greatly on R, i.e., on the width of the saturated parabolic voting kernel. A trade-off arises since, e.g., a narrow voting kernel reduces the effect of background noise but increases the sensitivity to linear edge point location errors.

In mathematical terms, the first problem is to ensure that

$$h_{lr}(\delta\theta, \delta\rho) < h_{lr}(0,0) \qquad \forall(\delta\theta, \delta\rho) \ \text{ outside } S_{\#}. \tag{45}$$

(Note that inside $S_{\#}$ the global maximum of $h_{lr}(\delta\theta, \delta\rho)$ is at $(\delta\theta, \delta\rho) = (0,0)$ and is the desired true peak). By analysis of $h_{lr}(\delta\theta, \delta\rho)$ outside $S_{\#}$ it can be shown [23] that even for $\sigma_d^2 = d^2/4$, the worst case in which all linear edge points are at the largest possible distance from the true line, selecting $R \approx 1.2d$ or greater ensures that condition (45) is satisfied. If the data points are uniformly distributed in the voting

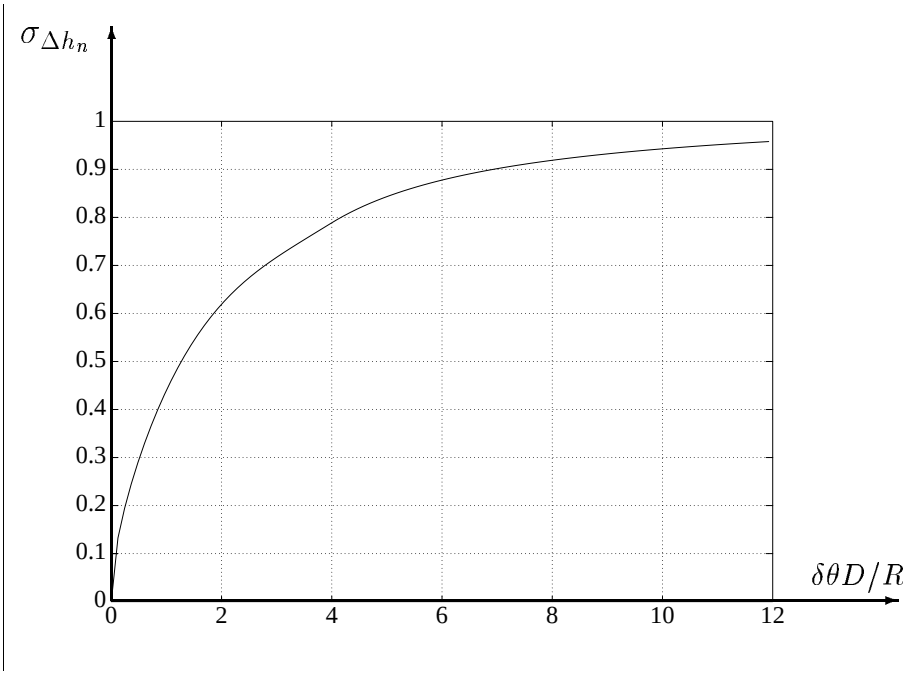

Figure 4: The standard deviation of the difference between the noise contributions at two points in the parameter space with similar ρ coordinates and different θ coordinates as a function of $\delta\theta D/R$, for $\overline{Q} = 1$.

strip, i.e., $\sigma_d^2 = d^2/12$, selecting $R \approx 0.8d$ or greater is sufficient. Let γ denote the "gain margin":

$$\frac{h_{lr}(\delta\theta, \delta\rho)}{h_{lr}(0,0)} < 1 - \gamma \qquad \forall(\delta\theta, \delta\rho) \ \text{ outside } S_{\#}. \tag{46}$$

The dependence of γ on R and σ_d^2/d^2 is described in Fig. 5.

Given that condition (45) is satisfied, the shape of the peak in the Hough Transform will still be distorted due to the presence of background noise in the edge image, so the maximum will be shifted. Formally, it is possible that for some $(\delta\theta, \delta\rho)$

$$h_{lr}(\delta\theta, \delta\rho) + h_{nr}(\delta\theta, \delta\rho) > h_{lr}(0,0) + h_{nr}(0,0), \tag{47}$$

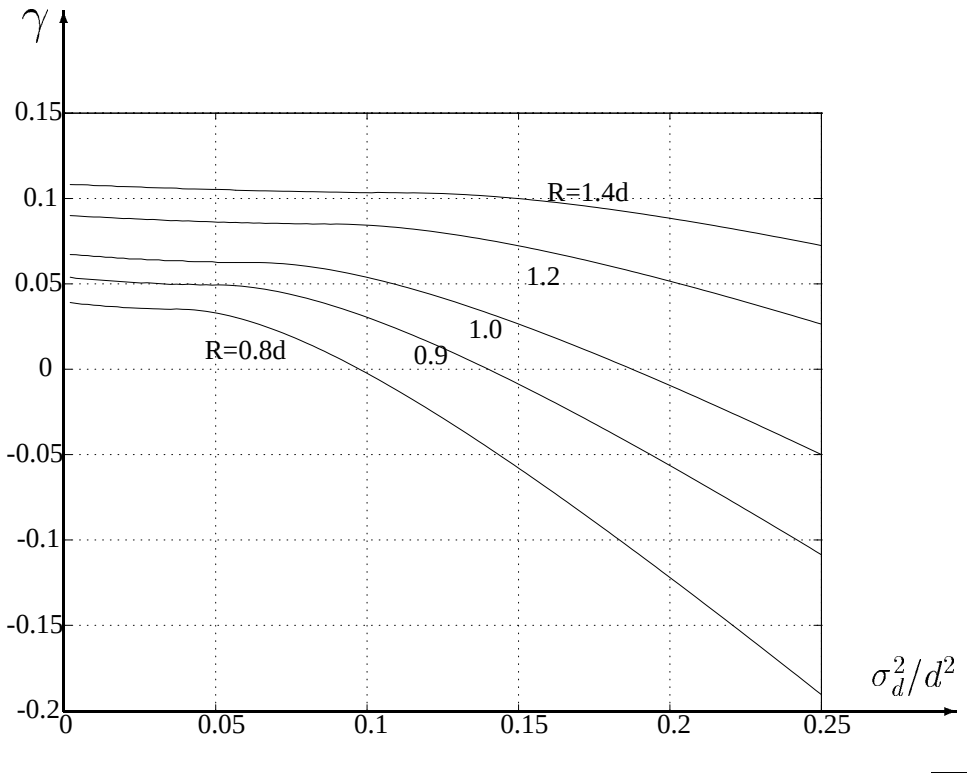

Figure 5: The dependence of the gain margin γ (vertical axis) on the relative lateral scatter of the linear edge points σ_d^2/d^2 (horizontal axis) and on the voting range R.

where h_{nr} is the background noise contribution in relative coordinates $(\delta\theta, \delta\rho)$. This will happen where

$$h_{nr}(\delta\theta, \delta\rho) - h_{nr}(0,0) > h_{lr}(0,0) - h_{lr}(\delta\theta, \delta\rho), \tag{48}$$

i.e., where the noise difference will out-balance the decay of the peak.

