Maybank, Stephen J. (2004) Detection of image structures using the Fisher information and the Rao metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26** (12) 1579-1589.

Detection of Image Structures Using the Fisher Information and the Rao Metric

Stephen J. Maybank, *Member*, *IEEE*

Abstract—In many detection problems, the structures to be detected are parameterized by the points of a parameter space. If the conditional probability density function for the measurements is known, then detection can be achieved by sampling the parameter space at a finite number of points and checking each point to see if the corresponding structure is supported by the data. The number of samples and the distances between neighboring samples are calculated using the Rao metric on the parameter space. The Rao metric is obtained from the Fisher information which is, in turn, obtained from the conditional probability density function. An upper bound is obtained for the probability of a false detection. The calculations are simplified in the low noise case by making an asymptotic approximation to the Fisher information. An application to line detection is described. Expressions are obtained for the asymptotic approximation to the Fisher information, the volume of the parameter space, and the number of samples. The time complexity for line detection is estimated. An experimental comparison is made with a Hough transform-based method for detecting lines.

Index Terms—Analysis of algorithms, clustering, edge and feature detection, multivariate statistics, robust regression, sampling, search process.

1 Introduction

One of the main tasks in computer vision is the detection of image structures such as lines and circles, as well as more abstract structures such as epipolar transforms, collineations, and fundamental matrices [13]. This paper describes a class of probabilistic models which are applicable to a wide range of image structures and shows how a probabilistic model can supply the information needed to design an algorithm for detecting the relevant structure. Each probabilistic model is a conditional density $p(x|\theta)$, where x is a measurement in a measurement space D and θ corresponds to a structure and takes values in a parameter space T. The example of line detection is discussed in detail in order to show how the general theory is applied to a particular case.

The advantages of this approach are: 1) The parameters required by the algorithm can be calculated from $p(x|\theta)$ and a user defined threshold e_f on the probability of a false detection. 2) The algorithm is simple: The space T is sampled at a finite number of points and the structures corresponding to one or more of these points are detected if they have a sufficient number of inliers. 3) In the case of line detection, the time complexity is only $O(N(\gamma t)^{-1/2}\ln(N(\gamma t)^{-1/2}))$, where N is the number of measurements, $\gamma = O(1)$, and $2t$ is the variance of the measurement noise.

The class of probabilistic models is an extension of the class of models defined by Werman and Keren [24]. In the absence of noise, the measurements compatible with the structure corresponding to the parameter value θ in T form a subset $M(\theta)$ of D. In the special case of lines, D coincides with the image, T is a two-dimensional manifold, and $M(\theta)$ is a line in D. The probability density function $p(x|\theta)$ for a measurement x given $M(\theta)$ is obtained from a solution $(s,x)\mapsto p_s(x|\theta)$ to the heat equation on D, where s is the time parameter in the heat equation. At time 0, $p_0(x|\theta)$ is zero outside $M(\theta)$. As s increases, the density $p_s(x|\theta)$ takes larger values away from $M(\theta)$. The heat flow is stopped at a time t. The density $p(x|\theta)$ is given by $p(x|\theta) = p_t(x|\theta)$. The density $p(x|\theta)$ is a familiar one, in spite of its elaborate definition: If t is small, then $\ln p(x|\theta)$ is proportional to the squared distance from x to $M(\theta)$. Further information about $p(x|\theta)$ is given in Section 3.1 and in Appendices A and B.

The density $p(x|\theta)$ contains information which has not so far been used in applications to computer vision. The source of the information is a Riemannian metric [5], [9] defined on T by $p(x|\theta)$ and known in statistics as the Rao metric [15], [22]. The Rao metric is the distance metric for comparing parameter vectors wished for in the Introduction to [11]. It is defined at each point θ of T by a matrix $J(\theta)$ which is the Fisher information [1], [4], [7], [19] of $p(x|\theta)$. Under the Rao metric, the space T has a volume $V(T,J)$. The volume $V(T,J)$ is a measure of the difficulty of searching D for occurrences of the structures $M(\theta)$. If $V(T,J)$ is small, then, in a sense to be made precise in Section 4.2 below, D contains only a few distinct structures and it is possible to search D quickly for those structures which are supported by the measurements.

The Rao metric leads to an upper bound on the probability of a false detection. The upper bound exists because it is, in some sense, possible to make a finite list of "all the false detections that might occur." If the probability of each individual false detection is small, then the probability of obtaining any false detection on the list is also small. False detections are often discussed in the literature, see [6], [11], [23] for example, but, until now, the discussion has not included any quantitative description of all the false detections that might occur. The results on false detection obtained in this paper support the claim in [23] that "...fits with an arbitrarily low inlier percentage...may be found, as long as the bad data are random and the good data are close enough to the correct fit." Numerical evidence presented in Section 6.2 suggests but does not

• *The author is with the School of Computer Science and Information Systems, Birkbeck College, University of London, Malet Street, London, WC1E 7HX, UK. E-mail: sjmaybank@dcs.bbk.ac.uk.*

Manuscript received 11 Apr. 2003; revised 2 Apr. 2004; accepted 3 May 2004. Recommended for acceptance by A. Rangarajan.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0030-0403.

prove that, at a fixed noise level and a fixed probability of false detection, the ratio r/N of the least number r of inliers sufficient for detection to the total number of measurements N tends to 0 as N tends to infinity.

An advantage of using the Rao metric is that the results, including the volume $V(T, J)$, the number of distinct structures, and the upper bound for the probability of a false detection, are independent of the choice of parameterization of T. If a tractable parameterization of T exists, then it can be used with no loss of information.

Unfortunately, the densities $p(x|\theta)$ which arise in practice only rarely yield mathematically tractable or closed form expressions for $J(\theta)$. This difficulty can sometimes be overcome when the measurement noise is small by replacing $J(\theta)$ with an asymptotic approximation $K(\theta)$ which is more likely to have a closed form expression. A formula (34) for $K(\theta)$ is given below in Appendix B. The strategy of approximating $J(\theta)$ by $K(\theta)$ is successful in the case of lines: $K(\theta)$ takes a particularly simple form and it is straightforward to calculate $V(T, K)$ and to implement a line detection algorithm based on $K(\theta)$.

Related work on line detection is discussed in Section 2. Background material from statistics is covered in Section 3. In Section 4, it is shown in detail how the theory outlined above leads to structure detection algorithms based on sampling the parameter space and an upper bound is obtained for the probability of false detection. In Section 5, the theory developed in Sections 3 and 4 is applied to line detection. False detections of lines are discussed in Section 6 and experimental results are reported in Section 7. Some suggestions for future research are made in Section 8.

2 Line Detection

It is assumed that the data for line detection consist of a set of measurements of image points. The task is to find those subsets of the measurements which support the presence of a line. It is not assumed that a line is present and, even if there is a line present, it is not assumed to be unique. The measurements supporting the presence of a line l are known as the inliers for l. The remaining measurements are outliers for l, but some of them may be inliers for other lines in the image. In this section, four methods for detecting lines are described, namely RANSAC [6], MINPRAN [23], the Hough transform [8], [10], [14], [17], and the new method based on the Rao metric on T.

The RANSAC method of line detection is described first. Suppose that there are N measurements $x(1), \ldots, x(N)$. A pair of distinct measurements $x(i)$, $x(j)$ is chosen and the remaining measurements are checked to see how many are inliers for the line $\langle x(i), x(j)\rangle$. If there are a sufficient number of inliers, then $\langle x(i), x(j)\rangle$ is detected. Ideally, each of the $N(N-1)/2$ pairs of measurements should be checked to see if enough of the remaining measurements are inliers to the line defined by the pair. However, this is inefficient if N is large. Instead, RANSAC takes a series of random samples from the set of $N(N-1)/2$ pairs of measurements. The number of random samples depends on a prior estimate of the number of measurements which are inliers to a given line. RANSAC detects lines in the presence of large numbers of outliers. It has the disadvantage that it is not possible to calculate the probability of a false detection.

MINPRAN builds on RANSAC by making a careful investigation of the criteria for deciding 1) if a measurement is an inlier to a particular line and 2) if there are enough inliers to justify detecting the line. The resulting improvements make it possible to detect lines and other structures reliably even if there are a large number of outliers and even if the variance of the measurement noise is unknown. The problem remains of calculating the probability of a false detection over all the lines that might occur.

In the Hough transform method, the set of all lines in the image is parameterized by a two-dimensional parameter space T. Each point θ in T corresponds to a line in the image. An example of a parameter space is given: Let l be an image line and let $\rho(\cos(\alpha), \sin(\alpha))$ be the vector of minimum Euclidean length from the origin to l. Then, $\theta = (\rho, \alpha)$ and T is the region of the plane defined by $0 \leq \rho < b$, $0 \leq \alpha < 2\pi$, where b is an upper bound depending on the size of the image. The space T is divided into small regions called buckets [8] or accumulator cells [10]. Each bucket B is assigned an integer $a(B)$ which is initialized to zero. The measurements $x(i)$ are examined, in turn, for $1 \leq i \leq N$. If $x(i)$ is on a line corresponding to a point θ in B, then $a(B)$ is increased by 1. The final value of $a(B)$ is equal to the number of measurements which are on lines corresponding to points in B. If $a(B)$ is large, then a line is detected in the image with parameter vector θ in B. The disadvantage of the Hough transform is that there is no probabilistic model for deducing the values of the key parameters. These parameters include the size and number of the buckets and the threshold on $a(B)$ for detecting a line.

