Near-Optimal Detection of Geometric Objects by Fast Multiscale Methods

Ery Arias-Castro, David L. Donoho, *Member, IEEE*, and Xiaoming Huo, *Senior Member, IEEE*

Abstract—We construct detectors for "geometric" objects in noisy data. Examples include a detector for presence of a line segment of unknown length, position, and orientation in two-dimensional image data with additive white Gaussian noise. We focus on the following two issues.

 i) The optimal detection threshold—i.e., the signal strength below which no method of detection can be successful for large dataset size n.

 ii) The optimal computational complexity of a near-optimal detector, i.e., the complexity required to detect signals slightly exceeding the detection threshold.

We describe a general approach to such problems which covers several classes of geometrically defined signals; for example, with one-dimensional data, signals having elevated mean on an interval, and, in d-dimensional data, signals with elevated mean on a rectangle, a ball, or an ellipsoid. In all these problems, we show that a naive or straightforward approach leads to detector thresholds and algorithms which are asymptotically far away from optimal. At the same time, a multiscale geometric analysis of these classes of objects allows us to derive asymptotically optimal detection thresholds and fast algorithms for near-optimal detectors.

Index Terms—Beamlets, detecting hot spots, detecting line segments, Hough transform, image processing, maxima of Gaussian processes, multiscale geometric analysis, Radon transform.

I. INTRODUCTION

SUPPOSE we have noisy image data $x(i,j), 1 \leq i, j \leq n$. We believe that hidden in the noise there is a line segment, of unknown location, orientation, and length. Applications include the following.

- *Detecting Moving Objects*. Suppose we have an infrared staring array, which looks in a fixed direction for long periods of time [2]. A distant moving object will create, upon lengthy exposure, an image of a very faint line segment against a noisy background. The length, orientation, and location of the segment are *a priori* unknown.

- *Detecting Ship Wakes*. The theory of bow waves says that a ship at sea leaves a wake in the shape of a "V," with an angle between the two wake arms of about 40°. In scanning aerial imagery for signs of ships, one expects that the imagery might sometimes contain a very faint "V" shape against a highly cluttered background; there is no real reason to expect any particular scale, orientation, or location of the "V" [22].

We envision that detecting line segments is a component task in more ambitious settings: detecting cracks in materials [54], identifying streams and roadbeds [5], [67], and so on. There is an extensive cognate literature on line detection in computer vision—extending back decades, for example, under the label "Hough transform"—for which citations in [26], [42], [61] may provide helpful pointers and applications references.

This paper derives asymptotic $(n \rightarrow \infty)$ properties of: i) the signal amplitude at which detection in noisy data is possible; and ii) the computational burden required to detect signals with amplitude approaching the limit of detectability. These are theoretical questions and are approached from an abstract viewpoint in this paper.

Our abstract theoretical approach has an advantage: it is not limited to detection of line segments in images; it can deal with a wide range of geometric detection problems, in one- and higher dimensional data. In such problems, it is desired to detect geometrically defined classes of objects (e.g., intervals in dimension 1, rectangles, disks, and ellipses in dimension $d > 1$) with near-optimal sensitivity, and to have computationally efficient algorithms for near-optimal detection in such settings.

In practice, detecting geometric objects is an important task in a variety of image-processing problems, including target recognition [18] and medical imaging. Detecting disks and tilted rectangles in very noisy data is currently a priority in cryo-electron microscopy [3], [36], [65], [70]. Our theory places fundamental limits on attainable performance, and provides a standard of comparison for practical work.

A. Line Segment Detection

Suppose we have an n-by-n array of pixels and a collection of $(n + 1)^2$ pixel corners. Let S denote a line segment connecting two pixel corners. Corresponding to this is an array $(\xi_S(i,j) : 1 \leq i, j \leq n)$ where $\xi_S(i,j)$ is proportional to the arc length of S incident on $\mathrm{Pixel}(i,j)$, but which has been standardized so that ξ_S has ℓ^2 norm 1. Let $\mathbb{S}_n$ be the collection of all such line segments.

Our observed data obey

$$x(i,j) = A \cdot \xi_S(i,j) + z(i,j), \qquad 1 \leq i, j \leq n \qquad (1.1)$$

E. Arias-Castro and D. L. Donoho are with the Department of Statistics, Stanford University, Stanford, CA 94305 USA (e-mail: acery@stat.stanford.edu; donoho@stat.stanford.edu).

X. Huo is with the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205 USA (e-mail: xiaoming@isye.gatech.edu).

where A is the signal strength, $z(i,j)$ is an independent and identically distributed (i.i.d.) Gaussian white noise, and S is an unknown line segment in $\mathbb{S}_n$. Our goal is to test $H_0 : A = 0$ against $H_{1,n} : A > 0, S \in \mathbb{S}_n$.

Two questions present themselves immediately.

1) *Threshold of detectability:* At what amplitude A does the object become reliably detectable?
2) *Computational complexity:* What is the minimal order of complexity of algorithms achieving the desired threshold of detectability?

We will be able to give satisfying answers to both kinds of questions, at least when performance is measured from the asymptotic viewpoint, as $n \to \infty$.

To motivate our approach, we recall some classical detection theory. Suppose that in (1.1), the segment S which is hypothesized to be present, instead of being an unknown member of the class $\mathbb{S}_n$, is actually one specific, *known* member of that class, and that A is known as well. Then, the alternative becomes a simple hypothesis $H_{1,S,A}$, say. The statistic

$$X[S] = \langle \xi_S, X \rangle = \sum_{i,j} \xi_S(i,j)x(i,j)$$

allows us to conduct the Neyman–Pearson test of H_0 against $H_{1,S,A}$, simply by asking if

$$X[S] > t$$

for some threshold t. In effect, this is (a repackaging) of the likelihood ratio test.

For the composite alternative hypothesis, where $A > 0$ and $S \in \mathbb{S}_n$ are both unknown, we might consider a test based on

$$X_n^* \equiv \max_{S \in \mathbb{S}_n} X[S] > t_{n,\alpha}^*$$

where $t_{n,\alpha}^*$ is an appropriate threshold calculated to preserve the overall α-level of the resulting test under H_0. It seems plausible that such a test would behave well, although the issue of determining the size of the threshold $t_{n,\alpha}^*$ would have to be solved. For obvious reasons, and also historical precedent in such cases, we call this the generalized likelihood ratio test (GLRT).

Our first main result states that this test is approximately optimal. A fundamental aspect of this problem is the existence of a well-defined asymptotic detection threshold, such that on one side of the threshold no test is of any value, and on the other side, powerful testing becomes possible. To articulate this, we introduce some terminology.

Definition 1.1: In a sequence of testing problems $(H_{0,n})$ versus $(H_{1,n})$, we say that a sequence of tests (T_n) is **asymptotically powerful** if

$$P_{H_{0,n}}\{T_n \text{ rejects } H_0\} + P_{H_{1,n}}\{T_n \text{ accepts } H_0\} \to 0$$

as $n \to \infty$, and the sequence is **asymptotically powerless** if

$$P_{H_{0,n}}\{T_n \text{ rejects } H_0\} + P_{H_{1,n}}\{T_n \text{ accepts } H_0\} \to 1$$

as $n \to \infty$.

Theorem 1.2: Under the null hypothesis H_0

$$X_n^* = 2\sqrt{\log(n)}(1 + o_p(1)), \qquad n \to \infty.$$

Hence, the critical value obeys

$$\frac{t_{n,\alpha}^*}{2\sqrt{\log(n)}} \to 1, \qquad n \to \infty.$$

Say that $\alpha_n \to 0$ **slowly enough** if

$$t_{n,\alpha_n}^* \sim 2\sqrt{\log(n)}, \qquad n \to \infty.$$

Consider, for a given η, the sequence of alternative hypotheses

$$H_{1,n}^\eta : x(i,j) = A_n \xi_S(i,j) + z(i,j), \qquad 1 \le i, j \le n$$

where

$$A_n = 2 \cdot (1 + \eta)\sqrt{\log(n)}(1 + o(1))$$

and S is an arbitrary element of $\mathbb{S}_n$. For $\eta > 0$, the GLRT operated with level $\alpha_n \to 0$ slowly enough is asymptotically powerful for testing $H_{0,n}$ against $H_{1,n}^\eta$. Alternatively, let $\eta < 0$; then every sequence of tests (T_n) is asymptotically powerless for testing $H_{0,n}$ against $H_{1,n}^\eta$.

In words, $2\sqrt{\log(n)}$ is roughly the optimal detection threshold; if the signal is a bit stronger than this (say, $A_n = 2.01\sqrt{\log(n)}$) it is asymptotically detectable by X_n^*, but anything slightly weaker (e.g., $A_n = 1.99\sqrt{\log(n)}$) is not asymptotically detectable by any test. (The result says nothing about performance right at the boundary—a delicate question for future work.)

Motivated by the above near-optimality result, one might hope to use X_n^* as a simple, easily understood tool. Unfortunately, it is not a very practical tool, in the following sense. Implementation in the obvious fashion costs $O(n^5)$ flops: there are $\approx n^4/2$ segments to be tested and, for most of these, $> n/10$ flops to compute the sum $\sum_{(i,j) \sim S} \xi_S(i,j)x(i,j)$. Fortunately, better asymptotic performance is possible.

Theorem 1.3: For each $\eta > 0$, there is an algorithm operating in $C(\eta) \cdot n^2 \log(n)$ flops which is asymptotically powerful for detecting signals with amplitudes $A_n \ge 2 \cdot (1 + \eta) \cdot \sqrt{\log(n)}$.

In this paper, then, we identify fundamental limits to detectability and algorithmically effective ways to approach those limits.

B. The Multiscale Viewpoint

The results stated above show that the space of line segments is not as it seems on the surface. The collection of all line segments between pixel corners has order $O(n^4)$ elements, but this crude count is misleading; it does not reflect the true number of "independent tests" or "underlying dimensionality" of the space of line segments. If it did, the correct detection threshold in this problem would be $\sim \sqrt{2\log(n^4)}$, not $\sim \sqrt{2\log(n^2)}$, and the correct computational complexity would be $O(n^5)$ flops, not $O(n^2 \log(n))$.