The deviation of the peak from its noise-free position can be in principle bounded by comparing the structure of the noise-free linear edge peak against the noise difference statistics. Define

$$\max\{h_{nr}(\delta\theta, \delta\rho) - h_{nr}(0,0)\}_\xi \stackrel{\triangle}{=} \xi\sqrt{\overline{\sigma^2_{\Delta h_n}}(\delta\theta, \delta\rho)} \tag{49}$$

to be the maximal noise difference at confidence level $erf(\xi)$. At that confidence level, the shifted peak will be within the boundary defined by

$$h_{lr}(0,0) - h_{lr}(\delta\theta, \delta\rho) = \max\{h_{nr}(\delta\theta, \delta\rho) - h_{nr}(0,0)\}_{\xi} \tag{50}$$

Solving Eq. (50) for the complete 2D boundary seems to be intractable. It is easier to obtain one dimensional upper bounds $\delta\theta_e$ and $\delta\rho_e$ on the errors in the location of the peak in the presence of background noise. These bounds respectively satisfy

$$h_{lr}(0,0) - h_{lr}(0, \delta\rho_e) = \max\{h_{nr}(0, \delta\rho_e) - h_{nr}(0,0)\}_{\xi} \tag{51}$$

and

$$h_{lr}(0,0) - h_{lr}(\delta\theta_e, 0) = \max\{h_{nr}(\delta\theta_e, 0) - h_{nr}(0,0)\}_{\xi}, \tag{52}$$

where the maximum is at confidence level $erf(\xi)$.

Consider for example the calculation of $\delta\theta_e$. The derivation of $\delta\rho_e$ is similar. From Eq. (19), inside $S_{\#}$, the *l.h.s.* of Eq. (52) is

$$h_{lr}(0,0) - h_{lr}(\delta\theta_e, 0) = \frac{N_l(\sigma_l^2 - \sigma_d^2)\sin^2\delta\theta_e}{R^2}. \tag{53}$$

σ_d^2 is usually much smaller than σ_l^2 and will be neglected. Since $\delta\theta_e$ is small,

$$h_{lr}(0,0) - h_{lr}(\delta\theta_e, 0) \approx \frac{N_l\sigma_l^2}{R^2}\delta\theta_e^2. \tag{54}$$

Suppose that the distribution of data points along the voting strip is uniform, so $\sigma_l^2 = l^2/12$. Then

$$h_{lr}(0,0) - h_{lr}(\delta\theta_e, 0) \approx \frac{N_l l^2}{12R^2}\delta\theta_e^2. \tag{55}$$

Consider now the *r.h.s.* of Eq. (52). At confidence level erf(ξ), the upper bound on the difference between the noise contributions at $(\delta\theta_e, 0)$ and at $(0,0)$ is

$$\max\{h_{nr}(\delta\theta_e, 0) - h_{nr}(0,0)\}_\xi = \xi\sqrt{\overline{\sigma^2_{\Delta h_n}}}, \tag{56}$$

where $\overline{\sigma^2_{\Delta h_n}}$ is given by Eq. (44). Thus,

$$\max\{h_{nr}(\delta\theta_e, 0) - h_{nr}(0,0)\}_\xi = \xi \begin{cases} \sqrt{\overline{Q}\left(\frac{x^2}{9} - \frac{x^3}{48} + \frac{x^5}{5760}\right)} & : \quad x < 4 \\ \\ \sqrt{\overline{Q}\left(\frac{16}{15} - \frac{16}{9}\frac{1}{x}\right)} & : \quad x > 4 \end{cases} \tag{57}$$

where

$$x \triangleq \delta\theta_e \cdot \frac{D}{R}. \tag{58}$$

Substituting Eqs. (55) and (57) in Eq. (52) and defining

$$SN_\theta \triangleq \frac{N_l l^2}{12\xi D^2\sqrt{\overline{Q}}}, \tag{59}$$

Eq. (52) takes the form

$$SN_\theta x^2 = \begin{cases} \sqrt{\frac{x^2}{9} - \frac{x^3}{48} + \frac{x^5}{5760}} & : \quad x < 4 \\ \\ \sqrt{\frac{16}{15} - \frac{16}{9}\frac{1}{x}} & : \quad x > 4 \end{cases} \tag{60}$$

x, hence $\delta\theta_e$ that satisfies Eq. (52) (at confidence level erf(ξ)), is obtained by numerical solution of Eq. (60). The graphs describing $x = (\delta\theta_e D/R)$ as a function of SN_θ, and $(\delta\rho_e/R)$ as a function of $SN_\rho \equiv N_l/(\xi\sqrt{\overline{Q}})$ are shown in Figs. 6 and 7 respectively.

In order to validate these results, the following simulation was performed. A 350×350 noise-free image with 20 roughly linear edge pixels was used. The normal parameters of the line best fitting the 20 pixels was $(\tilde{\theta}, \tilde{\rho}) = (-2.348, 7.36)$. 200 noisy images were produced by adding noise pixels at random

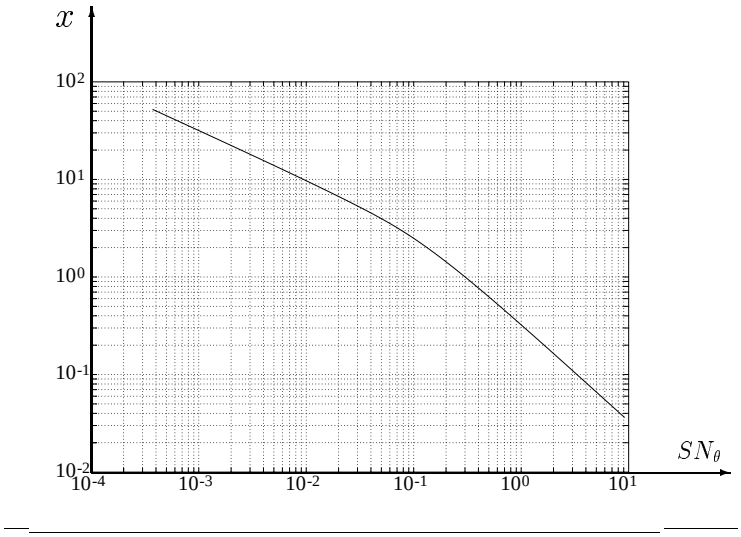

Figure 6: Continuous parameter plane analysis: $x = \delta\theta_e D/R$, the normalized upper bound on the θ peak location error, as a function of SN_θ.

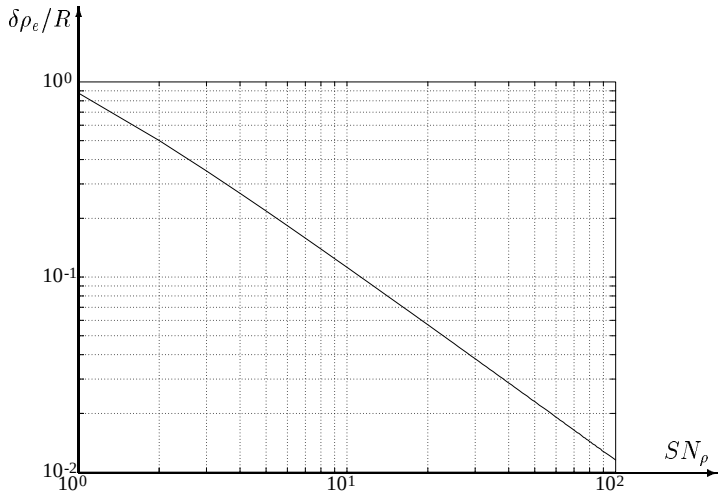

Figure 7: Continuous parameter plane analysis: $\delta\rho_e/R$, the normalized upper bound on the ρ peak location error, as a function of SN_ρ.

such that the expected number of noise pixels within a 24 pixel wide voting strip was $\overline{Q} = 5$. One of these images is shown in Fig. 8.