In the new method, T is sampled at a finite set of points $G \subset T$. The set G does not depend on the values of the measurements or on the number N of the measurements. The set G is searched for the set $L \subseteq G$ of points corresponding to lines with r or more inliers, where r is a threshold which depends on the noise level t, the number of measurements N, and a user specified probability e_f of false detection. The points in L correspond to the lines detected in the image.

3 Probabilistic Model for Image Structures

The aim in this section is to describe a general probabilistic model suitable for a wide range of detection problems, including line detection. The application to line detection is described in detail in Sections 5, 6, and 7.

3.1 The Model for $p(x|\theta)$

The definition of the probability density function $p(x|\theta)$ is an extension of a definition given in [24] for $D = \mathbb{R}^d$. The space D is given a Riemannian metric g with the associated canonical measure $d\mu$ [5], [9]. The measure $d\mu$ is obtained by using g to calculate the volumes of sets in D. For example, a d-dimensional cuboid at x with sides $\Delta x_1 \times \ldots \times \Delta x_d$ has a volume under $d\mu$ approximately equal to $|\det(g(x))|^{1/2}$ $\Delta x_1 \ldots \Delta x_d$. The probability that a measurement x is contained in a subset A of D, given the presence of a structure $M(\theta) \subseteq D$, is $\mathrm{Prob}(x \in A|\theta) = \int_A p(x|\theta)\, d\mu$.

It is assumed that each measurement x is obtained by adding noise to an underlying noise free measurement $\tilde{x}$ such that $p(x|\tilde{x})$ is the result of a heat flow or diffusion on D. The heat flow begins at time 0 as a delta function concentration of heat at $\tilde{x}$. The heat flow lasts for a time t, giving a probability density function $p(x|\tilde{x}) = p_t(x|\tilde{x})$. If $D = \mathbb{R}^d$ and g is the Euclidean metric, then $d\mu$ is the Lebesgue measure in $\mathbb{R}^d$ and $p_t(x|\tilde{x})$ is the Gaussian density with expected value $\tilde{x}$ and

covariance $2tI$, where I is the $d \times d$ identity matrix. In the case of line detection, D is the unit disc in $\mathbb{R}^2$, g is the Euclidean metric, and $d\mu$ is the Lebesgue measure on D.

Each subset $M(\theta)$ of D is given a probability measure dh which specifies the distribution of $\tilde{x}$ on $M(\theta)$. If the measurements are more likely to arise from certain parts of $M(\theta)$, then dh should be larger in those parts. Conversely, dh should be smaller in those parts of $M(\theta)$ less likely to give rise to measurements. If there is no information about the distribution of $\tilde{x}$ on $M(\theta)$, then the simplest default choice is to make dh equal to a scaled version of the measure induced on $M(\theta)$ as a submanifold of D. The scaling is chosen such that the total volume of $M(\theta)$ under dh is 1. The density $p(x|\theta) = p_t(x|\theta)$ is obtained by integrating the different contributions $p_t(x|\tilde{x})$ as $\tilde{x}$ ranges over $M(\theta)$ or, equivalently, by solving the heat equation on D with the condition that, at time 0, the distribution of the heat is given by dh. As the time increases away from 0, heat flows from $M(\theta)$ into the rest of D. If t is small, then the density $p_t(x|\theta)$ is concentrated in a neighborhood of $M(\theta)$.

3.2 Fisher Information and the Rao Metric

General arguments from probability theory show that the Fisher information for the family of densities $x \mapsto p(x|\theta)$, $\theta \in T$ gives rise to a statistically meaningful definition of volume on T. Intuitively, a subset B of T has a small volume if the densities $p(x|\theta)$, $\theta \in B$ are similar to each other. Let $n(T)$ be the dimension of T. The Fisher information is the symmetric $n(T) \times n(T)$ matrix $J(\theta)$ defined for $\theta \in T$ by

$$J_{ij}(\theta) = -\int_D \Big(\partial^2_{\theta_i,\theta_j} \ln(p(x|\theta))\Big) p(x|\theta)\, d\mu(x), \quad 1 \le i, j \le n(T),$$

where θ_i, θ_j are components of the vector θ and $\partial^2_{\theta_i,\theta_j}$ is the differential operator defined such that $\partial^2_{\theta_i,\theta_j} \ln(p(x|\theta)) = \partial^2 \ln(p(x|\theta))/\partial\theta_i\partial\theta_j$. The Fisher information for r measurements sampled independently from $p(x|\theta)$ is $rJ(\theta)$. However, $J(\theta)$ is used rather than a multiple such as $rJ(\theta)$ because it is not known a priori which sets of measurements are inliers to the same line.

The Fisher information defines a Riemannian metric on T, known as the Rao metric. The square of the length element ds for the Rao metric is $ds^2 = d\theta^\top J(\theta) d\theta$. The Rao metric has a statistical meaning [1], [2]. Let $x \mapsto p(x|\theta)$, $x \mapsto p(x|\theta')$ be two probability density functions and let x be a measurement sampled either from $p(x|\theta)$ with probability $1/2$ or from $p(x|\theta')$ with probability $1/2$. Suppose that x and θ, θ' are given but the information about which of $p(x|\theta)$, $p(x|\theta')$ provided the sample x is hidden. If θ, θ' are close together under the Rao metric, then any method for choosing the density which provided x has a high probability of error.

Let $\tau(\theta)\, d\theta$ be the canonical measure on T associated with the Rao metric. The measure $\tau(\theta)\, d\theta$ is defined by $\tau(\theta)\, d\theta = |\det(J(\theta))|^{1/2}\, d\theta$. The volume $V(B, J)$ of any subset B of T under the canonical measure is defined by $V(B, J) = \int_B \tau(\theta)\, d\theta$. The volume $V(B, J)$ is independent of the choice of parameterization of T. If $V(B, J)$ is large and if a measurement x is given, then there is a high probability that B contains a point θ for which $p(x|\theta)$ is large. The reason is that the $p(x|\theta)$ vary widely as θ ranges over B. In some sense, there are "many" $p(x|\theta)$ for $\theta \in B$ and, therefore, an increased probability that $p(x|\theta)$ is large for some $\theta \in B$. Conversely, if $V(B, J)$ is small, then there is a low probability of finding a point θ in B for which $p(x|\theta)$ is large.

Amari [1] shows that a wide range of metrics for comparing probability density functions reduce to simple functions of the Rao metric when the density functions are near to each other. Examples of such metrics include Kullback-Leibler, Bhattarcharrya, Matusita-Hellinger, Chernoff, and the Jensen-Shannon divergence. For example, the Kullback-Leibler distance $D(\theta||\theta')$ between $p(x|\theta)$ and $p(x|\theta')$ is approximated by

$$D(\theta||\theta') = \frac{1}{2}(\theta - \theta')^\top J(\theta)(\theta - \theta') + O(\|\theta - \theta'\|^3). \quad (1)$$

Further information is given in [19]. The connection between $D(\theta||\theta')$ and the Rao metric fails if $D(\theta||\theta')$ is large. The right-hand side of (1) is an approximation to half the square of the geodesic distance [5], [9] between θ and θ'. The geodesic distance is symmetric in θ, θ', but $D(\theta||\theta')$ is not symmetric, thus (1) cannot hold, in general, if $D(\theta||\theta')$ is large. To the author's knowledge, there is no known statistical or information theoretic interpretation of the geodesic distance between widely separated points θ, θ' of T.

4 Models

In Section 4.1, it is shown that T can be divided into small subsets $B(\theta)$, in which the central point θ is a single representative for all the points θ' in $B(\theta)$. The different θ together form a discrete approximation to T which is the basis of a simple algorithm for detecting the structures parameterized by T. In this approach to structure detection, it is possible to calculate upper bounds for the probability of false detection, as explained in Sections 4.3 and 4.4.

4.1 Models Represented by θ

Let γ be a constant of order 1. The set $B(\theta)$ of models represented by θ is the ellipsoid defined by

$$B(\theta) = \left\{\theta' \in T, \frac{1}{2}(\theta' - \theta)^\top J(\theta)(\theta' - \theta) \le \gamma\right\}. \quad (2)$$

The factor $1/2$ is included in (2) to ensure that γ is approximately equal to the Kullback-Leibler distance from θ to any point on the boundary of $B(\theta)$. If θ' is in $B(\theta)$, then the submanifolds $M(\theta)$, $M(\theta')$ of D are so close together that a measurement x arising from $\tilde{x}$ in $M(\theta)$ can, with a high probability, also arise from a nearby point $\tilde{y}$ in $M(\theta')$. The point θ is regarded as a representative model for all the points θ' in $B(\theta)$.