The fundamental point is that the effective dimensionality of the collection of all line segments in the square is $O(n^2 \log(n))$, not $O(n^4)$.

To understand this, we have developed tools that view the space of line segments from a multiscale viewpoint. Our approach has four components [1], [29].

1) *Beamlets*. A dictionary of multiscale line segments at a range of orientations, and at dyadically organized scales and locations. It has cardinality $O(n^2 \log(n))$.
2) *Beamlet chaining*. Arbitrary segments $S \in \mathbb{S}_n$ can be represented approximately as short chains of at most $O(\log(n))$ beamlets. Hence, the small cardinality set of beamlets can generate the $O(n^4)$ cardinality set of segments.
3) *Beamlet analysis*. The collection of all integrals of the image along line segments in the beamlet dictionary. Because of hierarchical relationships in the beamlet dictionary, this can be computed rapidly by a pyramid algorithm.
4) *Beamlet algorithms*. Using the preceding three ideas, we can develop fast algorithms to construct chains of beamlets and hence, optimize over line segments in far less than order $O(n^5)$ flops.

As will be seen, the $O(n^2 \log(n))$ nature of the beamlet dictionary is just as important for the analysis of the GLRT as it is for constructing fast algorithms to approximate it.

Multiscale approaches to detect linear features in two dimensions have been invoked for some time. Early examples include work of Brandt [16], [61], Horn and collaborators [4], [33]. Later we will discuss related literature and our relationship to it.

C. A Range of Geometric Problems

While our exposition began with the problem of detecting arbitrary line segments in two-dimensional images, there is a wide range of problems where geometrically defined classes of objects can be efficiently detected using multiscale methods. This paper aims to provide a coherent development of multiscale geometric detection, building up from simple examples to an abstract framework, and then deriving numerous corollaries of the abstract framework.

Section II begins the development by considering a one-dimensional case, where we have an array $x(i)$, $i = 0, \ldots, n-1$, which we believe to possibly have an elevated mean in some unknown interval $a \leq i < b$. We wish to detect the existence of such an interval with optimal sensitivity and computational efficiency. It turns out that this setting, although far simpler than many other geometric detection problems we will consider, exhibits all the basic theoretical elements in their simplest and cleanest form.

The collection of intervals has apparent dimension n^2; taking this at face value, it suggests that the optimal detection threshold should be $\sim \sqrt{2 \log(n^2)}$, the cost of computing sums over all of the intervals is apparently $O(n^3)$ flops. However, deeper study shows that the collection of intervals actually has effective dimension $O(n \log(n))$, the optimal detection threshold is $\sim \sqrt{2 \log(n)}$, and the complexity of an effective test has order $O(n)$ flops.

These conclusions are reached in Section II by introducing a multiscale analysis of the problem. The central idea for this multiscale analysis is to focus attention on dyadic intervals, a natural multiscale system of intervals, spanning a range of lengths and positions. A metric appropriate to the detection problem is introduced, and it is shown that, in this metric, one can approximate arbitrary intervals by short chains of dyadic intervals. Hence, search over the space of all intervals is replaced by constrained search over disjoint unions of a small number of dyadic intervals.

In effect, we learn that the space of intervals is roughly n dimensional because there are n dyadic intervals, that the detection threshold behaves roughly as the maximum of n independent Gaussians, and that fast algorithms need only examine a collection of $O(n)$ intervals to decide with statistical confidence whether one of the $O(n^2)$ intervals has an elevated mean.

The multiscale approach is quite general and applies in many geometric detection problems. The key ingredients from the interval case—dyadic intervals, short chains of dyadic intervals, search over short chains—will, under the guise of suitable analogs, reappear in other cases, and a series of theorems will follow, giving in each case lower detection thresholds and lower algorithmic complexity than a naive dimensionality argument would suggest.

Thus, in Section III, we consider the problem of detecting, in high-dimensional noisy data, arbitrary rectangles having elevated means. Once again, the difficulty of the problem is heavily overestimated by naive analysis, while a representation based on products of dyadic intervals reveals that the apparent dimensionality of the collection of all rectangles—$O(n^{2d})$—vastly exceeds the effective dimensionality, which is $O(n^d)$. However, rather than pedantically developing such results in Section III, we skip the proofs entirely, looking ahead to Section IV from which they follow effortlessly.

Our main result, stated in Section IV, considers an abstract framework, with a general class of geometric objects in d-dimensional noisy data, and places general assumptions about such a class. These assumptions permit a multiscale decomposition of the objects in the class and allow to calculate the asymptotic detection threshold and asymptotic computational complexity. The framework will be seen to be an abstraction and generalization of the ideas first developed in Section II. Section IV-C contains several tables summarizing how the abstract framework generates the key ideas and results of this paper, and summarizing the paper's results on optimal detection thresholds and optimal computational complexity.

We then apply this abstract theory in Sections V–VII, obtaining detection thresholds and efficient algorithms for detecting disks, tilted rectangles, and tilted ellipses, and for line segments. In Section VIII, we consider more complex settings, including detection of articulated objects and star-shaped "blobs" belonging to a nonparametric class.

For all these detection tasks performed on a n-by-n pixel image, the detection threshold is $\sim \sqrt{2 \log(n^2)}$. Our multiscale methodology provides a way to derive the asymptotic behavior of this detection threshold and, at least in principle, construct a near-optimal detector that runs in order $O(n^{2+\eta})$ flops, for all $\eta > 0$.

Conceptually, all these problems, although different in detail, share a common abstract structure, a common multiscale nature, and common sense in which the superficial dimensionality of the problem is far away from the effective dimensionality.

D. Related Work

1) Two Traditions: Our work may be taken as addressing problems in "computer vision"; more specifically "target detection" or "object recognition." An extensive literature has developed in those topic areas over the years, only a tiny fraction of which can be cited here. Some of that literature even uses cognate language: "detectors" for edges go back to classical work of the late 1970s, e.g., [55]; more recent literature is even said to develop "object detectors" which are multiscale and run rapidly [7], [35], [38], [66], [69].

At the same time, there is a very extensive literature on statistical image reconstruction from noisy data. Since the mid-1980s, work of Besag [11] and Geman and Geman [39], Bayesian methods have been widely used in reconstruction of images from noisy data. Based on a model for the object(s) to be estimated/identified in the image, together with a noise model, Bayesian methods rely on a posterior mode or posterior expectations to find a reconstructed object, often computed by Monte Carlo techniques [12], [39]. Deformable templates [8], [21], [34], [44], [45], [56], [58]–[60] and marked point processes [64], [68] are often used as models for deriving Bayesian methods.

2) Differences in Goals: The prior literature in these two fields, while fundamental and very useful, seems to pursue different goals than we do here. We consider a) *detection* of b) *geometric* objects in a c) *theoretical* setting. In our approach, optimality is carefully defined, as is the class of objects under consideration, and our efforts are oriented toward understanding how to analyze the problem theoretically and to determine the optimal detection threshold and the optimal computational complexity implied by a certain class of objects.

Our approach may be contrasted with the computer vision approach, which takes an engineering/practical viewpoint: What are useful algorithmic architectures for today's practical challenges? We take instead classical signal detection theory/mathematical decision theory as a starting point, and ask about intrinsic difficulty rather than about currently competitive algorithms.

Our approach may also be compared with the Bayesian image reconstruction literature, which is inspired by an aesthetic/philosophical viewpoint: Are there models for images under which optimal inferences would have especially beautiful and/or convenient forms? Bayesian methods are often not fast in terms of flops, but they are often optimal—within a very narrowly specified decision-theoretic model. However, Bayesians often do not study the detection problem (as opposed to the reconstruction problem) and they often do not study theoretical properties of models (e.g., size of detection thresholds). Finally, general geometric classes are typically not very convenient for specifying Bayesian prior distributions—geometric models do not often support aesthetically pleasing Bayesian priors as least favorable distributions. Bayesians typically prefer more special stochastic models on grounds of beauty or computational tractability. Such

alternative models would not lead to theoretically interesting methods in our framework.

Our approach therefore diverges from these important and very well-established traditions, asking different questions—about geometric objects—seeking different kinds of answers—more theoretical ones—and using different kinds of tools—asymptotic analysis, multiscale decomposition of objects of interest.

3) Notable Recent Work in the Two Traditions: The most closely related methodological literature seems to be Geman and Jedynak [38], Fleuret and Geman [35], and Viola and Jones [69], all of which in some way detect using multiscale features and fast algorithms. Our work indirectly lends political support to such methodological approaches.

Within the Bayesian literature, the marked point process models of van Lieshout and of Stoica [64], [68] and the deformable template models of Rue and coauthors [58]–[60] seem to provide the closest approach to our geometric detection problem. Unfortunately, reliance on Monte Carlo Markov chain methods for Bayesian inference seems to prevent any near-optimality from the computational complexity viewpoint. Perhaps a multiscale geometric analysis approach could speed up such methods, as Brandt [15] has often argued for methods inspired by statistical physics. Moreover, the detection thresholds and other theoretical properties implied by such methods seem not to have been studied so far.

4) Intellectual Precursors: The intellectually closest predecessors of our approach can be found, first, in line segment detection.

- Our algorithmic viewpoint on line segment detection (beamlets) has been anticipated by work of Brandt and Dym [16] and Horn and coworkers [4], [33].
- Our decision-theoretic viewpoint for line segment detection was anticipated by Desolneux, Moisan, and Morel [26] who considered a Poisson/binomial stochastic model for noise instead of our Gaussian one.

Second, there are relations with work in the mathematical sciences. In sequential analysis, there has been interest for some time in probability distributions for quantities of the form

$$W_n = \max\left\{\sum_s^t z(u)/\sqrt{t-s} : 1 \le s < t \le n\right\}$$

where $z(u)$ is a white Gaussian noise. See, for example, work of Siegmund and Venkatraman [63] and Lai and Chan [19], [20]. This problem is closely related to our Section II, which involves detecting a one-dimensional signal with an elevated mean. The techniques developed in sequential analysis, based on Markov processes, seem quite different than those used here; in particular, our multiscale approach allows to formulate a fast algorithm and also an abstract strategy that works for much more general classes of sets than intervals.