The Hough Transform of each of the 200 noisy images was computed with a voting range of $R = 12$, and the respective global maxima were determined at high resolution. The lines detected in the noisy images are shown in Fig. 9. Fig. 10 is the histogram of the detected θ values.

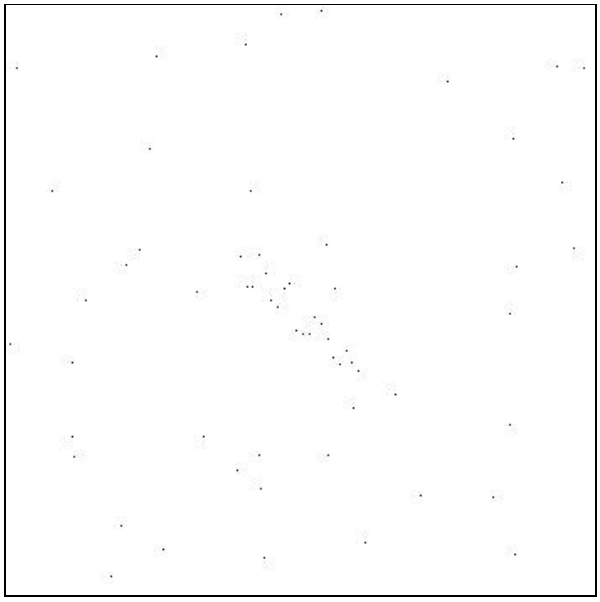

Figure 8: A 350×350 image edge that contains 20 roughly linear edge pixels and additional random noise pixels generated such that the expected number of noise pixels in a voting strip was $\overline{Q} = 5$.

The distribution of the detected θ values was compared to the theoretical result by calculating $\delta\theta_e$ for several values of $\mathrm{erf}(\xi)$, and comparing $\mathrm{erf}(\xi)$ to the fraction of images for which the deviation in the detected θ from $\tilde{\theta}$ was less than $\delta\theta_e$. The results given in Table 1 corroborate the theoretical analysis.

6 Strict Convergence of the Hough transform

The Hough transform is usually implemented by calculating $h(\theta, \rho)$ on a grid of sampling points in the parameter plane that is represented by an accumulator array. The maximum is then found by exhaustive

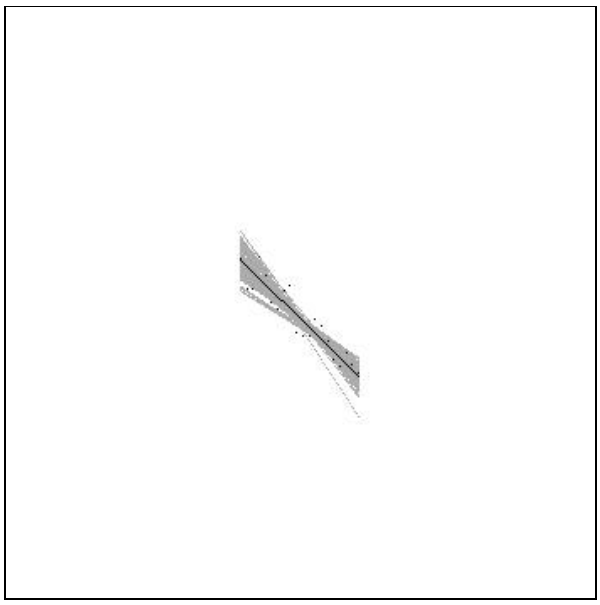

Figure 9: The 20 roughly collinear pixels, the best fitting line to these pixels (black) and the lines obtained from the 200 noisy images (gray).

ξ	$\delta\theta_e$	erf(ξ)	simulation
0.5	0.084	0.69	0.69
1.0	0.12	0.84	0.87
1.5	0.16	0.93	0.97
2.0	0.18	0.98	0.98

Table 1: Comparison of theoretic statistical bounds and experimental results on the deviation of the continuous parameter space peak in the θ direction due to noise. Each row corresponds to a confidence level erf ξ. The respective calculated upper bound on the deviation is in the second column. The fraction of images in which the bound is not exceeded is in the fourth column. As expected, it is upper bounded by the confidence level.

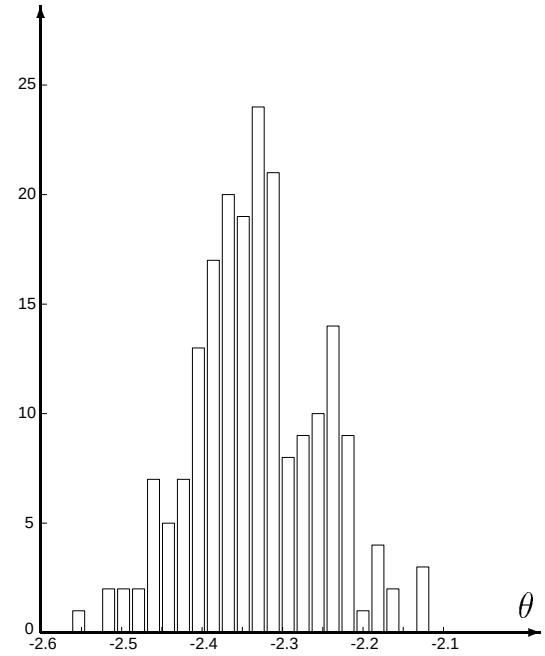

Figure 10: The histogram of detected θ values in 200 images of a roughly linear edge with added noise.

search. The hidden assumption is that the global maximum of $h(\theta, \rho)$ is near the sampling point that corresponds to the highest accumulator. This assumption is referred to as *strict convergence*.

There is an obvious tradeoff between resolution and computational load in selecting the density of the sampling points, i.e. the size of the accumulator array. In order to improve resolution at reduced computational costs, coarse to fine computation of the Hough transform was suggested, e.g. in [8]. In principle, the Hough transform is first calculated on a sparse grid that covers the whole domain P in the parameter plane. It is then recalculated on a denser grid over a subset of the parameter plane around the sampling point(s) that gained the highest value(s) in the previous stage. The process can be continued, using in each stage finer sampling grids that cover smaller subsets of the parameter plane.

Coarse to fine computation of the Hough transform is justified only if strict convergence holds at each stage in the process. Strict convergence will not take place if at any stage the sampling grid is too sparse. In this section the possibility of guaranteeing strict convergence is studied. It is important to realize

that the analysis is needed even for standard single-stage Hough algorithms to ensure that the highest accumulator corresponds to the global maximum.