The value of γ should not be too small, otherwise there would exist points θ' well outside $B(\theta)$ but such that $M(\theta)$, $M(\theta')$ are indistinguishable given a single measurement. On the other hand, an upper bound of the form $\gamma \approx 1$ is needed to ensure that, if $\theta', \theta'' \in B(\theta)$, then any inlier x for $M(\theta')$ is also an inlier for $M(\theta'')$. It may be possible to deduce an exact value of γ using a constraint on the probability of missed detection. An argument is given in Section 5.4 to show that, in the case of line detection, it is reasonable to choose $\gamma = 1/2$.

4.2 Number of Models

Let $b(n(T))$ be the volume of the unit ball in the Euclidean space $\mathbb{R}^{n(T)}$. It can be shown that the volume $V(B(\theta), J)$ of $B(\theta)$ in T is $V(B(\theta), J) \approx (2\gamma)^{n(T)/2} b(n(T))$. In particular, $V(B(\theta), J)$ is independent of θ to leading

order. The number, $n(T,J,\gamma)$, of models in T is defined, as in [2], [21], by $n(T,J,\gamma) = (2\gamma)^{-n(T)/2} b(n(T))^{-1} V(T,J)$. The quantity $n(T,J,\gamma)$ is independent of the choice of parameterization of T. It is a measure of the complexity of the problem of detecting the structures $M(\theta)$. If $n(T,J,\gamma)$ is small, then it is easy to detect the structures $M(\theta)$ by first covering T with sets $B(\theta(i))$, $1 \le i \le n(T,J,\gamma)$ and then checking each $\theta(i)$, in turn, to see if it is supported by the measurements.

4.3 Threshold for Detecting a Model

Let $A(\gamma,\theta) \subseteq D$ be defined by

$$A(\gamma,\theta) = \{x, x \in M(\theta') \text{ for some } \theta' \in B(\theta)\}. \tag{3}$$

The inliers for $M(\theta)$ are defined to be the points of $A(\gamma,\theta)$. Let $x(i)$, $1 \le i \le N$, be the measurements and let $n(\theta)$ be the number of measurements in $A(\gamma,\theta)$, $n(\theta) = \#\{x(i), x(i) \in A(\gamma,\theta), 1 \le i \le N\}$. If $n(\theta)$ is large, then this is evidence in favor of the detection of $M(\theta)$.

A false detection arises if there are, by chance, many measurements which are inliers for $M(\theta)$ but which do not arise from a "true" image structure, where "true" might mean a structure as seen by a human observer. If the measurements are chosen randomly and uniformly in D, then there is a small but nonzero probability that a large number of measurements will be inliers to $M(\theta)$ for some value of θ. A human observer, knowing the origin of the measurements, would not agree that $M(\theta)$ is detected.

As in [23], the problem of false detections is reduced by using a threshold r: If $n(\theta) \ge r$, then $M(\theta)$ is detected. If r is sufficiently large, then the probability of a false detection is small. However, r should not be too large, otherwise it may happen that $n(\theta) < r$ even when the structure $M(\theta)$ is present in the image. The threshold r is chosen such that, if the $x(i)$, $1 \le i \le N$, are independent samples from a random variable taking values uniformly distributed on D, then there is only a small probability that there exists $\theta \in T$ for which $n(\theta) \ge r$.

4.4 Upper Bound for Probability of False Detection

Let A be a subset of D, let $E(j,A)$ be the event that j or more of the measurements $x(i)$ are in A and let F be the probability of a false detection. An upper bound for F is obtained. The strategy in the proof is to express T as a union of sets $B(\theta(i))$, $1 \le i \le n(T,J,\gamma)$, and to sum the contribution of each $B(\theta(i))$ to F. The result is only an upper bound because the interdependencies between the events $E(r,A(\gamma,\theta))$, $\theta \in T$ are not fully taken into account.

The probability F is given by $F = \text{Prob}(\cup\{E(r,A(\gamma,\theta)), \theta \in T\})$. Let T be covered by the sets $B(\theta(i))$, $1 \le i \le n(T,J,\gamma)$. It follows that

$$\begin{aligned} F &= \text{Prob}\left(\bigcup_{i=1}^{n(T,J,\gamma)} \bigcup_{\theta \in B(\theta(i))} E(r,A(\gamma,\theta))\right) \\ &\le \sum_{i=1}^{n(T,J,\gamma)} \text{Prob}\left(\bigcup_{\theta \in B(\theta(i))} E(r,A(\gamma,\theta))\right). \end{aligned} \tag{4}$$

If r measurements are contained in $A(\gamma,\theta)$ for some $\theta \in B(\theta(i))$, then the same r measurements are contained in $\cup\{A(\gamma,\theta), \theta \in B(\theta(i))\}$, thus

$$\cup\{E(r,A(\gamma,\theta)), \theta \in B(\theta(i))\} \subseteq E(r, \cup\{A(\gamma,\theta), \theta \in B(\theta(i))\}). \tag{5}$$

If $x \in A(\gamma,\theta)$ for some $\theta \in B(\theta(i))$, then there exists $\theta' \in B(\theta)$ such that $x \in M(\theta')$. It is assumed that $\theta(i)$, θ, θ' are close enough to ensure that $J(\theta(i)) \approx J(\theta) \approx J(\theta')$. With this assumption, $D(\theta(i)||\theta) \le \gamma$, $D(\theta||\theta') \le \gamma$ and $D(\theta(i)||\theta) \le 4\gamma$. It follows that $x \in A(4\gamma,\theta(i))$ and

$$E(r, \cup\{A(\gamma,\theta), \theta \in B(\theta(i))\}) \subseteq E(r, A(4\gamma,\theta(i))). \tag{6}$$

Equations (4), (5), and (6) yield

$$\begin{aligned} F &\le \sum_{i=1}^{n(T,J,\gamma)} \text{Prob}\left(E\left(r, \bigcup_{\theta \in B(\theta(i))} A(\gamma,\theta)\right)\right) \\ &\le \sum_{i=1}^{n(T,J,\gamma)} \text{Prob}(E(r, A(4\gamma,\theta(i)))). \end{aligned} \tag{7}$$

Let $p(\gamma,\theta)$ be the probability that a random variable uniformly distributed in D takes a value in $A(\gamma,\theta)$. Recall from Section 3.2 that $V(A(\gamma,\theta),g)$ is the volume of $A(\gamma,\theta)$ under the canonical measure defined on D by the metric g. It follows that

$$p(\gamma,\theta) = V(A(\gamma,\theta),g)/V(D,g), \tag{8}$$

$$\text{Prob}(E(j,A(\theta))) = \sum_{i=j}^{N} \binom{N}{i} p(\gamma,\theta)^i (1-p(\gamma,\theta))^{N-i}. \tag{9}$$

Let $p_m(\gamma)$ be defined by $p_m(\gamma) = \sup\{p(\gamma,\theta), \theta \in T\}$. An argument similar to that used in [23] Lemma 2 establishes that $\text{Prob}(E(j,A(4\gamma,\theta)))$ is an increasing function of $p(4\gamma,\theta)$, thus (9) yields

$$\text{Prob}(E(r,A(4\gamma,\theta))) \le \sum_{i=r}^{N} \binom{N}{i} p_m(4\gamma)^i (1-p_m(4\gamma))^{N-i}. \tag{10}$$

The upper bound F_{up} for F is obtained from (7) and (10),

$$F_{up} = n(T,J,\gamma) \sum_{i=r}^{N} \binom{N}{i} p_m(4\gamma)^i (1-p_m(4\gamma))^{N-i}. \tag{11}$$

The upper bound, F_{up}, on F is general in the following sense: Let $\mathcal{A}$ be any algorithm which samples T at points θ and checks without error to see if $n(\theta) \ge r$. Then, F_{up} bounds the probability of false detection by $\mathcal{A}$ regardless of the number of samples θ and regardless of whether $\mathcal{A}$ uses the subsets $B(\theta)$ of T. A numerical investigation of F_{up} is made in Section 6.2.

5 Application to Line Detection

The theory developed in Sections 3 and 4 is applied to the detection of lines in two-dimensional images. As noted in Section 1, the theory is made mathematically tractable by replacing the Fisher information $J(\theta)$ with an asymptotic approximation $K(\theta)$. The error in the approximation is small provided the measurement noise is small.

5.1 Parameter Space for Lines

The mathematical calculations are simplified by choosing the image to be a disc rather than the usual square or rectangle. The disc is scaled to have unit radius and Cartesian

coordinates x_1, x_2 are chosen with the origin of coordinates at the center of the disc. The image coincides with the measurement space D, the metric g on D is the usual Euclidean metric, and $d\mu$ is the Lebesgue measure on D.

Let l be any line in D. If l does not contain the origin, then l is specified by giving the polar coordinates (ρ, α) of the point on l nearest to the origin. If l contains the origin, then l is specified by the coordinates $(0, \alpha)$, where α is the angle between the x_1 axis and the normal to l. If $0 \le \alpha < \pi$, then $(0, \alpha)$ and $(0, \alpha + \pi)$ specify the same line. The parameter space is $T = [0, 1) \times [0, 2\pi)$ and the equation of the line l with parameter vector $\theta = (\rho, \alpha)$ is $x_1 \cos(\alpha) + x_2 \sin(\alpha) = \rho$, $0 \le \rho < 1, 0 \le \alpha < 2\pi$.