Farther afield, but sure to come up in a close examination of our topic, is work in genome sequence analysis. Waterman and collaborators [9], [23], [24] studied asymptotic behavior of detectors searching for squares and/or rectangles of specified values in discrete-valued data. They define matching not in terms of statistical significance (as we do), but by agreement

within a specified proportion of errors. More generally, the scan statistic is a widely used tool in statistical sequence analysis; see the books [40], [41]. It measures the maximal number of occurrences in any interval of a given length. The statistic we are using is a multiscale scan statistic, which considers all possible lengths and positions and compares different lengths by putting all comparisons on a common statistical scale.

E. Contributions

It may help the reader to briefly summarize the main contributions we see in our work.

- *Geometric Detection Framework*. We formally state, and give an asymptotically optimal solution to, the problem of detecting an object buried in noisy data, when the object comes from one of several classes of geometric objects.
- *The Metric δ*. This metric between sets, defined in Section II, drives our theoretical analysis. Unlike the Hausdorff metric, δ measures intrinsic statistical distinguishability.
- *Multiscale Geometric Analysis*. We show that one can approximate elements in a geometric class, in δ metric, by a disjoint union of a small number of multiscale generating elements. Thus, dyadic intervals combine to create arbitrary intervals, beamlets combine to approximate arbitrary line segments, and axoids combine to create ellipsoids, and, more generally, convex shapes.
- *Approximate GLRT*. Suppose that one can approximate (in δ metric) a general class of geometric objects using disjoint unions from a simpler class of "generators." We show that a simple search procedure, exhaustively exploring disjoint unions of small numbers of generators, gives a near-optimal detector.
- *Cardinality of Generators*. The multiscale geometric analysis provides sets of generators for the classes of interest which are small, with not more than $O(n^d \log(n))$ elements in dimension d. Thus, at most $O(n^2 \log(n))$ dyadic rectangles are needed as generators for the class of disks. This, and some extra work, enables search over combinations of few generators to use effectively no more than $O(n^d \log(n))$ flops.

These basic ideas underly a general theory encompassing numerous classes of geometric objects, giving precise orders of computational complexity and precise asymptotic detection thresholds for those classes.

From our theory, we draw several surprising conclusions:

- *Apparent Versus Effective Dimension*. In all the cases we consider, the apparent dimension of the object class is twice as big as the effective dimension of the class. This means that the optimal thresholds are a factor $1/\sqrt{2}$ smaller than the thresholds one might expect based on apparent dimension, and that the computational complexity is roughly the square root of the complexity anticipated on the grounds of apparent dimensionality. Doing things optimally is *much* better than doing the "obvious" thing.
- *Invariance of Optimal Detection Threshold*. In the cases we consider, the detection threshold takes the form $\sim \sqrt{2 \log(n^d)}$ where d is the dimension of the ambient

space. This is true for very simple parametric object classes, like rectangles, and more complex nonparametric classes, like the class of convex sets with regular boundaries. Presumably, the difference between classes is buried inside second-order asymptotic terms not visible to us at this level of asymptotic analysis. The rough invariance of the detection threshold was not an expected outcome of this study; it would have seemed more in line with years of research in nonparametric estimation to have the results depend on assumed smoothness. Perhaps the cause for this difference is the fact that we discuss detection rather than estimation.

- *Invariance of Optimal Computational Complexity*. In the cases we consider, the computational complexity is $O(n^{d+\eta})$ for each $\eta > 0$, independently of the class itself, so that it is roughly the same for simple classes like rectangles and for complex classes like convex sets. This again was not an expected outcome of the study.

F. Practical Implications

Using our theory, it becomes possible to ask whether any of the standard methods of "object detection" are either computationally or statistically efficient. How do their detection thresholds and computational burdens compare with the theoretical limits we derive? Although the question of sensitive object detection is central for target recognition and face detection, we do not know of any work evaluating the empirical detection thresholds of actual detectors. Our theory provides a starting point for such efforts, suggesting a specific standard for comparison. In the long run, this could have an impact in sorting out the many possible competing methods, focusing attention on the best ones; it might offer computer vision and image processing researchers a common theoretical evaluation framework.

One suspects that our framework, with effort, could provide theoretical support to some of the existing multiscale approaches, such as Fleuret and Geman [35] and Viola and Jones [69]. A thorough analysis might show that, when deployed appropriately, those techniques could be asymptotically efficient both in the statistical sense and in the computational complexity sense. This seems a fruitful question for further work.

II. DETECTING INTERVALS IN DIMENSION ONE

Let $X = (x(i) : 0 \le i < n)$ be an array of random variables which contains a white noise, except possibly on an interval where the mean might be elevated

$$x(i) = \mu 1_{\{a \le i < b\}} + z(i), \qquad i = 0, \ldots, n-1.$$

Here, the endpoints a, b of the interval obey $0 \le a < b \le n$ but are assumed to be unknown *a priori*; and μ is the amplitude of the signal. This is a clear one-dimensional analog of the two-dimensional case studied later in this paper; in this section, we will state and prove one-dimensional analogs of Theorems 1.2 and 1.3. It turns out to be most natural to work with the *normalized amplitude* $A = \mu/\sqrt{b-a}$.

In this setting, the analog of GLRT is the following. Let

$$\xi_{a,b}(i) = 1_{\{a \le i < b\}}/\sqrt{b-a}$$

be the normalized prototype of an interval, and let

$$X_n^* = \max_{0 \le a < b \le n} \langle \xi_{a,b}, X \rangle.$$

We will be interested in four main issues about X_n^*. First, how large is the detector threshold? Second, how small an amplitude A is reliably detectable? Third, how effective is this test? Fourth, how can this be calculated rapidly? We capture the answers in four formal results.

Theorem 2.1: For each $\eta > 0$

$$P_{H_0} \left\{ X_n^* > \sqrt{2(1+\eta)\log(n)} \right\} \to 0, \qquad n \to \infty.$$

Theorem 2.2: Let $t_{n,\alpha}^1$ denote the $1 - \alpha$ quantile of X_n^*. Let $\alpha_n \to 0$ slowly enough such that

$$t_{n,\alpha_n}^1 \backsim \sqrt{2\log(n)}, \qquad n \to \infty.$$

Let $A_n = \sqrt{2(1+\eta)\log(n)}$ be a sequence of signal amplitudes, with $\eta > 0$. Then the GLRT operated at level α_n is asymptotically powerful.

In words: the GLRT reliably detects the presence of an interval of strength $\sqrt{2.01 \log(n)}$.

Theorem 2.3: Let $A_n = \sqrt{2(1-\eta)\log(n)}$ be a sequence of signal amplitudes, with $\eta > 0$. Then there is a sequence of distributions on intervals $[a, b) \subset \{0, \dots n-1\}$ so that every sequence of tests (T_n) is asymptotically powerless.

In words: no test reliably detects the presence of an interval of normalized amplitude $\sqrt{1.99 \log(n)}$.

Theorem 2.4: For each $\eta > 0$, there is an algorithm running in $O(n)$ time with the ability to detect intervals of strength $\sqrt{2(1+\eta)\log(n)}$.

In short, we can efficiently and reliably detect intervals of amplitude roughly $\sqrt{2\log(n)}$, but not smaller. We note that Theorems 2.1 and 2.2 can also be obtained by other approaches, as shown in [20].

A. Viewpoint

Let $\mathbb{I}_n$ denote the collection of all subintervals of $\{0, \dots, n-1\}$. We are studying a random field $X[I]$ defined on $I \in \mathbb{I}_n$; the statistic X_n^* is the maximum of X over this collection. Our approach singles out for attention the dyadic intervals as a special subset of the collection of all intervals. Dyadics are special because they have cardinality $2n$ rather than $\approx n^2/2$ and yet furnish an ϵ-net for the space of intervals in a special "detection metric" we define below, with $\epsilon < 1$. Thus, although the dyadics are not actually close to every interval, they are "within shouting distance"; this shouting distance property will be the main one determining the effective dimension of the space of intervals.

B. The Metric Space of Intervals

Let n be a dyadic integer $n = 2^J$ and let $\mathbb{J}_n$ denote the collection of all dyadic subintervals

$$I_{j,k} = \{k2^j, \dots (k+1)2^j - 1\}$$

where $0 \le j \le \log_2(n)$ and $0 \le k < n/2^j$. We will let $|I|$ denote the cardinality of an interval I.

Any interval is naturally associated with its maximal dyadic subinterval. The example of the interval $\{1, \dots, n-2\}$, which has $\{n/2, \dots, 3n/4 - 1\}$ for (one choice of) maximal dyadic interval, shows that the dyadic subinterval can be roughly one-fourth as long as the the whole interval. In fact, this is extremal, and we have

$$\min_I \max_{I_{j,k} \subset I} \frac{|I_{j,k}|}{|I|} = \frac{n/4}{n-1} \sim \frac{1}{4}. \qquad (2.1)$$

Define now a measure of *affinity* between intervals

$$\rho(I, J) = \frac{|I \cap J|}{\sqrt{|I||J|}}.$$

This is precisely the correlation coefficient between the functions 1_I and 1_J. Hence, $0 \le \rho \le 1$, with $\rho = 0$ indicating disjointness and $\rho = 1$ indicating complete identity. We can also define a metric by

$$\delta(I, J) = \sqrt{1 - \rho(I, J)}$$

ranging from 0 for identical intervals, to 1 for disjoint ones. (The relevance of these measures to our setting is made clear later in Lemma 2.7). From (2.1) we have

$$\min_I \max_{I_{j,k} \subset I} \rho(I, I_{j,k}) > 1/2$$

and

$$\max_I \min_{I_{j,k} \subset I} \delta(I, I_{j,k}) < \sqrt{1/2}.$$

This proves the following.

Lemma 2.5: The dyadic intervals $\mathbb{J}_n$ make an ϵ-net for the collection $\mathbb{I}_n$ of all intervals, with $\epsilon = \sqrt{1/2}$.