Assume that the Hough Transform is implemented using a rectangular sampling grid with sampling intervals $\Delta\theta$ and $\Delta\rho$, such that the (p, q) accumulator in the accumulator array represents $h(p\Delta\theta, q\Delta\rho)$. By strict convergence of the Hough Transform we formally mean that the accumulator with the highest value should correspond to one of the four parameter plane sampling points that are closest to the global maximum of $h(\theta, \rho)$. Then the global maximum would be within a $2\Delta\theta \times 2\Delta\rho$ neighborhood around the highest accumulator. If strict convergence could be guaranteed, super-resolution would be easy to achieve by using a coarse to fine algorithm that assigns all computational resources to search for the global maximum in that neighborhood of the highest accumulator.

Consider first the Hough Transform of edge images without background noise. In order to guarantee strict convergence, we require that the sampling intervals should be small enough such that the lower bound on $h(\theta, \rho)$ at one of the four sampling points (accumulators) that surround the peak will be higher than the upper bound at all other sampling points. Comparison is made with sampling points inside and outside of $S_{\#}$.

6.1 Comparison with sampling points inside $S_{\#}$

Inside $S_{\#}$ $h_{lr}(\delta\theta, \delta\rho)$ is given by Eq. (19) and resembles a ridge. The position of the ridge with respect to the sampling grid might result in a potentially misleading situation in which the four sampling points that surround the peak are down the slopes, but another nearby sampling point is on the ridge, as shown in Fig. 11.

In the worst case let $(\delta\theta, \delta\rho) = (\delta\theta^*, \Delta\rho/2)$, for a certain $\delta\theta^* \in [0, \Delta\theta/2]$, be the coordinates of one of the four sampling points that surround the peak. In order to prevent the misleading situation, one can

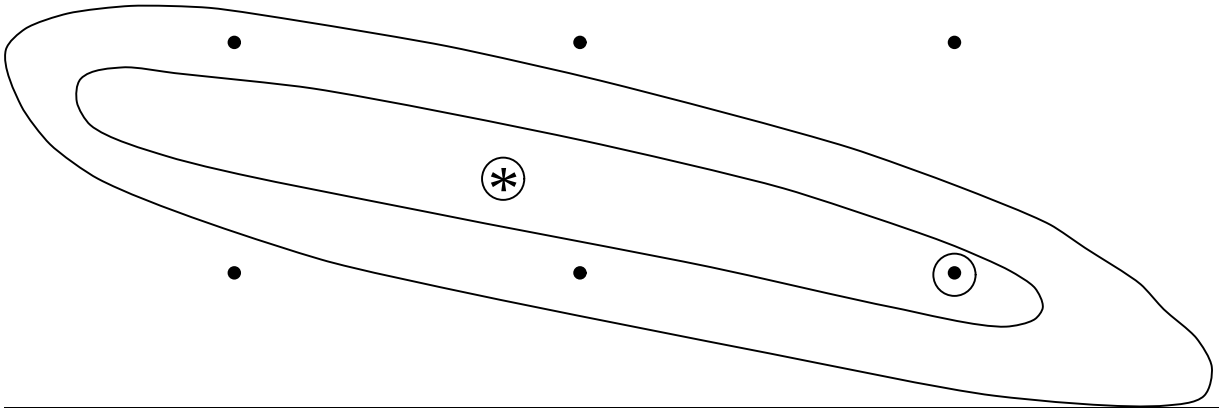

Figure 11: A misleading situation in which the four sampling points that surround the peak are down the slopes, but another nearby sampling point is on the ridge.

demand

$$\forall\, \delta\theta^* \in [0, \Delta\theta/2]: \quad h_{lr}\,(\delta\theta^*, \Delta\rho/2) > h_{lr}(\delta\theta^* + \Delta\theta, 0). \tag{61}$$

By substituting Eq. (19) and derivating with respect to $\delta\theta^*$, we find that for reasonable model parameters the minimal difference between the l.h.s. and the r.h.s. of inequality (61) is obtained for $\delta\theta^* = 0$. Substituting that value and rearranging the inequality yields the following requirement on the ratio between $\Delta\rho$ and $\Delta\theta$:

$$\Delta\rho < 2\sqrt{\sigma^2_{l\,\mathrm{min}} - \sigma^2_{d\,\mathrm{max}}} \cdot \sin\Delta\theta \tag{62}$$

6.2 Comparison with sampling points outside $S_{\#}$

According to inequality (46), the value of $h_{lr}(\delta\theta, \delta\rho)$ outside $S_{\#}$ is upper bounded by

$$h_{lr}(0,0)[1 - \gamma(R, \sigma_d^2)]$$

where $\gamma(R, \sigma_d^2)$ can be obtained from Fig. 5.

The relative coordinates of at least one of the four sampling points that surround the peak satisfy

$|\delta\theta| < \Delta\theta/2$ and $|\delta\rho| < \Delta\rho/2$, so the value of h_{lr} at that sampling point exceeds $h_{lr}\,(\Delta\theta/2, \Delta\rho/2)$. Using Eq. (19), for strict convergence we require that

$$N\left(1 - \frac{\sigma_d^2 \cos^2\frac{\Delta\theta}{2} + \sigma_l^2 \sin^2\frac{\Delta\theta}{2} + (\frac{\Delta\rho}{2})^2}{R^2}\right) > N\left(1 - \frac{\sigma_d^2}{R^2}\right)[1 - \gamma(R^2, \sigma_d^2)]. \tag{63}$$

Rearranging and taking the worst case model parameters we get:

$$\Delta\rho < 2\sqrt{\gamma \cdot (R^2 - \sigma_{d\,\max}^2) - (\sigma_{l\,\max}^2 - \sigma_{d\,\max}^2) \cdot \sin^2(\Delta\theta/2)} \tag{64}$$

6.3 Discussion on strict convergence

Without background noise, strict convergence is guaranteed if conditions (62) and (64) on the sampling intervals are satisfied. They respectively ensure that there would be no accumulators inside and outside $S_\#$ higher than an accumulator near the global maximum of $h(\theta, \rho)$. Note that in order for condition (64) to yield real values of $\Delta\rho$, the following condition must be also be satisfied:

$$\Delta\theta < 2\,\sin^{-1}\sqrt{\frac{\gamma \cdot (R^2 - \sigma_d^2)}{\sigma_{l\,\max}^2 - \sigma_d^2}}. \tag{65}$$

In order to reduce the computational complexity, feasible values of $\Delta\theta$ and $\Delta\rho$ should be selected such that $\Delta\theta \cdot \Delta\rho$ would be maximized. Actual calculation of $\Delta\theta$,$\Delta\rho$ for several examples gave quite small values. Hence, even without background noise, insisting on guaranteed strict convergence is computationally expensive. This is not surprising since the conditions for guaranteed convergence refer to worst cases and bounds and are very conservative with respect to typical images.

Suppose for example that the parameters of the linear edge model are $d < 3$, $\sigma_d^2 < 1.35$, $60 < l < 280$, $200 < \sigma_l^2 < 6200$. The best sampling intervals for guaranteed strict convergence are shown in Table 2.

R	$\Delta\theta$	$\Delta\rho$
$1.0d_{\max} = 3.0$	0.0116	0.33
$1.2d_{\max} = 3.6$	0.0231	0.65
$1.4d_{\max} = 4.2$	0.0316	0.89
$1.6d_{\max} = 4.8$	0.0393	1.11

Table 2: The best sampling intervals for guaranteed strict-sense convergence as a function of the voting range R for images of edges with $d < 3$, $\sigma_d^2 < 1.35$, $60 < l < 280$ and $200 < \sigma_l^2 < 6200$, in the absence of background noise.