5.2 Approximation to the Fisher information

The line $l = M(\theta)$ has the arc length parameterization

$$\begin{aligned} s \mapsto \rho(\cos(\alpha), \sin(\alpha)) + s(-\sin(\alpha), \cos(\alpha)), \\ -(1-\rho^2)^{1/2} < s < (1-\rho^2)^{1/2}. \end{aligned} \tag{12}$$

As noted in Section 3.1, it is assumed that each measurement x arises from an underlying point $\tilde{x}$ in $M(\theta)$. It is assumed that $\tilde{x}$ is uniformly distributed on $M(\theta)$. This is the most general assumption that can be made in default of any additional information about $\tilde{x}$. The probability measure dh on $M(\theta)$ is equal to the Lebesgue measure on $M(\theta)$, scaled such that $M(\theta)$ has one-dimensional volume equal to 1,

$$dh(s) = 2^{-1}(1-\rho^2)^{-1/2}\, ds, \quad -(1-\rho^2)^{1/2} < s < (1-\rho^2)^{1/2}. \tag{13}$$

The following notation is employed: $x = (x_1, x_2)$ is a point of D, $y(x)$ is the point on $M(\theta)$ closest to x and $u(x) = x - y(x)$. (More technically, D is identified with a subset of the tangent space $T_{y(x)}D$ and $\exp_{y(x)}(u(x)) = x$, where exp is the exponential map from a neighborhood of 0 in $T_{y(x)}D$ to D [9].) The Euclidean norm of $u(x)$ is denoted by $\|u(x)\|$. The asymptotic approximation $K(\theta)$ to $J(\theta)$ is obtained using (34) in Appendix B. The point $y(x) \in M(\theta)$ has a parameter value s given by $s = x.(-\sin(\alpha), \cos(\alpha)) = -x_1 \sin(\alpha) + x_2 \cos(\alpha)$. A short calculation yields

$$\begin{aligned} y(x) &= \rho(\cos(\alpha), \sin(\alpha)) + (-x_1 \sin(\alpha) \\ &\quad + x_2 \cos(\alpha))(-\sin(\alpha), \cos(\alpha)), \\ u(x) &= (x_1 \cos(\alpha) + x_2 \sin(\alpha) - \rho)\ (\cos(\alpha), \sin(\alpha)), \end{aligned}$$

$$\begin{aligned} g_{y(x)}(u(x), u(x)) &= \|u(x)\|^2 = (x_1 \cos(\alpha) + x_2 \sin(\alpha) - \rho)^2, \\ \partial^2_{\rho,\rho} g_{y(x)}(u(x), u(x)) &= 2, \\ \partial^2_{\rho,\alpha} g_{y(x)}(u(x), u(x)) &= 2(x_1 \sin(\alpha) - x_2 \cos(\alpha)), \\ \partial^2_{\alpha,\alpha} g_{y(x)}(u(x), u(x)) &= 2(x_1 \sin(\alpha) - x_2 \cos(\alpha))^2 \\ &\quad - 2(x_1 \cos(\alpha) + x_2 \sin(\alpha))^2 \\ &\quad + 2\rho(x_1 \cos(\alpha) + x_2 \sin(\alpha)). \end{aligned} \tag{14}$$

If $x = y(x)$, then x is on $M(\theta)$. In this case,

$$\begin{aligned} s &= -x_1 \sin(\alpha) + x_2 \cos(\alpha), \\ \rho &= x_1 \cos(\alpha) + x_2 \sin(\alpha). \end{aligned} \tag{15}$$

It follows from (13), (14), (15), and (34) that the asymptotic approximation $K(\theta)$ to the Fisher information $J(\theta)$ is given by

$$\begin{aligned} K(\theta) &= \frac{1}{4t} \int_{-\sqrt{1-\rho^2}}^{\sqrt{1-\rho^2}} \begin{pmatrix} 1 & -s \\ -s & s^2 \end{pmatrix} \frac{ds}{(1-\rho^2)^{1/2}} \\ &= \frac{1}{2t} \begin{pmatrix} 1 & 0 \\ 0 & (1-\rho^2)/3 \end{pmatrix}, \quad \theta \in T. \end{aligned} \tag{16}$$

5.3 Curvature and volume of T

The parameter space $T = [0, 1) \times [0, 2\pi)$, as defined in Section 5.1, is flat under the Euclidean metric in the plane. However, the Euclidean metric on T has no statistical meaning. It is simply a consequence of the choice of parameterization of the lines in D. The Rao metric on T does have a statistical meaning, as described in Section 3.2. Under the Rao metric, T is curved. The scalar curvature [20] R of T under the approximation $K(\theta)$ to the Rao metric is $R = 4t/(1-\rho^2)^2$, $0 \le \rho < 1$, thus T has positive scalar curvature at every point. The Gaussian curvature of T is $R/2$.

The definition (2) of the sets $B(\theta)$, $\theta = (\rho, \alpha)$, becomes

$$B(\theta) = \{(\rho', \alpha'), (4t)^{-1}(\rho' - \rho)^2 + (12t)^{-1}(1-\rho^2)(\alpha' - \alpha)^2 \le \gamma\}. \tag{17}$$

It follows from (17) that the boundary of $B(\theta)$ is an ellipse with its axes aligned with the coordinate axes. As ρ increases toward 1, the length of the ρ axis of the ellipse is constant but the length of the α axis increases, showing that estimates of the orientations of lines decrease in accuracy for lines near to the circumference of D.

The canonical measure $\tau(\theta)\, d\theta$ defined on T by K is

$$\tau(\theta)\, d\theta = |\det(K(\theta))|^{1/2}\, d\theta = (2\sqrt{3}\, t)^{-1}(1-\rho^2)^{1/2} d\rho d\alpha \quad \theta \in T. \tag{18}$$

The volume $V(T, K)$ of T and the number $n(T, K, \gamma)$ of models are

$$V(T, K) = \int_T \tau(\theta)\, d\theta = \pi^2/(4\sqrt{3}\, t), \tag{19}$$

$$n(T, K, \gamma) = V(T, K)/(2\gamma b(2)) = \pi/(8\sqrt{3}\, \gamma t). \tag{20}$$

5.4 Value of γ

The theory developed in Sections 3 and 4 does not assign a value to γ. In the case of line detection, a plausible value is found as follows: Note first that, if x is away from the boundary of D, then $p(x|\tilde{x})$ is very closely approximated by a Gaussian density $p(x|\tilde{x}) = (4\pi t)^{-1} \exp(-\|x - \tilde{x}\|^2/(4t))$. It follows that $2t = \sigma^2$, where σ is the standard deviation of the measurement noise for the components x_1, x_2 of x.

The sets $A(\theta)$ are narrow strips containing the line θ. To a first approximation in t, $A(\theta)$ is symmetric about the line θ. If γ is large, then the sets $A(\gamma, \theta) \subseteq D$ are relatively large and it is difficult to justify including the boundary points of $A(\gamma, \theta)$ as inliers to $M(\theta)$. This suggests choosing γ such that all the points of $A(\gamma, \theta)$ are within 2σ of $M(\theta)$, where $\sigma = (2t)^{1/2}$ is the standard deviation of the measurement noise. Let θ be a line through the origin, i.e., $\rho = 0$. It follows from (26) below that the furthest distance of a point of $A(\gamma, \theta)$ from $M(\theta)$ is $4(t\gamma)^{1/2}$ to leading order in t. On setting $2\sigma = 4(t\gamma)^{1/2}$, it follows that $\gamma = 1/2$.

5.5 Comparison with the Hough Transform

The Hough transform buckets are similar to the sets $B(\theta)$ defined by (2). The innovation in this paper is that the size and shape of each $B(\theta)$ and the total number $n(T, K, \gamma)$ of sets $B(\theta)$ needed to cover T are calculated using the approximation $K(\theta)$ to the Rao metric on T. The Hough transform is analyzed in [11], but the measurements in [11] are line segments rather than points. Yuen and Hlavac [26] and Lam et al. [16] use geometric arguments to obtain values for the sides $\Delta\rho$, $\Delta\alpha$ of the Hough transform buckets.