While, of course, $\epsilon = \sqrt{1/2}$ is not small, the fact that any $\epsilon < 1$ is possible, independent of n, should attract our attention. To get small-ϵ approximations, we use dyadic intervals as our "base" and form compound intervals by attaching additional dyadic intervals at the ends. Formally, we say that the interval J_ℓ is an *ℓ-level extension* if it can be constructed as follows.

1) Start from a base J_0 which is either a dyadic interval $I_{j,k}$ or the union of two adjacent dyadic intervals $I_{j,k}$ and $I_{j,k+1}$, where k is odd (so that $I_{j,k}$ and $I_{j,k+1}$ are not siblings).
2) At stages $g = 1, \dots, \ell$, extend J_{g-1} to produce J_g by attaching dyadic intervals of length $2^{-g}|I_{j,k}|$ at either, or both, ends of J_{g-1} or by doing nothing (so that $J_g = J_{g-1}$).

The result will be an interval as depicted in Fig. 1. The collection of all ℓ-level extensions of a dyadic interval I will be denoted $\mathbb{J}_\ell[I]$; the collection of all ℓ-level extensions will be denoted $\mathbb{J}_{n,\ell}$. The next result is proved in Appendix I.

Lemma 2.6:

$$\#\mathbb{J}_{n,\ell} \le n4^{\ell+1} \qquad (2.2)$$

$$\rho_{n,\ell}^* = \min_{I \in \mathbb{I}_n} \max_{J \in \mathbb{J}_{n,\ell}} \rho(I, J) \ge 1/\sqrt{1 + 2 \cdot 2^{-\ell}}. \quad (2.3)$$
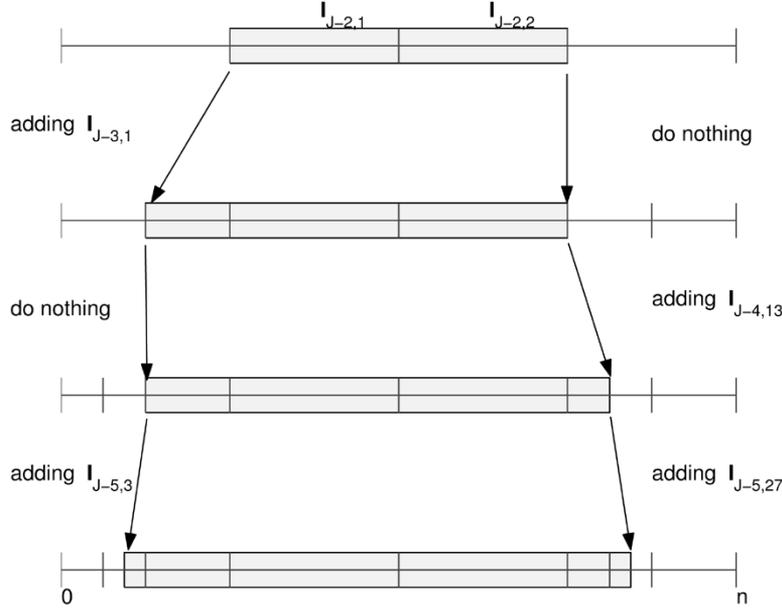
Fig. 1. Extending dyadic intervals. Starting from a dyadic base, dyadic intervals of shorter length are appended (or not) in a stagewise fashion.

It follows that, for each $I \in \mathbb{I}_n$, there is an interval in $\mathbb{J}_{n,\ell}$ with

$$\delta(I, J) \le 2^{-\ell/2}. \tag{2.4}$$

C. Maxima of Gaussian Processes on $\mathbb{I}_n$

Let now $\{z_i : i = 0, \ldots, n-1\}$ be a collection of i.i.d. $\mathcal{N}(0,1)$ random variable, and, for $I \in \mathbb{I}_n$, set

$$Z[I] = |I|^{-1/2} \sum_{i \in I} z_i. \tag{2.5}$$

Then for each fixed I, $Z[I]$ is $\mathcal{N}(0,1)$. We will view Z as a set-indexed random field, indexed by elements in $\mathbb{I}_n$, $\mathbb{J}_n$, or $\mathbb{J}_{n,\ell}$. We are interested in two maximal quantities

$$Z^*[\mathbb{I}_n] = \max\{Z[I] : I \in \mathbb{I}_n\}$$
$$Z^*[\mathbb{J}_{n,\ell}] = \max\{Z[I] : I \in \mathbb{J}_{n,\ell}\}$$

and in showing that these behave similarly. As every $I \in \mathbb{I}_n$ has an approximation by $\mathbb{J}_{n,\ell}$ with δ distance $\le 2^{-\ell/2}$, we can write

$$Z^*[\mathbb{J}_{n,\ell}] \le Z^*[\mathbb{I}_n] \le Z^*[\mathbb{J}_{n,\ell}] + \Delta_{n,\ell} \tag{2.6}$$

where

$$\Delta_{n,\ell} = \max\{Z[I] - Z[J] : I, J \in \mathbb{I}_n; \delta(I, J) \le \epsilon_\ell\}$$

and $\epsilon_\ell = 2^{-\ell/2}$. We now justify our interest in δ-distance, showing that it is intrinsically connected with the size of oscillations $Z[I] - Z[J]$.

Lemma 2.7:

$$\operatorname{var}(Z[I] - Z[J]) = 2\delta^2(I, J) \tag{2.7}$$
$$\operatorname{cov}(Z[I], Z[J]) = \rho(I, J). \tag{2.8}$$

Proof: In view of $\delta^2 = 1 - \rho$, the second relation implies the first. Now

$$\operatorname{cov}\left(\sum_{i \in I} Z_i, \sum_{j \in J} Z_j\right) = \operatorname{var}\left(\sum_{i \in I \cap J} Z_i\right) = |I \cap J|$$

and the second relation follows from the definition (2.5) of $Z[I]$ and $Z[J]$. $\square$

Thus, if two fixed intervals I and J obey $\delta(I, J) \le \epsilon_\ell$, we expect $Z[I] - Z[J]$ to be of size roughly $\sqrt{2}\epsilon_l$. The following lemma shows that, even maximizing over all I and J obeying $\delta(I, J) \le \epsilon_\ell$, the largest $Z[I] - Z[J]$ we will observe is at most a logarithmic factor times ϵ_ℓ.

Lemma 2.8:

$$P\left(\delta_{n,\ell} > 4\sqrt{\log(n)} \cdot 2^{-\ell/2}\right) \le 1/\sqrt{\log(n)}. \tag{2.9}$$

This, together with (2.6), will allow us to show that $Z^*[\mathbb{I}_n]$ and $Z^*[\mathbb{J}_{n,\ell}]$ behave similarly. To prove (2.9), we need the following well-known fact about extreme values of Gaussian processes ([52], [57], [62]).

Lemma 2.9: Let $w_1, \ldots, w_m$ be (possibly dependent) $\mathcal{N}(0, \sigma_i^2)$ variables with all $\sigma_i \le \sigma$. Then

$$P\left(\max\{w_1, \ldots, w_m\} > \sqrt{2\log(m)} \cdot \sigma\right) \le 1/\sqrt{4\pi \log(m)}. \tag{2.10}$$

Applying this to $\Delta_{n,\ell}$, set

$$m = \#\{I - J : I, J \in \mathbb{I}_n\},$$
$$\sigma^2 = \max\{\operatorname{var}(Z[I] - Z[J]) : I, J \in \mathbb{I}_n; \delta^2(I, J) \le \epsilon_\ell^2\} = 2\epsilon_\ell^2.$$

As $m \le (n^2)^2$, we get from (2.10) that for $n \ge 8$

$$P\left\{\Delta_{n,\ell} > \sqrt{2\log(n^4)}\sqrt{2}\epsilon_\ell\right\} \le 1/\sqrt{16\pi \log(n)}$$

which yields (2.9).

D. Behavior of Maxima Under H_0

We now demonstrate Theorem 2.1. Note that, under H_0, $X_n^* =_D Z^*[\mathbb{I}_n]$. Hence, we need to show that, for $\eta > 0$

$$P\left(Z^*[\mathbb{I}_n] > \sqrt{2(1+\eta)\log(n)}\right) \to 0, \qquad n \to \infty. \quad (2.11)$$

To do this, we prove a similar result for $Z^*[\mathsf{J}_{n,\ell}]$ and use (2.6) and Lemma 2.8. First apply Lemma 2.9 to $Z^*[\mathsf{J}_{n,\ell}]$. With $\sigma = 1$ and $m = \#\{\mathsf{J}_{n,\ell}\} \le n4^{l+1}$, we get

$$P\left(Z^*[\mathsf{J}_{n,\ell}] > \sqrt{2\log(n) + (2\ell+2)\log(4)}\right) \le 1/\sqrt{\log(n)}.$$

Now pick $\ell = \ell_n \to \infty$ as $n \to \infty$, with

$$\ell_n/\log(n) \to 0, \qquad n \to \infty$$

so that

$$\sqrt{2\log(n) + (2\ell_n+2)\log(4)} = \sqrt{2\log(n)} \cdot (1 + o(1)),$$
$$n \to \infty.$$

From this and (2.9) we have that $\Delta_{n,\ell_n} = o(\sqrt{\log(n)})$, so that with overwhelming probability, for large n

$$\sqrt{2\log(n) + (2\ell_n+2)\log(4)} + \Delta_{n,\ell_n} < \sqrt{2(1+\eta)\log(n)}$$

and so (2.11) follows. □

E. Proof of Theorem 2.2

We now suppose that $A_n \ge \sqrt{2(1+\eta)\log(n)}$, and that

$$x_n(i) = A_n|\tilde{I}_n|^{-1/2}1_{\{i \in \tilde{I}_n\}} + z(i) \quad (2.12)$$

for some interval $\tilde{I}_n$. We will show that X_n^* rejects $H_{0,n}$ with overwhelming probability. Now obviously

$$X^*[\mathbb{I}_n] \ge X[\tilde{I}_n].$$

But

$$X[\tilde{I}_n] \frown \mathcal{N}(A_n, 1).$$

Moreover, Theorem 2.1 implies that, for all sufficiently large n

$$t_{n,\alpha_n}^* < \sqrt{2(1+\eta/2)\log(n)}$$

however,

$$P\left(X[\tilde{I}_n] < \sqrt{2(1+\eta/2)\log(n)}\right)$$
$$= P\left(\mathcal{N}(0,1) < -\gamma\sqrt{\log(n)}\right), \qquad n \to \infty$$

where $\gamma = \sqrt{2(1+\eta)} - \sqrt{2(1+\eta/2)} > 0$. Now, applying Mills' ratio

$$P\left(\mathcal{N}(0,1) < -\gamma\sqrt{\log(n)}\right) \le 2n^{-\gamma^2/2} \to 0, \qquad n \to \infty.$$

It follows that

$$P_{H_1}\left(X^*[\mathbb{I}_n] \le t_{n,\alpha_n}^*\right) \to 0, \quad n \to \infty. \qquad \square$$

F. Proof of Theorem 2.3

Consider the collection $\mathcal{P}_n$ of pairs $\{2k, 2k+1\}$. Then $\mathcal{P}_n \subset \mathbb{I}_n$, while the elements of $\mathcal{P}_n$ have disjoint support and so are orthogonal when viewed as vectors in R^n. Let $H_{0,n}$ be the hypothesis $X_i \overset{\text{i.i.d.}}{\frown} \mathcal{N}(0,1)$ and $H'_{1,n}$ be the hypothesis

$$x(i) = A_n|I|^{-1/2}1_I(i) + z(i), \qquad i = 0, \ldots, n-1$$

where I is chosen uniformly at random from $\mathcal{P}_n$, and $A_n = \sqrt{2(1-\eta)\log(n)}$.