The validity of these theoretic conditions has been corroborated by simulations and experiments with real but low noise images [23]. In order to handle edge images that contained several straight edges and thus do not conform to our image model, a simple criterion for discrimination between true edges and inter-edge arrangements was suggested and successfully used, see subsection 7.4.

It has not been possible to provide useful guarantees for strict convergence in the presence of background noise. Even with little noise the necessary sampling intervals become extremely small and with additional noise feasible sampling intervals cease to exist. It is thus necessary to develop a peak search strategy that would not be based on the strict convergence assumption. A suitable strategy, based on *wide-sense convergence*, is described in the next section.

7 Wide-sense convergence

Let $h_{\max}$ denote the maximal value in the accumulator array, i.e., the value of the maximal sampling point in the parameter space. The strict convergence assumption is that the maximum of $h(\theta, \rho)$ is in the neighborhood of that maximal sampling point. *Wide-sense convergence* of the Hough Transform is the weaker requirement that the global maximum of $h(\theta, \rho)$ is in the neighborhood of any of the sampling points that exceed a given threshold T, where $T < h_{\max}$.

The importance of a wide-sense convergence guarantee depends on the threshold T being sufficiently

high so that large parts in the parameter space P will be excluded further search. High resolution search can then be performed just near the sampling points that exceed the threshold. This is referred to as *dynamic* multi-resolution computation of the Hough Transform. The number of processing stages and the rate of refinement can be optimized for minimal computation.

Useful wide-sense convergence guarantees can be obtained even in the presence of background noise. However, since the background noise is random, a wide-sense convergence guarantee based on a threshold T is only at a certain statistical confidence level $erf(\xi)$. We proceed to obtain a threshold T of the form

$$T = \min(h_{t1}, h_{t2}), \tag{66}$$

where h_{t1} and h_{t2} respectively correspond to $h_{\max}$ being at a sampling point outside or inside of $S_{\#}$.

7.1 If $h_{\max}$ is outside $S_{\#}$

Suppose that the highest sampling point (accumulator) is not in the immediate neighborhood of the global maximum of $h(\theta, \rho)$. Assume first that it is outside $S_{\#}$. If the number of linear edge points had been known to be $\tilde{N}_l$, then at confidence level of at least $erf(\xi)$, the value of $h(\theta, \rho)$ in an accumulator outside $S_{\#}$ would have been upper bounded by

$$\tilde{N}_l \left[1 - \gamma(R, \sigma_d^2)\right] \frac{R^2 - \sigma_d^2}{R^2} + \frac{2}{3}\overline{Q} + \xi\sqrt{\frac{8}{15}\overline{Q}}$$

where $\overline{Q}$ is the expected number of background noise points in the voting strip. The contribution of the linear edge points follows from the definition of the gain margin (46) and Eq. (19) (with $\delta\theta = 0$ and $\delta\rho = 0$), and the noise contribution follows from equations (34) and (35). Here $h_{\max}$ is known, so by rearrangement we obtain that at confidence level of at least $erf(\xi)$ the number of linear edge points N_l

is at least

$$\frac{R^2}{\left[1-\gamma(R,\sigma_d^2)\right](R^2-\sigma_d^2)}\left[h_{\max}-\frac{2}{3}\overline{Q}-\xi\sqrt{\frac{8}{15}\overline{Q}}\right]$$

Given this lower bound on N_l, the value of the accumulator that corresponds to the sampling point that is nearest to the global maximum can also be lower bounded. Since the location $(\delta\theta, \delta\rho)$ of that sampling point satisfies $\delta\theta < \Delta\theta/2$ and $\delta\rho < \Delta\rho/2$, (using Eq. (19) as a lower bound on the contribution of the linear edge points) we obtain

$$h_{t1} = \frac{(R^2-\sigma_d^2)-(\sigma_{l\,\max}^2-\sigma_d^2)\sin^2(\frac{\Delta\theta}{2})-(\frac{\Delta\rho}{2})^2}{\left[1-\gamma(R,\sigma_d^2)\right](R^2-\sigma_d^2)}(h_{max}-\frac{2}{3}\bar{Q}-\xi\sqrt{\frac{8}{15}\bar{Q}}\ )+\frac{2}{3}\bar{Q}-\xi\sqrt{\frac{8}{15}\bar{Q}} \qquad (67)$$

7.2 If $h_{\max}$ is inside $S_\#$

Suppose again that the highest sampling point (accumulator) is not in the immediate neighborhood of the global maximum of $h(\theta,\rho)$, but assume now that it is inside $S_\#$. Using a similar approach, if the number of linear edge points had been known to be $\tilde{N}_l$, then at confidence level of at least $erf(\xi)$, the value of $h(\theta,\rho)$ at a sampling point inside $S_\#$ would have been upper bounded by

$$\tilde{N}_l\left[1-\frac{\sigma_d^2+(\sigma_{l\min}^2-\sigma_d^2)\sin^2\Delta\theta}{R^2}\right]+\frac{2}{3}\overline{Q}+\xi\sqrt{\frac{8}{15}\overline{Q}}.$$

where the upper bound on the contribution of the linear edge points is obtained by substituting $\delta\rho = 0$ and $\delta\theta = \Delta\theta$ in Eq. (19). (Since it is assumed that the sampling point is not in the immediate neighborhood of the global maximum, $\delta\theta > \Delta\theta$). Again, here $h_{\max}$ is known, so by rearrangement we obtain that at confidence level of at least $erf(\xi)$ the number of linear edge points N_l is at least

$$\frac{R^2}{(R^2-\sigma_d^2)(\sigma_{l\min}^2-\sigma_d^2)\sin^2\Delta\theta}\left[h_{\max}-\frac{2}{3}\overline{Q}-\xi\sqrt{\frac{8}{15}\overline{Q}}\right].$$

As above, given this lower bound on N_l, the value of the accumulator that corresponds to the sampling point that is nearest to the global maximum can be lower bounded. We obtain

$$\begin{aligned} h_{t2} &= \frac{(R^2-\sigma_d^2)-(\Delta\rho/2)^2}{(R^2-\sigma_d^2)-(\sigma_{l\,\min}^2-\sigma_d^2)\sin^2\Delta\theta}\left[h_{\max}-\frac{2}{3}\overline{Q}-\xi\sqrt{\frac{8}{15}\overline{Q}}\right] \\ &+ \frac{2}{3}\overline{Q}+\xi\sqrt{\frac{8}{15}\overline{Q}}-\xi\sqrt{\overline{\sigma_{\Delta h_N}^2(\Delta\theta)}} \end{aligned} \tag{68}$$

where $\overline{\sigma_{\Delta h_N}^2(\Delta\theta)}$ is given by Eq. (44) with $\delta\theta=\Delta\theta$.