The dimensions of $B(\theta)$ are now compared with $\Delta\rho$, $\Delta\alpha$. It follows from (17) that, if $\gamma = 1/2$, $\theta = (\rho, \alpha)$, then

$$\begin{aligned}\max\{|\rho' - \rho|, (\rho', \alpha') \in B(\theta)\} &= (4t\gamma)^{1/2} = (2t)^{1/2},\\ \max\{|\alpha' - \alpha|, (\rho', \alpha') \in B(\theta)\} &= \left((12t\gamma)/(1-\rho^2)\right)^{1/2} \geq (6t)^{1/2},\end{aligned} \tag{21}$$

which suggest $\Delta\rho = 2(2t)^{1/2}$, $\min\{\Delta\alpha\} = 2(6t)^{1/2}$. The equation $\Delta\rho = 2(2t)^{1/2}$ is similar to an expression for $\Delta\rho$ given as (7) in [26], provided the $(N+1)\times(N+1)$ pixel^2 image in [26] is scaled to the unit square. The equation $\min\{\Delta\alpha\} = 2(6t)^{1/2}$ is similar to an expression for $\Delta\alpha$ given as (13) in [26]. It follows from (21) and $\Delta\rho = 2(2t)^{1/2}$ that $(1-\rho^2)^{1/2}\Delta\alpha = \sqrt{3}\,\Delta\rho$ which is similar to the formula $2(1-\rho^2)^{1/2}\sin(\Delta\alpha/2) = \Delta\rho$ given as (13) in [16].

6 Probability of a False Detection

A major advantage of the approach to line detection described in Section 5 is that it allows the calculation of upper bounds for the probability of false detection. An upper bound is obtained in Section 6.1 and checked using synthetic data in Section 6.2.

6.1 Upper Bound for the Probability of a False Detection

Let x be a measurement sampled using the uniform density on D. The probability $p(\gamma, \theta)$ that x is an inlier for $M(\theta)$ is given by (8). The two-dimensional volume $V(A(\gamma,\theta), g)$ is the usual Euclidean area of $A(\gamma,\theta)$. Let ds be the length element on $M(\theta)$ and let $\max\{|u|\}$ be the distance from a point on $M(\theta)$ to the boundary of $A(\gamma,\theta)$ in the direction normal to $M(\theta)$. It is assumed that $1-\rho^2 \neq O(t)$. Then, to a first approximation, the line $M(\theta)$ runs down the middle of $A(\gamma,\theta)$ and the area of $A(\gamma,\theta)$ can be estimated by integrating $2\max\{|u|\}\,ds$ along $M(\theta)$. The points on the boundary of $A(\gamma,\theta)$ are on lines with parameter values θ' in the boundary of $B(\theta)$, thus the first step in estimating $V(A(\gamma,\theta), g)$ is to examine the boundary points of $B(\theta)$.

It follows from (17) that the boundary of $B(\theta)$ is the ellipse $(4t)^{-1}(\rho'-\rho)^2 + (12t)^{-1}(1-\rho^2)(\alpha'-\alpha)^2 = \gamma$. The points $\theta' = (\rho', \alpha')$ on the boundary of $B(\theta)$ are parameterized by $\beta \in [0, 2\pi)$ as follows:

$$\begin{aligned}\rho' &= \rho + (4t\gamma)^{1/2}\cos(\beta),\\ \alpha' &= \alpha + (12t\gamma)^{1/2}(1-\rho^2)^{-1/2}\sin(\beta).\end{aligned} \tag{22}$$

Let y be a point on $M(\theta)$. The coordinates of y in the arc length parameterization (12) of $M(\theta)$ are

$$y = \rho(\cos(\alpha), \sin(\alpha)) + s(-\sin(\alpha), \cos(\alpha)). \tag{23}$$

Let x be a point in $A(\gamma,\theta)$ such that $x-y$ is normal to $M(\theta)$ and let u be the signed distance from y to x, $u = \pm\|x-y\|$. It follows from (23) that

$$x = (\rho+u)(\cos(\alpha), \sin(\alpha)) + s(-\sin(\alpha), \cos(\alpha)). \tag{24}$$

The expression $|u|$ is a maximum when x is on a line $M(\theta')$ for some θ' in the boundary of $B(\theta)$. It follows that $\theta' = (\rho', \alpha')$, where ρ', α' are given by (22). The condition that x is on $M(\theta')$ is $x.(\cos(\alpha'), \sin(\alpha')) = \rho'$. It follows from this equation and (24) that $(\rho+u)\cos(\alpha'-\alpha) + s\,\sin(\alpha'-\alpha) = \rho'$, which yields

$$\begin{aligned}u &= \rho' - \rho - s(\alpha'-\alpha) + O(t)\\ &= 2(t\gamma)^{1/2}\left(\cos(\beta) - \sqrt{3}\,s\,\sin(\beta)(1-\rho^2)^{-1/2}\right) + O(t).\end{aligned} \tag{25}$$

On taking the maximum of $|u|$ over $0 \leq \beta < 2\pi$, it follows from (25) that

$$\max_{\beta}\{|u|\} = 2(t\gamma)^{1/2}\left(1 + 3s^2(1-\rho^2)^{-1}\right)^{1/2} + O(t). \tag{26}$$

Now that $\max\{|u|\}$ is obtained, $V(A(\gamma,\theta), g)$ is estimated by the following integral along $M(\theta)$,

$$\begin{aligned}V(A(\gamma,\theta), g) &= 2\int_{M(\theta)} \max_{\beta}\{|u|\}\,ds + O(t)\\ &= 8(t\gamma)^{1/2}\int_0^{\sqrt{1-\rho^2}}\left(1+3s^2(1-\rho^2)^{-1}\right)^{1/2} ds + O(t)\\ &= 4(t\gamma(1-\rho^2))^{1/2}\left(2 + (\sqrt{3})^{-1}\sinh^{-1}\left(\sqrt{3}\right)\right) + O(t).\end{aligned} \tag{27}$$

It follows from (8) and (27) that

$$\begin{aligned}&p(\gamma,\theta) =\\ &4\pi^{-1}\left(t\gamma(1-\rho^2)\right)^{1/2}\left(2 + (\sqrt{3})^{-1}\sinh^{-1}\left(\sqrt{3}\right)\right) + O(t),\end{aligned} \tag{28}$$

thus the supremum, $p_m(\gamma)$, of the probabilities $p(\gamma,\theta), \theta \in T$ is

$$\begin{aligned}p_m(\gamma) &= \sup\{p(\gamma,\theta), \theta \in T\}\\ &= 4\,\pi^{-1}(t\gamma)^{1/2}\left(2 + (\sqrt{3})^{-1}\sinh^{-1}\left(\sqrt{3}\right)\right) + O(t).\end{aligned} \tag{29}$$

When $p_m(\gamma)$ is used in numerical calculations, the $O(t)$ term is omitted from (29).

It follows from (11), (20), and (29) that the upper bound F_{up} for the probability of a false detection of a line is

$$F_{up} = \frac{\pi}{8\sqrt{3}\,\gamma t}\sum_{i=r}^{N}\binom{N}{i} p_m(4\gamma)^i(1-p_m(4\gamma))^{N-i}. \tag{30}$$

6.2 Numerical Results

The expression (30) for F_{up} was evaluated for a range of values of r with $\gamma = 1/2$, $t = 1/2 \times 10^{-4}$. The graphs of $\ln(F_{up})$ as a function of r are shown in Fig. 1 for the two cases $N = 20$ (lower graph) and $N = 40$ (upper graph). The graphs show that to achieve $F_{up} \leq 1$ for $N = 20$, a threshold of $r = 6$ almost suffices and for $N = 40$, $r = 9$ suffices. It is apparent from Fig. 1 that F_{up} decreases rapidly as r increases. For example, the

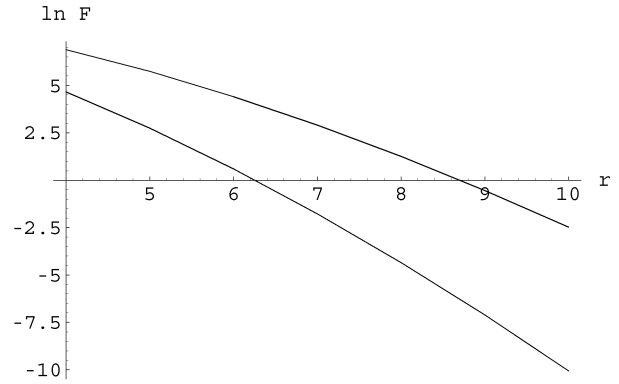


Fig. 1. Graphs of $\ln(F_{up})$ as a function of r for $N = 40$ (upper) and $N = 20$ (lower).

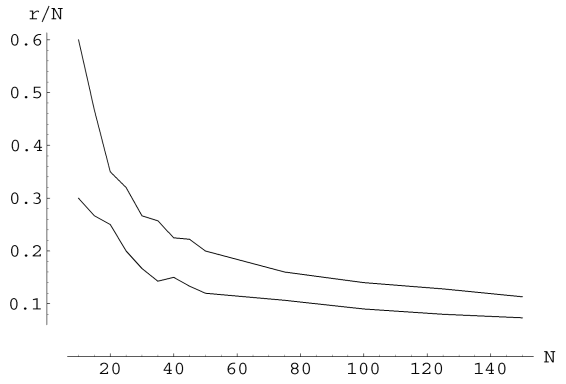


Fig. 2. Calculated (upper) and empirical (lower) graphs of r/N as a function of N.

upper graph in Fig. 1 can be approximated by a straight line with gradient $-1.56\ldots$. If r is increased by 1 in the region where the straight line approximation is accurate, then $F_{up}(r+1)/F_{up}(r) \approx \exp(-1.56) = 0.21\ldots$.