Reduce our data by summing adjacent pairs, producing

$$w_i = (x(2i) + x(2i+1))/\sqrt{2} = \theta_i + z'_i, \quad i = 0, \ldots, n/2-1$$

where $\theta_i = 0$ in all but one randomly chosen position and $\theta_i = A_n$ in that position, and z'_i are i.i.d. $\mathcal{N}(0,1)$. By sufficiency, our ability to discriminate on the basis of $(x(i))$ is the same as on the basis of the $(w(i))$.

The problem of testing whether a sequence otherwise thought to be Gaussian white noise might have a single nonzero mean in an unknown position may be called the "needle in a haystack" problem, and it has been known, according to [17], since work of Ibragimov and Khas'minskii [47], [48]. For $\eta > 0$ and $A_n = \sqrt{2(1-\eta)\log(n)}$, hypotheses $H_{0,n}$ and $H'_{1,n}$ merge asymptotically, meaning that every sequence of tests has Type I + Type II errors tending to 1. See also related work of Ingster [49], [50], of Donoho and Johnstone [31], and of Jin [30], which elaborates this fact in various ways.

Now the hypothesis $H'_{1,n}$ is a subhypothesis of the $H_{1,n}$ laid out in the statement of Theorem 2.3, so the problem of distinguishing $H_{0,n}$ from $H_{1,n}$ must be at least as hard. Theorem 2.3 follows. □

G. Proof of Theorem 2.4

We propose an algorithm based on extension of promising dyadic intervals. Fix $\eta > 0$. Recalling the definition (2.3) of $\rho_{n,\ell}^*$, fix ℓ so that

$$\rho_{n,\ell}^* > \sqrt{\frac{1+\eta/2}{1+\eta}}. \quad (2.13)$$

Proceed in the following three stages.

1) *Find promising dyadic intervals*. Identify all dyadic intervals $I \in \mathbb{I}_n$ with

   $$X[I] > 1/3\sqrt{\log(n)}. \quad (2.14)$$

   If there are more than $n^{19/20}$ such intervals among the $2n$ dyadic ones, reject $H_{0,n}$.
2) *Extend promising intervals*. For each interval I found at Stage 1, enumerate all level-ℓ extensions

   $$X_\ell^*[I] = \max\{X[I'] : I' \in \mathsf{J}_\ell[I]\}.$$

3) *Decide*. If the maximum $X[I]$ at Stage 1 exceeds $\sqrt{2\log(n)}$ or if the maximum $X_\ell^*[I]$ at Stage 2 exceeds $\sqrt{2(1+\eta/3)\log(n)}$, reject $H_{0,n}$.

By design, the algorithm is computationally efficient. Nominally, the cost of evaluating X_n^* by brute force is of order $O(n^2)$ flops, whereas our procedure can be seen to take order $O(n)$ flops. The main algorithmic work is at Stage 1, where $X[I]$ must be evaluated for every dyadic interval. This requires only $O(n)$ flops. Although there are $2n$ such intervals, and hence apparently $O(n^2)$ work (as many intervals are of length $n/10$ or larger), the full array of dyadic sums can actually be obtained in order $O(n)$ flops. To see this, notice that dyadic sums obey a recursion

$$S[I_{j,k}] = S[I_{j-1,2k}] + S[I_{j-1,2k+1}],$$
$$1 \le j \le \log_2(n), \ 0 \le k < n/2^j$$

so one is able to reuse sums at finer levels to compute the sums at coarser levels. The recursion starts at the finest level by

$$S[I_{0,k}] = x(k), \qquad k = 0, \ldots, n-1.$$

Once the $2n$ dyadic sums are available, one obtains the detector statistics by rescaling

$$X[I_{j,k}] = 2^{-j/2} S[I_{j,k}].$$

At most $n^{19/20}$ promising intervals will be harvested in the original thresholding (2.14). Then one proceeds to Stage 2. Each interval processed at Stage 2 generates $O(4^l)$ work; the work bound at this stage is $O(4^\ell n^{19/20})$; as ℓ is constant, this is $o(n)$.

We now show that this algorithm is able to detect line segments of amplitude

$$A_n = \sqrt{2(1+\eta)\log(n)}.$$

Let $\tilde{I}_n$ denote the underlying line segment generating the specific data (x_i). Let $I^{(d)}$ denote its maximal included dyadic subinterval. Let $\tilde{I}_{n,\ell}$ denote the ℓ-level extension of $I^{(d)}$ which best approximates $\tilde{I}_n$ in δ-distance, so that, by (2.4)

$$\delta\left(\tilde{I}_n, \tilde{I}_{n,\ell}\right) \le 2^{-\ell/2}.$$

Now ℓ has been chosen so that

$$\rho\left(\tilde{I}_n, \tilde{I}_{n,\ell}\right)\sqrt{1+\eta} > \sqrt{1+\eta/2}.$$

It follows from Lemma 2.5 that

$$EX[I^{(d)}] = \left[A_n \xi_{\tilde{I}_n}, \xi_{I^{(d)}}\right] = A_n \rho\left(\tilde{I}_n, I^{(d)}\right) > 1/2 \, A_n$$

and from Lemma 2.6, we have

$$EX[\tilde{I}_{n,\ell}] = \left[A_n \xi_{\tilde{I}_n}, \xi_{\tilde{I}_{n,\ell}}\right] = A_n \rho\left(\tilde{I}_n, \tilde{I}_{n,\ell}\right)$$
$$> \sqrt{\frac{1+\eta/2}{1+\eta}} A_n.$$

Of course, both $X\left[I^{(d)}\right]$ and $X\left[\tilde{I}_{n,\ell}\right]$ have variance 1.

Now the algorithm succeeds in detecting H_n^1 if both

$$X\left[I^{(d)}\right] > 1/3\sqrt{\log(n)}$$

and

$$X\left[\tilde{I}_{n,\ell}\right] > \sqrt{2(1+\eta/3)\log(n)};$$

in that case, both Stages 1 and 2 function as they were designed: $I^{(d)}$ indeed shows up as a promising interval, and its extension $\tilde{I}_{n,\ell}$ generates a rejection of $H_{0,n}$. Now for sufficiently large n

$$P\left(X\left[I^{(d)}\right] < 1/3\sqrt{\log(n)}\right)$$
$$\le P\left(\mathcal{N}(A_n/2, 1) \le 1/3\sqrt{\log(n)}\right)$$
$$= P\left(\mathcal{N}(0,1) \le -\gamma\sqrt{\log(n)}\right)$$
$$\le n^{-\gamma^2/2} \to 0, \qquad n \to \infty$$

where $\gamma = \sqrt{2}/2 - 1/3 > 0$. Similarly

$$P\left(X\left[\tilde{I}_{n,\ell}\right] \le \sqrt{2(1+\eta/3)\log(n)}\right)$$
$$\le P\left(\mathcal{N}\left(\sqrt{2(1+\eta/2)\log(n)}, 1\right) \le \sqrt{2(1+\eta/3)\log(n)}\right)$$
$$= P\left(\mathcal{N}(0,1) \le -\gamma\sqrt{\log(n)}\right) \le n^{-\gamma^2/2} \to 0, \quad n \to \infty$$

where

$$\gamma = \sqrt{2}\left(\sqrt{1+\eta/2} - \sqrt{1+\eta/3}\right) > 0.$$

The two "failure modes" for the detector thus are seen to occur each with vanishingly small probability under $H_{1,n}$.

We now verify that under $H_{0,n}$, the procedure successfully accepts H_0 with overwhelming probability. There are three ways to reject H_0.