7.3 Dynamic multi-resolution Hough transform

The wide sense convergence strategy provides some freedom in selecting the sampling intervals in a multi-resolution Hough transform. Note however that if the sampling intervals are very large at one stage, the computational savings at that stage will be out-balanced by the need for a large search area in the next stage. A reasonable rule of thumb is to select $\Delta\theta, \Delta\rho$ for the first, coarsest stage, such that $h_{lr}(\Delta\theta/2, \Delta\rho/2)$ will be attenuated with respect to $h_{lr}(0,0)$ by a factor of $f \approx 2$, and reduce the attenuation in the following stages. If, in the final stage, the sampling intervals satisfy Eq. (62) and (64), then the accumulator with the highest value in that stage can be associated with the global maximum. Note that at the final stage R is smallest and the effect of noise is minimal. The attenuation f is given by

$$f=\frac{h_{lr}(\Delta\theta/2,\Delta\rho/2)}{h_{lr}(0,0)}=\frac{(R^2-\sigma_d^2)-(\sigma_l^2-\sigma_d^2)\sin^2(\Delta\theta/2)-(\Delta\rho/2)^2}{R^2-\sigma_d^2}$$

Thus, for a given f, the sampling intervals should satisfy the constraint:

$$\Delta\rho=\sqrt{(1-f)(R^2-\sigma_d^2)-(\sigma_l^2-\sigma_d^2)\sin^2(\Delta\theta/2)} \tag{69}$$

Among all pairs $(\Delta\theta, \Delta\rho)$ that satisfy (69) it is most economical to select the pair that maximizes the area $\Delta\theta\cdot\Delta\rho$ and minimizes the total number of accumulators.

R		f	$\Delta\theta$	$\Delta\rho$
$(1.4d)$	4.2	0.5	0.052	4
$(1.2d)$	3.6	0.85	0.024	1.85
$(1.0d)$	3.0	0.98	0.007	0.55

Table 3: Suggested parameters for a three-stage multi-resolution Hough algorithm based on the wide search convergence strategy for an image with $d < 3$, $\sigma_d^2 < 1.35$, $60 < l < 280$ and $200 < \sigma_l^2 < 6200$.

Consider for example the linear edge parameters given in subsection 6.3. Reasonable parameters for a three-stage multi-resolution Hough algorithm based on the wide search convergence strategy are shown in Table 3.

Comparing the sampling intervals to those needed for strict convergence (Table 2), we notice the potential for considerable reduction in the computational load. The actual savings will depend on the size of the search areas that follow from the specified thresholds. These depend on the noise level and the required confidence level.

7.4 Complex images

The convergence guarantees obtained in this paper are based on the image model described in section 2. The model includes a roughly collinear set of points and uniformly distributed background points. In applying the Hough transform to images that conform well to the model one needs to detect just the single collinear set of points and the convergence guarantees can be directly applied. Complex edge images may however contain several straight line segments and other curves. This leads to two problems. First, one may want to detect more than a single collinear subset of points. Second, one might encounter accidental collinear arrangements of edge points that belong to several *objects* in the image. These are different than the random arrangements of background noise points that are accounted for in the analysis.

In our implementation of the dynamic multi-resolution Hough transform, multiple straight edges are

detected one by one. In principle, whenever a peak in the parameter space is identified and located, the associated edge pixels are discarded and the Hough transform repeated. This well known method is seemingly very wasteful, but in the spirit of Gerig's approach [5], the process can be implemented very efficiently by subtractive re-voting of just of those edge pixels.

In standard implementations of the Hough transform, the discrimination between genuine straight edges and accidental inter-object arrangements of edge points is a post-processing operation carried out in the image plane. In the dynamic multi-resolution Hough transform it is usually possible to identify the accidental arrangements online.

Let R_c and R_f respectively denote the voting ranges in the coarse stage and in the fine stage of the dynamic multi-resolution Hough Transform. In the absence of noise, the ratio between the values of a given peak in the two stages is (see Eq. (19):

$$\frac{1-\sigma_d^2/R_f^2}{1-\sigma_d^2/R_c^2} = \frac{R_c^2}{R_f^2}\,\frac{R_f^2-\sigma_d^2}{R_c^2-\sigma_d^2} \tag{70}$$

With noise, the ratio becomes

$$\frac{1-\frac{\sigma_d^2}{R_f^2}+\frac{4}{3}\frac{R_f\hat{Q}}{N_l}}{1-\frac{\sigma_d^2}{R_c^2}+\frac{4}{3}\frac{R_c\hat{Q}}{N_l}} \tag{71}$$

where $\hat{Q}$ is the expected number of noisy pixels a one pixel wide strip and N_l is the number of edge points in the linear edge. These ratios are usually quite sensitive to σ_d. Accidental inter-object collinear arrangements of edge points can be expected to be much more scattered (in terms of σ_d) than genuine straight edges. It is thus possible to distinguish between wanted and unwanted maxima by comparing the ratio between their values in the fine and coarse stages. Formal thresholds can be obtained by substituting image model parameters in Eqs. (70) or (71).

Stage	R	$\overline{Q}$	γ	$\sqrt{\overline{\sigma^2_{\delta h_n}(\Delta\theta)}}$	$h_{t,1}$	$h_{t,2}$
1	4.2	46	0.11	6.1	$0.56\,h_{\max} + 5.7$	$0.78\,h_{\max} + 1.7$
2	3.6	40	0.08	4.4	$0.93\,h_{\max} - 7.1$	$0.94\,h_{\max} - 2.5$
3	3.0	33	0.03	2.6	$1.01\,h_{\max} - 8.7$	$0.99\,h_{\max} - 2.7$

Table 4: The necessary thresholds in a three stage dynamic multi-resolution Hough transform based on the wide sense convergence strategy with the sampling intervals suggested in Table 3. Other parameters are also shown.

8 Example

We have demonstrated [23] the design and operation of a three-stage dynamic multi-resolution Hough algorithm on real images containing several linear edges as well as other features and corrupted by noise. Consider for example the 350×350 image edge shown in Fig. 12. It consists of six straight lines as well as other shapes and noise. The straight lines satisfy the linear edge model discussed in subsection 6.3. The noise is real, due to poor illumination. It is quite uniformly distributed, and its level is such that the expected number of noise pixels within a one pixel wide rectangle along the image diagonal is $\hat{Q} = 5.5$.

Using a three stage wide-sense convergence approach, the sampling intervals suggested in Table 3 were used in order to minimize computation while ensuring convergence. The necessary thresholds were obtained using Eqs. (67) and (68) with $\xi = 3$ (confidence level erf(3)) and are summarized in Table 4. Only the neighborhoods of accumulators that exceeded the combined threshold $h_t = \min\{h_{t1}, h_{t2}\}$ in a given stage had to be searched in the following stage. The computational load of the second and third stages was negligible with respect to the first, coarse stage.

The linear edges were detected one by one as described in subsection 7.4. Eleven collinear subsets of points were detected, but using the suggested online discrimination method the five accidental arrangements were identified and all the six genuine linear edges were detected. See Fig. 13.