The fact that F_{up} is an upper bound for the probability of false detection suggests that the values of r predicted using F_{up} are too high. This suggestion is supported by the graphs shown in Fig. 2. As in Fig. 1, $\gamma = 1/2, t = 1/2 \times 10^{-4}$. Let $r(N)$ be the integer such that $F_{up}(r(N)) \leq 1$ and $F_{up}(r(N)-1) > 1$. The upper graph in Fig. 2 shows $r(N)/N$ as a function of N for $10 \leq N \leq 150$. The lower graph is obtained as follows: A set of N points is sampled from the uniform distribution on D. Let r_{min} be the least value of r for which no lines are detected by Algorithm 1 which is described in Section 7.2 below. The sampling is repeated three times for each value of N, yielding $r_{min}(N,1)$, $r_{min}(N,2)$, $r_{min}(N,3)$. Let $r_{av}(N)$ be defined by $r_{av}(N) = (r_{min}(N,1) + r_{min}(N,2) + r_{min}(N,3))/3$. The lower graph in Fig. 2 shows $r_{av}(N)/N$ as a function of N for $10 \leq N \leq 150$.

It is clear from Fig. 2 that $r_{av}(N)/N < r(N)/N$. For example, $r(150) = 17$, $r_{min}(150) = 7$. The downward slope of the graph for $N \mapsto r_{av}(N)/N$ supports the conjecture made in Section 1 that the ratio of the minimum acceptable number of inliers to N tends to 0 as N tends to infinity. The fact that the graph of $N \mapsto r(N)/N$ also has a downward slope suggests that the bound F_{up} might be accurate enough to support a proof of the conjecture.

7 Experiments

An algorithm for detecting lines was implemented in Mathematica [25] and tested by comparing its results with those obtained from a publicly available Matlab implementation of the Hough transform [12]. The algorithm is described in Sections 7.1 and 7.2. Experimental results are reported in Section 7.3 and the time complexity of the algorithm is estimated in Section 7.4.

7.1 Preliminaries

In the Mathematica program, the parameter space $T = [0,1) \times [0, 2\pi)$ is sampled at the points of a grid G which is square in the usual Euclidean metric on T. The grid G is chosen to be fine, i.e., with more than $n(T, K, \gamma)$ points, in order to make sure that each subset $B(\theta)$ of T contains at least one point of G. The choice of a square grid is not optimal, but it has the advantage of simplicity.

The size of the grid G is $n_g \times n_g$, where $n_g = \lceil 2\pi/h \rceil$ and

$$h = \min_{\rho} \max\{|\alpha' - \alpha|, (\rho', \alpha') \in B(\theta)\} = \pi/(\gamma t)^{1/2}. \quad (31)$$

The points of G are labeled by pairs of integers (i,j), $0 \leq i, j < n_g$. The point (i,j) in G has coordinates $\theta(i,j) = (\rho(i), \alpha(j))$ in T, where $\rho(i) = i/n_g, 0 \leq i < n_g, \alpha(j) = 2\pi j/n_g$, $0 \leq j < n_g$. Each point (i,j) in G is the center of a set $B(i,j)$ of points of G defined by

$$B(i,j) = \{(k,l), (k,l) \in G \text{ and } \theta(k,l) \in B(\theta(i,j))\}, \quad 0 \leq i < n_g, 0 \leq j < n_g. \quad (32)$$

Let $x(k)$ be one of the measurements. The curve $C(k)$ of points (ρ, α) in T corresponding to the lines in D containing $x(k)$ is defined by $\rho = x(k).(\cos(\alpha), \sin(\alpha))$, $0 \leq \rho < 1$, $0 \leq \alpha < 2\pi$. For each measurement, $x(k)$ define the function $j \mapsto i(k,j)$, $0 \leq j < n_g$, and the set $S(k)$ by $i(k,j) = \text{Round}\ (n_g x(k).(\cos(\alpha(j)), \sin(\alpha(j))))$, $0 \leq j < n_g$, $S(k) = \bigcup_{j=0}^{n_g - 1} B(i(k,j), j)$, $1 \leq k \leq N$. The set $S(k)$ contains the points of G close to $C(k)$ in the following sense: If $\theta(l,m) \in S(k)$, then $x(k)$ is an inlier for the line $M(\theta(l,m))$.

7.2 Algorithm 1

The parameters for Algorithm 1 are t, e_f, where e_f is a user defined threshold for the probability of a false detection. The variable γ is assigned the value $1/2$, as discussed in Section 5.4. The threshold r for detection is $r = r_{av}(N)$, where $r_{av}(N)$ is as defined in Section 6.2. The threshold r can be calculated at runtime using γ, t, e_f and the number N of measurements, but, to increase efficiency, it is assumed that a suitable table of values $N \mapsto r_{av}(N)$ is computed offline and r is obtained as an input from the table at run time. The output of Algorithm 1 is a list L of points of G corresponding to lines with r or more inliers.

By definition, a run in a list S is a sequence of successive identical elements of S. The function $\text{maxrun}(S)$ returns the length of the largest run in S and the function $\text{maxrunentry}(S)$ returns an element of S which belongs to a run in S with length equal to $\text{maxrun}(S)$.

Fig. 3. Gray-scale version of the original image 0017.jpg.

Algorithm 1

1. Input t, e_f, r and the measurements $x(i)$, $1 \leq i \leq N$.
2. Compute the sets $S(i)$, $1 \leq i \leq N$.
3. $W \leftarrow \emptyset$.
4. While True

   4.1. $S \leftarrow \mathrm{Sort}(\mathrm{List}(S(1), \ldots, S(N)))$.
   4.2. If $\mathrm{maxrun}(S) < r$, Goto 5.
   4.3. $W \leftarrow W \cup \{\mathrm{maxrunentry}(S)\}$.
   4.4. If $\mathrm{maxrunentry}(S) \in S(i)$, $S(i) \leftarrow \emptyset$, $1 \leq i \leq N$.
5. EndWhile.
6. $L \leftarrow \emptyset$;
7. While $W \neq \emptyset$,

   7.1. $(l, m) = \mathrm{argmax}_{(i,j)}\{|B(i, j) \cap W|, (i, j) \in W\}$;
   7.2. $L \leftarrow L \cup \{(l, m)\}$;
   7.3. $W \leftarrow W \setminus B(l, m)$;
8. EndWhile.
9. Output L.
10. Stop.

Line 4.4 in Algorithm 1 removes all the measurements which are inliers to a line, once the line has been detected. If the inliers are not removed in this way, then the algorithm fails when a large number of measurements are grouped close to a point x in the image: All the points in G which correspond to lines passing near to x are added to L. The While loop at line 7 extracts from W the set L of representative points of G.

7.3 Results

Algorithm 1 was applied to image 0017.jpg from the publicly available PETS 2001 database.[1] A gray-scale version of this image is shown in Fig. 3. The parameters for Algorithm 1 are shown in Table 1. The image 0017.jpg was converted to gray scale and three images $I(1)$, $I(2)$, $I(3)$ were selected from it. Each $I(i)$ was square and centered at the center $(164, 123)$ of 0017.jpg. The Sobel edge detector was applied to each image $I(i)$ and the magnitude of the response calculated for each pixel. The N pixels with the largest responses were selected as measurements, where N is given for each $I(i)$ by the appropriate entry in Table 1. The Sobel edge magnitudes were not filtered or edited in any other way. The number N is proportional to the width of $I(i)$ rather than the area, because the structures to be detected, i.e., lines, are one-dimensional. The parameter t depends on the size $w(i) \times w(i)$ pixel^2 of $I(i)$, $t = 2w(i)^{-2}$. This is equivalent to assuming that the standard deviation of the measurement noise is equal to 1 pixel. The last column of Table 1 shows the lines detected in each $I(i)$.

The images $I(i)$ are shown in Fig. 4 with the detected lines superimposed on them. The white circles mark the boundary of the measurement space D. The lines include structures in the buildings as well as structures in the straight row of cars parked in front of the buildings. Some lines in the original gray-level images are undetected because they do not contribute to the set of N measurements. Fig. 5 shows the detected lines superimposed on the measurements. For comparison, the results obtained using a publicly available implementation of the Hough transform [12] are shown in Fig. 6. The measurements are the same as those shown in Fig. 5. Note that the implementation [12] contains a parameter p which controls the number of detected lines in the following way: Let B be a bucket for the Hough transform and let $a(B)$ be the integer defined in Section 2. A line is detected in the bucket B if $a(B) \geq p \max_C\{a(C)\}$. If p is small, then a large number of lines is detected. The value of p is chosen for each $I(i)$ such that the number of lines detected by the Hough transform is similar to the number of lines detected by Algorithm 1. The results for $I(1)$ and $I(2)$ in Fig. 6 suggest that the Hough transform, as implemented in [12], has a tendency to detect sets of near concurrent lines.