1) The number of dyadic intervals I where

$$X[I] > 1/3\sqrt{\log(n)}$$

   exceeds $n^{19/20}$; or

2) there is a dyadic interval I where

$$X[I] > \sqrt{2\log(n)};$$

   or

3) there is a level-ℓ extension I where

$$X[I] > \sqrt{2(1+\eta/3)\log(n)}.$$

We will see that the three corresponding "failure modes" become increasingly unlikely under H_0. Indeed, mode 1 has an asymptotically negligible probability

$$P_{H_0}\left(X[I] > 1/3\sqrt{\log(n)}\right) \le n^{-1/18}$$

so letting $S_n = \#\left\{I \in \mathbb{J}_n : X[I] > 1/3\sqrt{\log(n)}\right\}$

$$P_{H_{0,n}}\left(S_n \ge n^{19/20}\right) \le \frac{ES_n}{n^{19/20}}$$
$$\le \frac{2n \cdot n^{-1/18}}{n^{19/20}}$$
$$= 2n^{-1/180} \to 0, \qquad n \to \infty.$$

Mode 2 also has an asymptotically negligible probability: by applying Lemma 2.9 with $\sigma = 1$ and $m = \#\mathbb{J}_n = 2n$

$$P_{H_{0,n}}\left\{X^*[\mathbb{J}_n] > \sqrt{2\log(2n)}\right\} \le 1/\sqrt{4\pi\log(2n)} \to 0,$$
$$n \to \infty.$$

Mode 3 also has an asymptotically negligible probability. In fact

$$X^*[\mathbb{J}_{n,\ell}] \le X^*[\mathbb{l}_n]$$

while, by Theorem 2.1

$$P_{H_{0,n}}\left\{X^*[\mathbb{l}_n] \ge \sqrt{2(1+\eta/4)\log(n)}\right\} \to 0, \qquad n \to \infty.$$
$$\square$$

Note: The algorithm just discussed has a parameter η which must be supplied. A small refinement of the algorithm does away with this parameter, and allows to detect any amplitude

sequence $A_n \geq \sqrt{2(1+\eta)\log(n)}$ for any $\eta > 0$, while maintaining the $O(n)$ flops efficiency. In the algorithm, rather than fixing η and then ℓ based on our criterion (2.13), simply specify a sequence $\ell_n \to \infty$ in such a way that $4^{\ell_n} = o(n^{1/20})$. It then happens that the total work for the procedure remains $O(n)$ while the sensitivity to $H_{1,n}$ is asymptotically at least as good as any version of the unmodified algorithm.

III. DETECTING AXIS-ALIGNED RECTANGLES IN DIMENSION d

Suppose now that we have noisy data $X = \{x(i_1, \ldots, i_d)\}$ observed on a d-dimensional Cartesian grid, $d > 1$, and that the data contain a Gaussian white noise except that the mean may be nonzero throughout some unspecified d-dimensional axis-aligned rectangle $R = I_1 \times \cdots \times I_d$. More specifically, our data are of the form

$$x(i_1, \ldots, i_d) = A \cdot \xi_R(i_1, \ldots, i_d) + z(i_1, \ldots, i_d),$$
$$1 \leq i_1, \ldots, i_d \leq n$$

where $z(i_1, \ldots, i_d)$ is a Gaussian white noise, where the signal ξ_R is the ℓ^2 normalized indicator of rectangle R, and where A is the signal amplitude. Letting $\mathbb{I}_n^d$ denote the set of d-dimensional rectangles, we are interested in testing $H_{0,n} : A = 0$ versus $H_{1,n} : A > 0$, $R \in \mathbb{I}_n^d$. This is related to problems of finding "high activity regions" [6], [46], and multivariate "bumps" [37].

If a specific rectangle R is the only one under consideration, we would be interested in the simple hypothesis-testing problem $H_{0,n} : A = 0$ versus $H_{1,R,A}$, where the particular R is specified. The Neyman–Pearson test could be based on

$$X[R] = \langle \xi_R, X \rangle = \sum \xi_R(i_1, \ldots, i_d) x(i_1, \ldots, i_d).$$

With A and R unspecified, we consider instead the GLRT statistic

$$X_{n,d}^* = \max_{R \in \mathbb{I}_n^d} X[R].$$

In this setting, we have analogs of earlier results for intervals. To avoid tedious repetition, we formalize some terminology.

Definition 3.1: Considering families of detection problems $H_{0,n}$ versus H_{1,n,A_n}, with amplitude parameter A_n, we say that **the optimal detection threshold** $\sim \tau_n$ if, for each fixed $\eta > 0$, along a sequence of testing problems with $A_n \geq (1+\eta)\tau_n$, there are asymptotically powerful tests, while along a sequence of testing problems with $A_n \leq (1-\eta)\tau_n$, every sequence of tests is asymptotically powerless. We say that a specific sequence of tests is **near optimal** if, when applied to any sequence of testing problems with $A_n \geq (1+\eta)\tau_n$, it is asymptotically powerful, for every $\eta > 0$.

In these definitions we omit consideration of what happens exactly at the threshold $A_n = \tau_n(1 + o(1))$, where the story may be considerably more complicated.

Theorem 3.2: In the problem of detecting d-dimensional axis-aligned rectangles, the optimal detection threshold $\sim \sqrt{2\log(n^d)}$. The GLRT is near-optimal.

The formula for the detection threshold $\sim \sqrt{2\log(n^d)}$ is perhaps surprising; as there are $O(n^{2d})$ distinct rectangles in $\mathbb{I}_n^d$, superficially, it seems that a detection boundary $\sim \sqrt{2\log(n^{2d})}$ might initially have been expected.

The GLRT, while near-optimal, may not be a very practical strategy. Straightforward implementation requires $O(n^{3d})$ flops; as compared to the number $N = n^d$ of data, this is of algorithmic order $O(N^3)$ flops. By a multiscale geometric approach, one can obtain algorithmic order much closer to $O(N)$.

Theorem 3.3: There is an algorithm operating in $O(n^{d+\eta})$ flops for each $\eta > 0$ which is near-optimal for detecting axis-aligned rectangles. The algorithm is asymptotically powerful to detect axis-aligned rectangles of amplitude $(1+\eta) \cdot \sqrt{2\log(n^d)}$ or greater, for each $\eta > 0$.

Clearly, there is a very strong parallel between the statement of these results and the corresponding results stated in previous sections in the case of intervals in dimension one and line segments in dimension two. There is also a very strong parallel between the proofs in the two cases. Rather than argue for these results directly, we now develop a general viewpoint able to encompass these results and many others.

IV. ABSTRACTING THE ESSENTIALS

A basic abstract structure underlies several different concrete problems addressed in this paper. Consider the d-dimensional cube $[0, 1]^d$, and subdivide it into $1/n \times \cdots \times 1/n$ generalized pixels. We are interested in a class $\mathcal{S}_n$ of subsets of this cube; corresponding to each $S \in \mathcal{S}_n$, there is the corresponding pixel array $\xi_S(p) \propto |S \cap p|$, where p denotes a generalized pixel and $|\cdot|$ denotes length, area, volume, etc., according to the class $\mathcal{S}_n$. We normalize so that $\|\xi_S\|_{\ell^2} = 1$. We observe data $X = (x(p))$ obeying

$$X = A\xi_S + Z$$

where Z is white Gaussian noise and $A \geq 0$ is the signal amplitude. Under $H_0 : A = 0$, while under $H_{1,n} : A > 0$ and $S \in \mathcal{S}_n$ is arbitrary and unknown.

A. Asymptotics of the Optimal Detection Threshold

Suppose that H_0 holds. We then view the data X as a Gaussian field $X[S]$ indexed by sets $S \in \mathcal{S}_n$, according to

$$X[S] = \langle X, \xi_S \rangle.$$

The GLRT statistic in this setting is

$$X^*[\mathcal{S}_n] = \max_{S \in \mathcal{S}_n} X[S].$$

In a broad range of problems this statistic exhibits threshold phenomena parallel to those seen so far, and which can be seen to hold under rather general conditions, which we now state. In what follows, $C_{k_1,k_2,\ldots}$ denotes an unspecified positive term depending on $k_1, k_2, \ldots$, and varying from occurrence to occurrence.

[DT-1]   *Exponent of Apparent Dimension.* For a certain exponent a, the class of sets $\mathcal{S}_n$ has cardinality $O(n^a)$.

We metrize the class of sets according to the metric

$$\delta(S, S') = (1 - \rho(S, S'))^{1/2}$$

where the affinity $\rho(S, S') = \langle \xi_S, \xi_{S'} \rangle$.

[DT-2] *Exponent of Effective Dimension.* For a certain exponent $e \leq a$, there is a sequence $(\mathcal{N}_{n,\ell})$ of ϵ_ℓ-nets for $\mathcal{S}_n$ under metric δ. Here, $\epsilon_\ell \to 0$ as $\ell \to \infty$, and the sequence of nets has cardinality bounded in the form $\#\mathcal{N}_{n,\ell} \leq C_{\ell,\eta} \cdot n^{e+\eta}$, as $n \to \infty$, for each $\eta > 0$.

[DT-3] *Lower Bound on Effective Dimension.* There is a subcollection $\mathcal{P}_n \subset \mathcal{S}_n$ of disjoint sets with, for each $\eta > 0$, asymptotically at least $C_\eta n^{e-\eta}$ elements.

In the study of intervals in Section II, these ingredients are all clearly present. There are $O(n^2)$ intervals contained in $\{1, \ldots, n\}$, so that [DT-1] holds with $\mathcal{S}_n = \mathbb{I}_n$ and exponent $a = 2$. The extension process gives us a sequence of ϵ_ℓ-nets with controlled cardinality, so that [DT-2] holds with $\mathcal{N}_{n,\ell} = \mathbb{J}_{n,\ell}$ and $e = 1$. Finally, there are $n/2$ disjoint intervals of length 2, so that [DT-3] holds with $\mathcal{P}_n \equiv P_n$.

In the case of d-dimensional digital rectangles $\mathbb{I}_n^d$, there are, of course, $O(n^{2d})$ rectangles so that [DT-1] holds, with $a = 2d$. Taking systematic Cartesian products of dyadic intervals

$$R = I_1 \times \cdots \times I_d$$

we produce the class $\mathbb{J}_n^d$ of dyadic rectangles; this has $O(n^d)$ elements. From the obvious formula

$$\rho_d(I_1 \times \cdots \times I_d, J_1 \times \cdots \times J_d) = \Pi_{j=1}^d \rho_1(I_j, J_j)$$

and the one-dimensional result, we have that $\mathbb{J}_n^d$ is an ϵ_d-net for $\mathbb{I}_n^d$, with $\epsilon_d < 1$. Taking systematic Cartesian products of nondyadic intervals from $\mathbb{J}_{n,\ell}^1$ gives us a collection of nondyadic rectangles $\mathbb{J}_{n,\ell}^d$. These collections provide a sequence of $\epsilon_{d,\ell}$-nets for $\mathbb{I}_n^d$ with appropriately controlled cardinality, and with $\epsilon_{d,\ell} \to 0$ as $\ell \to \infty$ with d fixed, so [DT-2] holds, with $\mathcal{N}_{n,\ell} = \mathbb{J}_{n,\ell}^d$ and exponent $e = d$. Finally, we build a collection $P_{n,d}$ of $(n/2)^d$ disjoint $2 \times 2 \times \cdots \times 2$ rectangles. Then [DT-3] holds with $\mathcal{P}_n = P_{n,d}$.