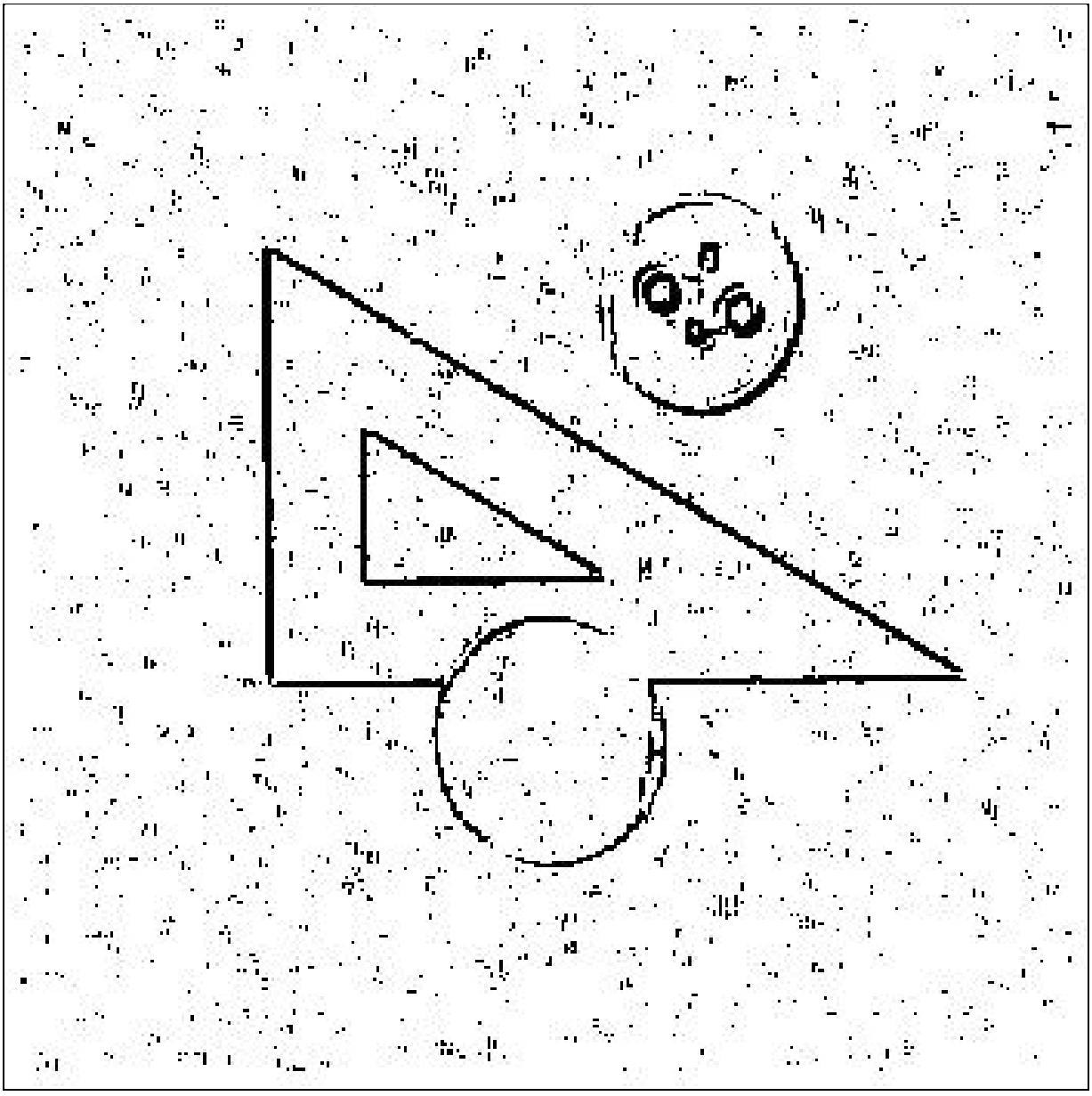

Figure 12: A 350×350 image edge that contains six straight lines that satisfy the linear edge model specified in subsection 6.3, as well as other shapes and real noise due to poor illumination. The expected number of noise pixels within a one pixel wide rectangle along the image diagonal is $\hat{Q} = 5.5$.

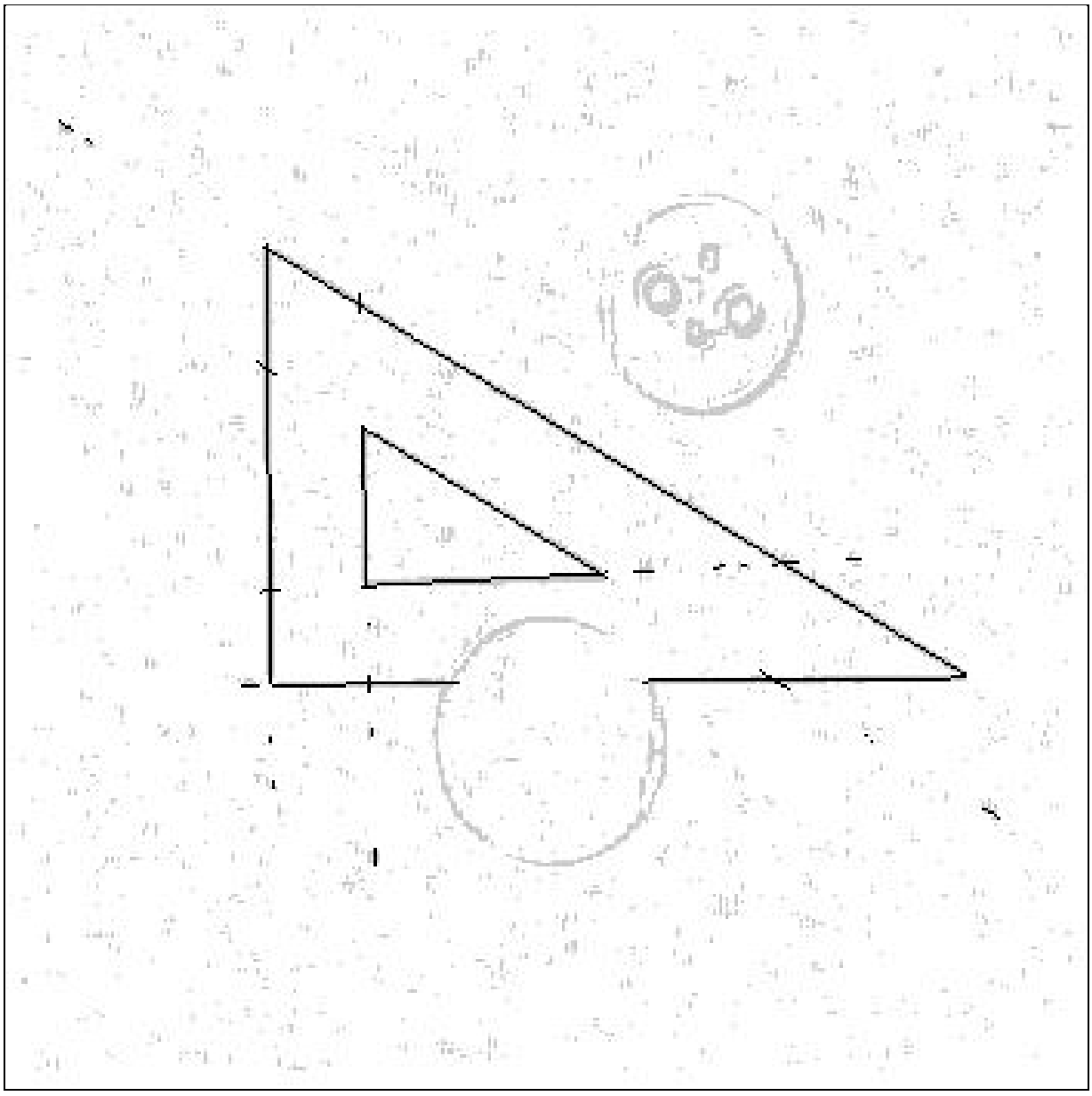

Figure 13: The six linear edges that were detected in the image shown in Fig. 12 are highlighted.