7.4 Time Complexity

The time complexity of Algorithm 1 is estimated. The length of the curve $C(k)$ in T under the Euclidean metric is $O(1)$. It follows that the number $|S(k)|$ of points in $S(k)$ is $O(n_g)$. The time taken to construct $S(k)$ is $O(|S(k)|)$. The length of S is $O(N n_g)$, thus the time taken to sort S is $O(N n_g \ln(N n_g))$, which is the leading order term in the time complexity of Algorithm 1. Equation (31) is used to substitute for $n_g = \lceil 2\pi/h \rceil$ to give the time complexity $O(N(\gamma t)^{-1/2} \ln(N(\gamma t)^{-1/2}))$.

For comparison, consider a second algorithm, Algorithm 2, which checks each of $n(T, K, \gamma)$ models in turn to see how many inliers it has. If each check has a time complexity $O(N)$, then the total time complexity for the second algorithm is $O(N(\gamma t)^{-1})$.

The time complexity of RANSAC is estimated. The probability that two measurements $x(i)$, $x(j)$ are inliers to the same line is, in the worst case, $r(N)^2/N^2$, where $r(N)$ is the threshold for detection. Let u be the number of random selections of pairs $x(i), x(j)$ sufficiently large to ensure that the probability of obtaining two inliers to the same line is $1 - \delta$, where δ is a small constant. It follows that $(1 - r(N)^2/N^2)^u \approx \delta$, thus $u \approx N^2 \ln(\delta^{-1})/r(N)^2$. If the time taken to find the inliers for a given line $\langle x(i), x(j)\rangle$ is $O(N)$, then the time complexity of RANSAC is $O(1)Nu = O(N^3 \ln(\delta^{-1})/r(N)^2)$.

The above estimates of time complexity suggest that Algorithm 1 and Algorithm 2 have a lower time complexity than RANSAC for large N, especially if $r(N)/N$ becomes small. On the other hand, Algorithm 1 and Algorithm 2 have a large time complexity if t is small.

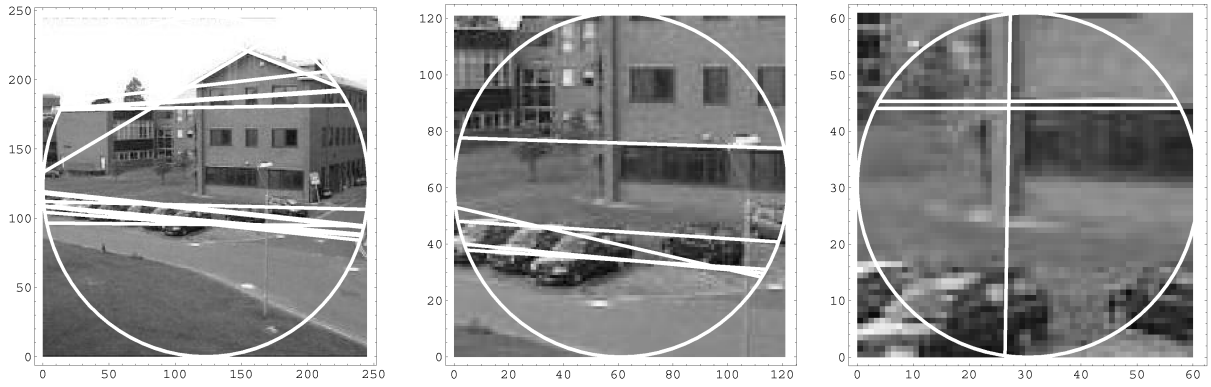
8 Conclusion

The probability density function $p(x|\theta)$ for a measurement x given an image structure θ contains information about the

1. ftp://pets2001.cs.rdg.ac.uk/PETS2001/DATASET1/TESTING/CAMERA1_JPEGS.

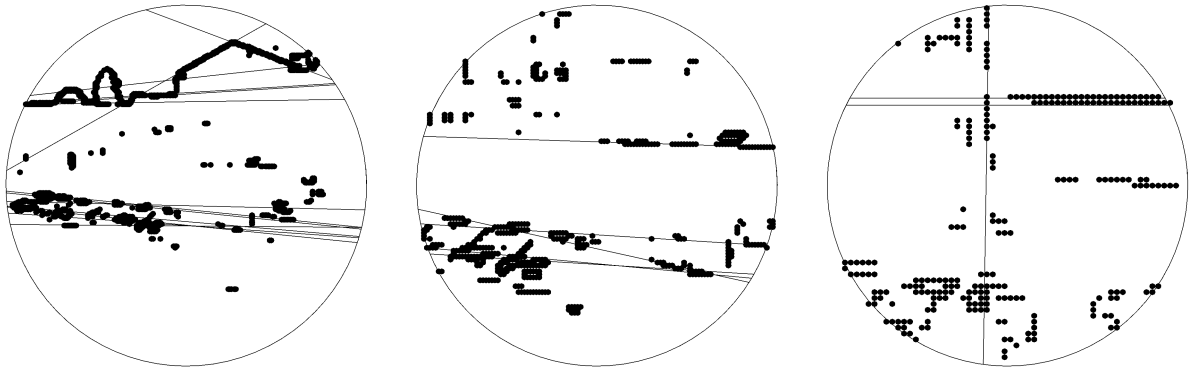
TABLE 1
Parameters Used to Obtain the Results Shown in Figs. 3, 4, 5

	size	N	γ	$t \times 10^4$	r	p	no.lines (Alg.1)	no.lines(HT)
I(1)	244×244	1000	0.5	0.34	32	0.45	12	13
I(2)	122×122	500	0.5	1.3	27	0.80	8	6
I(3)	61×61	250	0.5	5.3	26	0.70	4	3



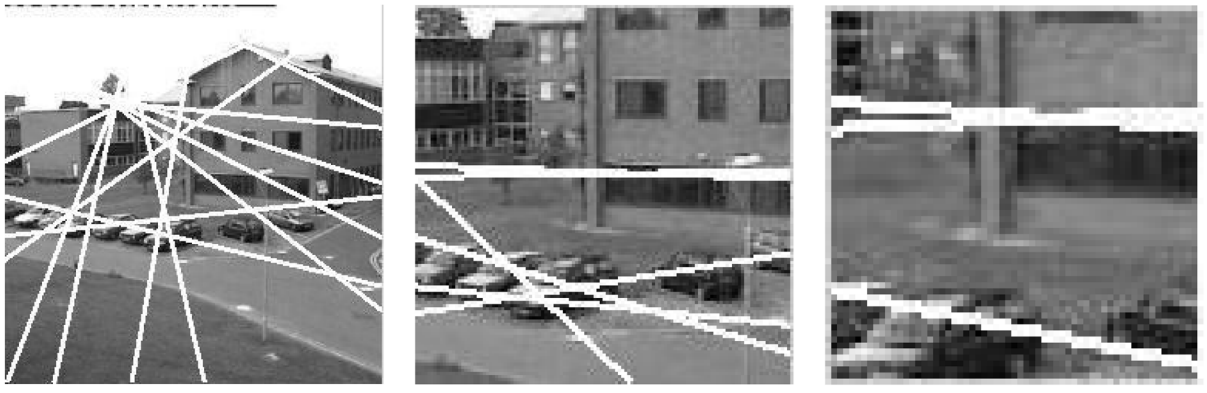
(a) (b) (c)

Fig. 4. Detected lines for (a) $I(1)$, (b) $I(2)$, and (c) $I(3)$.



(a) (b) (c)

Fig. 5. Measurements and detected lines for (a) $I(1)$, (b) $I(2)$, and (c) $I(3)$.



(a) (b) (c)

Fig. 6. Lines detected using the Hough transform for (a) $I(1)$, (b) $I(2)$, and (c) $I(3)$.

parameter space T in which θ takes values. This information leads to a metric on T known in statistics as the Rao metric. The Rao metric is used to define a class of optimal algorithms for detecting structures θ in an image. The algorithms are optimal in that they have the least detection threshold required to reduce the probability of false detection below a specified limit in the presence of uniformly distributed outliers. The prior information needed by the algorithms consists of the density $p(x|\theta)$ and a single additional parameter: an upper bound e_f on the probability of a false detection. All the other parameters in the algorithms are calculated from $p(x|\theta)$ and e_f. An upper bound for the probability of a false detection is obtained under the assumption that the outliers are uniformly distributed.

Line detection is a special case in which the structures θ are lines in the image. Experiments show that the new algorithm detects lines at least as well as Hough transform-based algorithms. The advantage of the new algorithm is that the parameters of the algorithm, apart from a user defined threshold on the probability of a false detection, are deduced from $p(x|\theta)$. The time complexity of the new algorithm is less than the time complexity of RANSAC if the number of measurements is large.

Possible directions for future research include:

1. Devise algorithms for detecting image structures other than lines;
2. Find better methods for choosing the sample points in the parameter manifold;
3. Improve the upper bound (11) on the probability of a false detection;
4. Improve the speed of the detection algorithms;
5. Improve the choice of density $p(x|\theta)$, for example by taking into account the statistics of images [18].