The following result, proved in Appendix II, generalizes the detection threshold arguments of Section II to the abstract setting of this section.

Theorem 4.1: When [DT-1]–[DT-3] hold, the optimal detection threshold $\sim \sqrt{2 \log(n^e)}$, and the GLRT is near-optimal.

As we have shown that [DT-1]–[DT-3] hold in the axis-aligned rectangle case, with $a = 2d$ and $e = d$, Theorem 3.3 follows immediately.

B. Asymptotics of the Optimal Algorithmic Complexity

We now state general conditions under which a simple algorithm, resembling the "extend promising intervals" approach of Section II, gives optimal-order asymptotic complexity. There is, however, an extra wrinkle: a *subpixel resolution* parameter $r \in \{0, 1, 2, \ldots\}$ which has been invisible until now, always taking the value 0.

[FA-1] *Sparse Generation.* There is a sequence $(\mathcal{G}_{n,r})$ of *generating* sets with cardinality $\#\mathcal{G}_{n,r} \leq C_{r,\eta} n^{e+\eta}$ for each $\eta > 0$; this class sparsely generates approximations to the ϵ_ℓ-nets $\mathcal{N}_{n,\ell}$ in

the following sense. There are constants $(\epsilon_{\ell,r})$ and $(k_{\ell,r})$, independent of n, so that for every $E \in \mathcal{N}_{n,\ell}$, there is a collection of $m(E) \leq k_{\ell,r}$ generators G_i in $\mathcal{G}_{n,r}$ with

$$\delta\left(E, \cup_{i=1}^{m(E)} G_i\right) \leq \epsilon_{\ell,r}.$$

The errors in the approximation obey, for a sequence $r_\ell \to \infty$, $\ell \to \infty$

$$\epsilon_{\ell,r} \to 0, \quad \ell \to \infty, \quad r \geq r_\ell.$$

[FA-2] *Fast Transform.* For each fixed r, it is possible to evaluate the random field X at every generator $G \in \mathcal{G}_{n,r}$, in $C_{r,\eta} n^{e+\eta}$ flops, for each $\eta > 0$.

[FA-3] *Effective Organization.* The generator classes $\mathcal{G}_{n,r}$ are effectively organized in this sense. With $C_{r,\ell}$ computations, we can enumerate all members of $\mathcal{G}_{n,r}$ needed to build up an approximation to any specific element $E \in \mathcal{N}_{n,\ell}$.

These assumptions were in effect shown to hold in the interval detection case, with r an invisible parameter set always to 0, and for generators $\mathcal{G}_{n,r} = \mathbb{J}_n$, the dyadic intervals. Indeed, the net $\mathcal{N}_{n,\ell} = \mathbb{J}_{n,\ell}$ has each net element built out of a limited number of dyadic intervals, giving [FA-1] with $\epsilon_{\ell,r} \equiv 0$. There is a fast transform for obtaining all sums over dyadic intervals, giving [FA-2]. The extension process is well structured, making it possible to effectively enumerate the ℓ-level extensions and to effectively identify which dyadic intervals go into the construction of any given member of $\mathbb{J}_{n,\ell}$, hence [FA-3] holds. Similar properties can be seen to hold for the rectangle detection case, with again r a hidden parameter always set to 0, and with $\mathcal{G}_{n,r} = \mathbb{J}_n^d$, the dyadic rectangles.

The role of the subpixel resolution parameter r will become more clear as it is used in sections below.

When these assumptions hold, we have a fast, near-optimal algorithm. The following is proved in Appendix III.

Theorem 4.2: When [DT-2]–[DT-3] and [FA-1]–[FA-3] hold, there is an algorithm which is near optimal and which operates in $O(n^{e+\eta})$ flops, for each $\eta > 0$. Moreover, no near-optimal algorithm can run in $O(n^{e-\eta})$ flops for any $\eta > 0$.

Theorem 3.3 essentially follows, although the specific complexity estimate stated in that theorem—$O(n^d)$—is slightly better than the one following immediately from the above result—$O(n^{d+\eta})$. The tightening of the estimate is based on a straightforward use of ideas already evident in our discussion of the intervals case in Section II.

C. Overview

Table I relates the abstract setting to the intervals and rectangles cases:

Summarizing the situation formally, we have the following definition.

Definition 4.3: If [DT-1]–[DT-3] and [FA-1]–[FA-3] hold with specific values of exponents a and e the same throughout,

TABLE I

Setting	$\mathcal{S}_n$	$\mathcal{N}_{n,\ell}$	$\mathcal{G}_{n,r}$	e	a
Intervals	$\mathbb{I}_n$	$\mathbb{J}_{n,\ell}$	$\mathbb{J}_n$	1	2
Rectangles	$\mathbb{I}_n^d$	$\mathbb{J}_{n,\ell}^d$	$\mathbb{J}_n^d$	d	$2d$

Setting	Threshold	Flops
Intervals	$\sim \sqrt{2\log(n)}$	$O(n)$
Rectangles	$\sim \sqrt{2\log(n^d)}$	$O(n^d)$

we say that n^a is the **apparent dimension** and n^e the **effective dimension** of the class $\mathcal{S}_n$.

Applying this terminology, we can characterize the optimal detection threshold and the algorithmic complexity of a near-optimal detector in terms of the effective dimension n^e

$$\tau_n \sim \sqrt{2\log(n^e)}, \qquad n \to \infty$$
$$\#\text{Flops} = O(n^{e+\eta}), \qquad \forall \eta > 0.$$

We also remark that in all these cases, the exponent in the effective dimensionality of the problem is apparently the same as the dimensionality of the space: $e = d$, rather than, as the apparent dimension might suggest, $2d$.

We will see other instances of this pattern below, as we consider a range of geometric object classes for "image" data with ambient dimension $d = 2$. We summarize the results to come in Table II, the terms all being defined later in the paper.

In each case, the exponent of apparent dimension is twice as large as the exponent of effective dimension, which in fact equals the ambient space dimension d.

V. DETECTING DISKS IN NOISY IMAGES

Let $\mathbb{D}_n$ denote the set of planar disks with maximal inscribed square belonging to $\mathbb{I}_n^2$, and whose interior does not intersect the boundary of $[0,1]^2$. The disks correspond to pixel arrays $\xi_F(p) \propto |F \cap p|$ where $|\cdot|$ denotes area and we normalize so that $\|\xi_F\|_{\ell^2} = 1$. The abstract approach gives the following results.

Theorem 5.1: For detecting disks in $\mathbb{D}_n$, the optimal detection threshold $\sim \sqrt{2\log(n^2)}$, and GLRT is near-optimal.

Theorem 5.2: For detecting disks in $\mathbb{D}_n$, there is a near-optimal detector which runs in $O(n^{2+\eta})$ flops for each $\eta > 0$.

To obtain these results, we let $\mathcal{S}_n = \mathbb{D}_n$, construct nets and generating sets satisfying conditions [DT-1]–[DT-3] and [FA-1]–[FA-3] with exponent of effective dimension $e = 2$, and then invoke Theorems 4.1 and 4.2.

A. Verifying [DT-1]–[DT-3] for Disks

For [DT-1], there are $O(n^3)$ digital disks, defined by $O(n^2)$ centers and $O(n)$ possible side lengths of the maximal inscribed squares, so for $\mathcal{S}_n = \mathbb{D}_n$, we have exponent of apparent dimension $a = 3$.

For [DT-2], we will map a sequence of increasingly fine nets for rectangles onto a corresponding sequence of nets for disks.

TABLE II

Setting	$\mathcal{S}_n$	$\mathcal{N}_{n,\ell}$	$\mathcal{G}_{n,r}$	e	a
Disks	$\mathbb{D}_n$	$D^*(\mathbb{J}_{n,\ell}^2)$	$\mathbb{J}_{n,r}^2$	2	4
Line Segments	$\mathbb{S}_n$	$\mathbb{B}_{n,2^{-r_\ell},\ell}$	$\mathbb{B}_{n,2^{-r}}$	2	4
Rectangles	$\mathbb{R}_n^2$	$R^*(\mathbb{X}_{n,2^{-r_\ell},\ell})$	$\mathbb{X}_{n,r}$	2	4
Ellipsoids	$\mathbb{E}_n^2$	$E^*(\mathbb{X}_{n,2^{-r_\ell},\ell})$	$\mathbb{X}_{n,r}$	2	4

Setting	Threshold	Flops
Disks	$\sim \sqrt{2\log(n^2)}$	$O(n^2)$
Line Segments	$\sim \sqrt{2\log(n^2)}$	$O(n^2 \log(n))$
Rectangles	$\sim \sqrt{2\log(n^2)}$	$O(n^2 \log^2(n))$
Ellipsoids	$\sim \sqrt{2\log(n^2)}$	$O(n^2 \log^2(n))$

Naturally associated to each rectangle R is a particular circular disk D—the minimal area disk containing R

$$D^*(R) = \operatorname{argmin}\{|D| : R \subset D\}.$$

We also introduce a continuum analog of δ: for sets $R, S \subset [0,1]^2$ define

$$\Delta(R,S) = \sqrt{1 - \frac{|R \cap S|}{\sqrt{|R|}\sqrt{|S|}}}$$

where now $|\cdot|$ denotes the Lebesgue measure.

Lemma 5.3: For all sets R, S, $\delta(R,S) \le \Delta(R,S)$.

This follows directly from

$$\sqrt{2}\delta(R,S) = \|\xi_R - \xi_S\|_2$$
$$= \|\text{Pix}(\mathbf{1}_R) - \text{Pix}(\mathbf{1}_S)\|_2$$
$$\le \|\mathbf{1}_R - \mathbf{1}_S\|_2$$
$$= \sqrt{2}\Delta(R,S)$$

where "Pix" denotes the pixelization operator

$$\text{Pix}(f)(p) = \text{Ave}\{f|\text{Pixel}(p)\}$$

and we used the fact that Pix is a linear operator from L^2 to ℓ^2 of norm 1.