9 Discussion

The Hough Transform for pattern recognition in edge images is an attempt to solve a continuous global optimization problem by grid sampling and association of the continuous global maximum (or maxima) with the globally highest grid sample(s). To our knowledge, this is the first paper in which formal conditions for this association to hold are provided. Even though the convergence guarantees rely on edge and noise models, experiments indicate that our results provide valuable guidelines for the selection of voting kernel widths, parameter space sampling intervals and focusing rates in multi-resolution as well as standard smooth kernel Hough algorithms. It is interesting to use the results of this paper to analyze and to provide theoretical support for hierarchical bottom-up processing as suggested in Princen *et al* [18] and Leavers [15].

In global optimization terminology the Hough Transform is a multidimensional covering method. The suggested approach can be categorized as a covering method with guaranteed convergence. Note that at least in 1988 there seemed to be no published results on the application of covering algorithms to practical problems (Törn and A. Žilinskas [26], page 85). The use of genetic algorithms for extracting geometric primitives has been suggested by Roth and Levine [22].

Standard M-estimation algorithms that rely on the availability of a reliable initial estimate of the line parameters as a starting point for local search can be regarded as *single-start* algorithms [26]. The strict-sense convergence approach discussed in section 6 is in fact rectangular sampling followed by single-start higher resolution grid search. In multi-resolution Hough based on wide-sense convergence (section 7), focusing follows the *multi-start* [26] paradigm. The use of hexagonal sampling grids in the Hough Transform has been suggested [10]; it is interesting to consider the possible advantages of adaptive sampling grids [26] in the Hough Transform. The inverse voting algorithm suggested by Chang and Hashimoto [2] is in fact a grid-less multi-start algorithm that is initiated at random locations in the parameter space and uses gradient search for local refinement.

Acknowledgments

This research was supported in part by the Israeli Ministry of Science and the Arts. It was carried out at the Ollendorff center of the Department of Electrical Engineering, Technion.

References

[1] D.H. Ballard, "Generalizing the Hough Transform to Detect Arbitrary Shapes", Pattern Recognit., Vol. 13, pp. 111-122, 1981.

[2] D. Chang and S. Hashimoto, "An Inverse Voting Algorithm for Hough Transform", Proceedings, IEEE 1st International Conference on Image Processing, Vol. I, pp. 223-227, Austin, Texas, November 1994.

[3] R.O. Duda and P.E. Hart,"Use of the Hough Transform to Detect Lines and Curves in Pictures", Commun. ACM, Vol. 15, pp. 11-15, 1972.

[4] W. Förstner, Tutorial Notes, 2nd International Workshop on Robust Computer Vision, Bonn, Germany, 1992.

[5] G. Gerig, "Linking Image Space and Accumulator Space: A New Approach for Object Recognition", Proceedings, IEEE 1st International Conference on Computer Vision, pp. 112-117, London, 1987.

[6] C. Goodall, "M-estimators of Location: An Outline of the Theory", in D.C. Hoaglin, F. Mosteller and J.W. Tukey (eds.), *Understanding Robust and Exploratory Data Analysis*, Wiley, 1983.

[7] P.V.C. Hough, "Method and Means for RecognizingComplex Patterns", U.S. Patent No. 3069654, 1962.

[8] J. Illingworth and J. Kittler, "The Adaptive Hough Transform", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 9, pp. 690-697, 1987.

[9] J. Illingworth and J. Kittler, "A Survey of the Hough Transform", Comput. Vision Graphics Image Process., Vol. 44, pp. 87-116, 1988.

[10] N. Kiryati and A.M. Bruckstein, "Anti-aliasing the Hough Transform", CVGIP: Graphical Models Image Process., Vol. 53, pp. 213-222, 1991.

[11] N. Kiryati and A.M. Bruckstein, "On Navigating between Friends and Foes", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 13, pp. 602-606, 1991.

[12] N. Kiryati and A.M. Bruckstein, "What's in a Set of Points?", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 14, pp. 496-500, 1992.

[13] N. Kiryati, Y. Eldar and A.M. Bruckstein, "A Probabilistic Hough Transform", Pattern Recognit., Vol. 24, pp. 303-316, 1991.

[14] V.F. Leavers, "Which Hough Transform?", CVGIP: Image Understanding, Vol. 58, pp. 250-264, 1993.

[15] V.F. Leavers, "Linking Low Level Perceptive Vision Processing to High Level Cognitive Processes Using the Dynamic Generalized Hough Transform", Proceedings, 7th Israeli Symposium on Artificial Intelligence and Computer Vision, pp. 315-324, Ramat Gan, Israel, December 1990.

[16] G. Li, "Robust Regression", in D.C. Hoaglin, F. Mosteller and J.W. Tukey (eds.), *Exploring Data Tables, Trends and Shapes*, Wiley, 1985.

[17] W. Niblack and D. Petkovic, "On Improving the Accuracy of the Hough Transform", Machine Vision and Applications, Vol. 3, pp. 87-106, 1990.

[18] J. Princen, J. Illingworth and J. Kittler, "A Hierarchical Approach to Line Extraction", Comput. Vision Graphics Image Process., Vol. 52, pp. 57-77, 1990.

[19] J. Princen, J. Illingworth and J. Kittler, "A Formal Definition of the Hough Transform: Properties and Relationships", Journal of Mathematical Imaging and Vision, Vol. 1, pp. 153-168, 1992.

[20] J. Princen, J. Illingworth and J. Kittler, "Hypothesis Testing: A Framework for Analyzing and Optimizing Hough Transform Performance", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 16, pp. 329-341, 1994.

[21] J. Princen, H.K. Yuen, J. Illingworth and J. Kittler, "A Comparison of Hough Transform Methods", Proceedings, IEE 3rd International Conference on Image Processing and its Applications, pp. 73-77, University of Warwick, UK, July 1989.

[22] G. Roth and M.D. Levine, "Extracting Geometric Primitives", CVGIP: Image Understanding, Vol. 58, pp. 1-22, 1993.

[23] M. Soffer, "Conditions for Convergence of the Hough Transform", M.Sc. Thesis, Dept. of Electrical Engineering, Technion - Israel Institute of Technology, August 1994 (Hebrew, English abstract).

[24] D.L. Snyder, "Random Point Processes", Wiley-Interscience, New York, 1975.

[25] P.R. Thrift and S.M. Dunn, "Approximating Point-Set Images by Line Segments Using a Variation of the Hough Transform", Comput. Vision Graphics Image Process., Vol. 21, pp. 383-394, 1983.

[26] A. Törn and A. Žilinskas, *Global Optimization*, Lecture Notes in Computer Science #350, Springer, 1989.

[27] T.M. Van Veen and F.C.A. Groen, "Discretization Errors in the Hough Transform", Pattern recognit., Vol. 14, pp. 137-145, 1981.