APPENDIX A

DEFINITION OF $p(x|\theta)$

As noted in Section 3.1, the conditional density $p(x|\theta)$ is obtained as the solution to the heat equation on the measurement space D. The space D has a Riemannian metric g which depends on the format of the measurements. The simplest case is $D \subseteq \mathbb{R}^d$ with g equal to the Euclidean metric. Let $\triangle_x$ be the Laplace-Beltrami operator on D [3]. The sign of $\triangle_x$ is chosen such that, in the special case $D \subseteq \mathbb{R}^d$, $\triangle_x$ is the negative of the Laplacian, i.e., if $f : \mathbb{R}^d \to \mathbb{R}$ is a C^2 function, then $\triangle_x f = -\sum_{i=1}^d \partial_{x_i}^2 f$. The heat equation on D is $\partial_t f(x,t) + \triangle_x f(x,t) = 0$, $x \in D, t \geq 0$. The density $p(x|\theta) = p_t(x|\theta)$ is obtained by solving the heat equation with the initial condition that that $f(x,0)$ is the generalized function defined by the measure dh on $M(\theta)$. If t is small, then $p_t(x|\theta)$ is concentrated in a neighborhood of $M(\theta)$.

APPENDIX B

ASYMPTOTIC APPROXIMATION TO $p(x|\theta)$

The Riemannian metric g defines an inner product on the tangent space $T_x D$ of D at x. If $u, v \in T_x D$, then the inner product of u, v is written as $g_x(u,v)$. The geodesic distance between $x, y \in D$ is $\mathrm{dist}_g(x,y)$. If $\mathrm{dist}_g(x,y)$ is small, then it is equal to the length of the shortest geodesic from x to y.

Let the function $w : D \times T \to \mathbb{R}$ be defined by $w(x,\theta) = \min_{y \in M(\theta)}\{\mathrm{dist}_g(x,y)\}$, $x \in D$, $\theta \in T$. If x is sufficiently close to $M(\theta)$, then there is a unique point $y(x) \in M(\theta)$ such that $w(x,\theta) = \mathrm{dist}_g(x, y(x))$ and there exists $u(x) \in T_{y(x)} D$ such that $\exp_{y(x)}(u(x)) = x$, where $\exp$ is the exponential map from an open neighborhood of 0 in $T_{y(x)} D$ to D. It follows from the properties of the exponential map that $w(x,\theta)^2 = g_{y(x)}(u(x), u(x))$. It can be shown that $\ln p_t(x|\theta)$ has the asymptotic approximation $\ln p_t(x|\theta) \sim -(4t)^{-1} w(x,\theta)^2 + O(1)$, $x \in D$, $t > 0$. This approximation is accurate provided $M(\theta)$ does not have large curvatures over an $O(t^{1/2})$ length scale. The resulting asymptotic approximation to the Fisher information $J(\theta)$ is

$$\begin{aligned} J_{ij}(\theta) &\sim \frac{1}{4t} \int_D \left(\partial^2_{\theta_i,\theta_j} w(x,\theta)^2 \right) p_t(x|\theta)\, d\mu(x) \\ &\sim \frac{1}{4t} \int_{M(\theta)} \left[\partial^2_{\theta_i,\theta_j} w(x,\theta)^2 \right]_{x=y} dh(y) \\ &\sim \frac{1}{4t} \int_{M(\theta)} \left[\partial^2_{\theta_i,\theta_j} g_{y(x)}(u(x),u(x)) \right]_{x=y} dh(y), \\ &\quad 1 \leq i,j \leq n(T). \end{aligned} \tag{33}$$

It follows from (33) that $J(\theta) \sim K(\theta)$, where $K(\theta)$ is the matrix defined by

$$K_{ij}(\theta) = \frac{1}{4t} \int_{M(\theta)} \left[\partial^2_{\theta_i,\theta_j} g_{y(x)}(u(x),u(x)) \right]_{x=y} dh(y), \quad 1 \leq i,j \leq n(T). \tag{34}$$

In the application to line detection in Section 5.2, g is the Euclidean metric on D and $g_{y(x)}(u(x),u(x)) = \|u(x)\|^2 = \|x - y(x)\|^2$.

ACKNOWLEDGMENTS

The author would like to thank G. Hamarneh, K. Althoff, and R. Abu-Gharbieh for making their implementation of the Hough transform publicly available on the Web. Thanks are also due to Mian Zhou for obtaining the experimental results shown in Fig. 6, to James Ferryman for permission to use an image from the PETS 2001 database, and to the referees for their comments, including the provision of references [6], [11], and [23].

REFERENCES

[1] S.-I. Amari, *Differential-Geometrical Methods in Statistics.* 1985.
[2] V. Balasubramanian, "A Geometric Formulation of Occam's Razor for Inference of Parametric Distributions," Report No. PUPT-1588, Dept. of Physics, Princeton Univ., http://www.arxiv.org/list/nlin.AO/9601, 1996.
[3] I. Chavel, *Eigenvalues in Riemannian Geometry.* Academic Press, 1984.
[4] T.M. Cover and J.A. Thomas, *Elements of Information Theory.* Wiley, 1991.
[5] M.P. Do Carmo, *Riemannian Geometry.* Birkhauser, 1993.
[6] M.A. Fischler and R.C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Comm. ACM,* pp. 381-395, 1981.
[7] R.A. Fisher, "On the Mathematical Foundations of Theoretical Statistics," *Philosophical Trans. Royal Soc. of London, Series A,* vol. 222, pp. 309-368, 1922.
[8] D.A. Forsyth and J. Ponce, *Computer Vision, a Modern Approach.* Prentice Hall, 2003.

[9] S. Gallot, D. Hulin, and J. LaFontaine, *Riemannian Geometry,* second ed. Universitext, Springer, 1990.
[10] R.C. Gonzalez and R.E. Woods, *Digital Image Processing.* Prentice Hall, 2002.
[11] W.E.L. Grimson and D.P. Huttenlocher, "On the Sensitivity of the Hough Transform for Object Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, no. 3, Mar. 1990.
[12] G. Hamarneh, K. Althoff, and R. Abu-Gharbieh, "Automatic Line Detection," http://www.cs.toronto.edu/ghassan/phd/compvis/cvreporthtml/CV_report.htm, 1999.
[13] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision.* Cambridge Univ. Press, 2000.
[14] J. Illingworth and J. Kittler, "A Survey of the Hough Transform," *Computer Vision, Graphics, and Image Processing,* vol. 43, pp. 221-238, 1988.
[15] *Breakthroughs in Statistics. Volume 1: Foundations and Basic Theory,* S. Kotz and N.L. Johnson, eds., Springer-Verlag, 1992.
[16] W.C.Y. Lam, L.T.S. Lam, K.S.Y. Yuen, and D.N.K. Leung, "An Analysis on Quantizing the Hough Space," *Pattern Recognition Letters,* vol. 15, pp. 1127-1135, 1994.
[17] V.F. Leavers, *Shape Detection in Computer Vision Using the Hough Transform.* Springer Verlag, 1992.
[18] A.B. Lee, D. Mumford, and J. Huang, "Occlusion Models for Natural Images: A Statistical Study of a Scale-Invariant Dead Leaves Model," *Int'l J. Computer Vision,* vol. 41, pp. 35-59, 2001.
[19] S.J. Maybank, "Fisher Information and Model Selection for Projective Transformations of the Line," *Proc. Royal Soc. of London, Series A,* vol. 459, pp. 1-21, 2003.
[20] C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation.* W.H. Freeman, 1973.
[21] J. Myung, V. Balasubramanian, and M.A. Pitt, "Counting Probability Distributions: Differential Geometry and Model Selection," *Proc. Nat'l Academy of Science,* vol. 97, pp. 11170-11175,
[22] C.R. Rao, "Information and the Accuracy Attainable in the estimation of Statistical Parameters," *Bull. Calcutta Math. Soc.,* vol. 37, pp. 81-91, 1945.
[23] C.V. Stewart, "MINPRAN: A New Robust Estimator for Computer Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, pp. 925-938, 1995.
[24] M. Werman and D. Keren, "A Novel Bayesian Method for Fitting Parametric and Non-Parametric Models to Noisy Data," *Proc. Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 552-558, 1999.
[25] S. Wolfram, *The Mathematica Book,* fourth ed. Cambridge Univ. Press, 1999.
[26] S.Y.K. Yuen and V. Hlavac, "An Approach to Quantization of Hough Space," *Proc. Seventh Scandinavian Conf. Image Analysis,* pp. 733-740, 1991.

Stephen J. Maybank received the BA degree in mathematics from King's College Cambridge in 1976 and the PhD degree in computer science from Birkbeck College, University of London in 1988. He was a research scientist at GEC from 1980 to 1995, first at MCCS, Frimley, and then, from 1989, at the GEC Marconi Hirst Research Centre in London. In 1995, he became a lecturer in the Department of Computer Science at the University of Reading and, in 2004, he became a professor in the School of Computer Science and Information Systems at Birkbeck College, University of London. His research interests include camera calibration, visual surveillance, tracking, filtering, applications of projective geometry to computer vision and applications of probability, statistics, and information theory to computer vision. He is the author of more than 85 scientific publications and one book. For further information, see http://www.dcs.bbk.ac.uk/~sjmaybank. He is a member of the IEEE.