The mapping $R \mapsto D^*(R)$, when restricted to the subset of rectangles which are squares, is surjective and uniformly continuous with respect to the Δ-metric. Because $\delta(F,G) \le \Delta(F,G)$ for disks $F, G \in \mathbb{D}_n$ and $\delta(R,S) = \Delta(R,S)$ for $R, S \in \mathbb{I}_n^2$, this implies that D^* is also uniformly (equi-)continuous with respect to each δ-metric. Here we say *equi-continuity* to make clear that although actually $\delta \equiv \delta_n$ depends on n, we actually have that D^* is uniformly continuous in every δ_n metric, uniformly so in n.

Recalling notions from the rectangle case, consider the collection of extended rectangles $\mathbb{J}_{n,\ell}^2$. It has cardinality at most $C_\ell \cdot n^2$. We saw that $\mathbb{J}_{n,\ell}^2$ offers an ϵ_ℓ-net (under the δ metric) for the space $\mathbb{I}_n^2$; here, $\epsilon_\ell \to 0$ as $\ell \to \infty$. Extract out of this collection of rectangles the collection of squares $\mathbb{Q}_{n,\ell}^2$ (say), and consider the collection of disks induced by pushing forward under D^*

$$\mathbb{D}_{n,\ell} = \left\{ D^*(Q) : Q \in \mathbb{Q}_{n,\ell}^2 \right\}.$$

Now, because D^* is onto and equi-continuous, $\mathbb{D}_{n,\ell}$ is a sequence of η_ℓ-nets for the space of disks $\mathbb{D}_n$, where $\eta_\ell \to 0$ as $\ell \to \infty$. At the same time

$$\#\mathbb{D}_{n,\ell} = \#\mathbb{Q}_{n,\ell}^2 \leq \#\mathbb{J}_{n,\ell}^2 \leq C_\ell \cdot n^2.$$

Setting now $\mathcal{N}_{n,\ell} = \mathbb{D}_{n,\ell}$, condition [DT-2] follows, with exponent $e = 2$.

To get [DT-3], simply break the n-by-n grid into 16-by-16 pixel blocks, each containing a 4-by-4 concentric square, and a concentric disk having that square as its maximal inscribed square. In this way, we get a collection $\mathcal{P}_n$ of $n^2/256$ disjoint disks with radius $\sqrt{8}/n$.

Having established [DT-1]–[DT-3], Theorem 5.1 follows.

B. Verifying [FA-1]–[FA-3] for Disks

The case of disks gives us our first opportunity to illustrate the role of the subpixel resolution parameter r. Let $\lambda = 2^{-r} \leq 1$ and consider the subpixel grid $V_{n,r}$ of points in the unit square whose coordinates are integer multiples of λ/n. Let $\mathbb{J}_n^{2,r}$ denote the collection of axis-aligned rectangles with vertices in the subpixel-resolved grid $V_{n,r}$. Forming unions of such rectangles (obviously) allows us finer approximations to the curved edges of disks. Notwithstanding the fact that this collection of rectangles, viewed as continuum objects, is subpixel resolved, we consider as corresponding pixel arrays $\xi_R \propto |R \cap p|$ only arrays which, as before, are defined by pixels of ordinary resolution $1/n$. We also introduce $\mathbb{Q}_{n,\ell}^2$ to denote the collection of squares based on products of intervals in $\mathbb{J}_{n,\ell}^2$.

To get condition [FA-1], take for our class of generators $\mathcal{G}_{n,r} = \mathbb{J}_n^{2,r}$. We claim the following.

i) $\#\mathcal{G}_{n,r} \leq (2n/\lambda)^2 \quad \forall n, \lambda$.

ii) There is a fast transform computing $(X[G], G \in \mathcal{G}_{n,r})$ in $\leq C(n/\lambda)^2$ flops.

iii) There are constants $(\epsilon_{\ell,r})$, and $(k_{\ell,r})$ so for any disk $D = D^*(Q)$ arising from a square $Q \in \mathbb{Q}_{n,\ell}^2$, there is a list $\mathcal{G}(D)$ of $m(D) \leq k_{\ell,r}$ generators $G_i \in \mathcal{G}_{n,r}$ such that

$$\delta\left(D, \cup_{i=1}^{m(D)} G_i\right) \leq \epsilon_{\ell,r}.$$

Moreover, if r_ℓ is a sequence tending to ∞ with ℓ

$$\epsilon_{\ell,r_\ell} \to 0, \quad \ell \to \infty.$$

iv) To construct the list $\mathcal{G}(D)$ takes at most $C_{\ell,r}$ operations for any D.

Claims i) and ii) are already obvious from our earlier discussion of rectangles. The idea behind iii) and iv) is to dissect the disk in a canonical way into (nondyadic) rectangles, to approximate those rectangles by unions of dyadic rectangles, and to glue together these unions.

Fig. 2 shows the "standard dissection" of the standard disk (center 0, radius 1) into rectangles. The j largest area terms in this dissection combine to approximate the disk to within distance $\leq \eta_j$ (say) in Δ metric.

Given any disk D as described in Claim iii), we can dissect it in a parallel fashion, simply by translating and scaling the rectangles in the standard dissection. Considering the j largest rectangles in the scaled and translated list, their union gives an
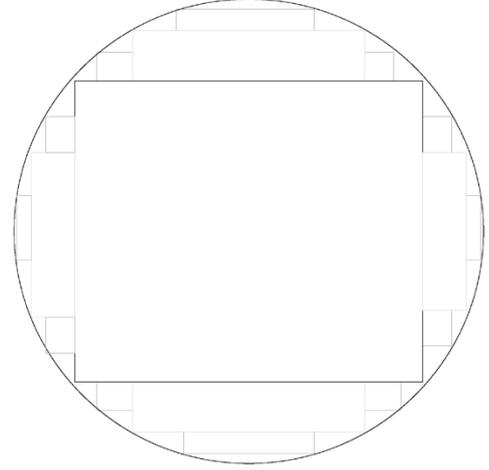

Fig. 2. The standard dissection of the disk into (nondyadic) rectangles.

approximation error $\leq \eta_j$, independent of the disk's center and radius. Indeed, this property holds for the standard dissection, and Δ is invariant under translation and scaling.

Each rectangle in the resulting list can itself be approximated by ℓ-level extensions of included dyadic rectangles. Lemma 5.4 shows that the union of approximants is a good approximation to the union of the approximands. The list of all the dyadic rectangles involved, for given D, ℓ, and r, can be constructed in $\leq C_{\ell,r}$ operations. Claims iii) and iv) follow, and Theorem 5.2 is proved. ∎

In this argument we invoked the following, proved in Appendix IV.

Lemma 5.4: Let (R_i) and (S_i), be sequences of sets of the same Hausdorff dimension, with

1) $|R_i \cap R_j| = 0$ and $|S_i \cap S_j| = 0$, for all $i \neq j$;
2) $\Delta(R_i, S_i) \leq \eta$ for all i.

Then, for η small enough

$$\Delta(\cup_i R_i, \cup_i S_i) \leq \sqrt{3}\eta.$$

VI. DETECTING LINE SEGMENTS IN IMAGES

As in the Introduction, consider line segments buried in noise in two-dimensional images. We will show that the above abstract ingredients hold in this setting. The generators will be based on beamlets, and the nets will be based on chains of beamlets.

A. Beamlets

For a dyadic $n = 2^J$ and a dyadic fraction $\lambda = 2^{-r} < 1$, let $\mathbb{B}_{n,\lambda}$ denote a family of line segments in $[0,1]^2$ defined as follows.

Let Q be a dyadic square in $[0,1]^2$ with side length 2^{-j} $j = 0, \ldots, J$, and on the boundary ∂Q of each such square, mark out vertices equally spaced with spacing $\nu = \lambda/n$, starting at the corners. See Fig. 3.

Let $\mathbb{B}_\lambda[Q]$ be the collection of all line segments joining pairs of marked vertices in ∂Q, and let $\mathbb{B}_{n,\lambda} = \cup_Q \mathbb{B}_\lambda[Q]$. $\mathbb{B}_{n,\lambda}$ is a dyadically organized set of line segments, occupying a range of orientations, scales, and locations. This system has been discussed at length by Donoho and Huo [28], [29]; see also [27]
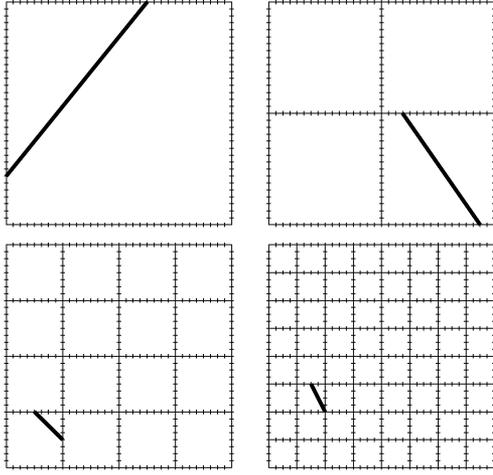
Fig. 3. Dyadic squares marked with vertices, and several beamlets.
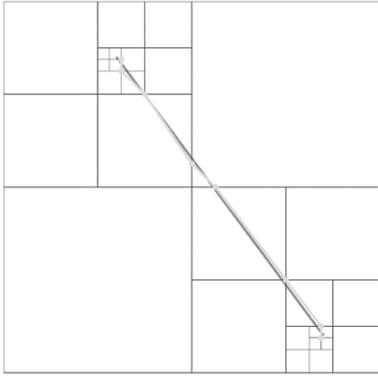

Fig. 4. Approximating a line segment (dark) by a chain of beamlets (light).

where the same system was called *edgelets*. It has the following properties:

1) $\#\mathbb{B}_{n,\lambda} \leq C \cdot \lambda^{-2} \cdot n^2 \log(n)$, with C not depending on n or λ; and
2) each line segment connecting two pixels can be approximated within Hausdorff distance $\nu = \lambda/n$ by a connected chain of at most $4\log(n)$ beamlets $b \in \mathbb{B}_{n,\lambda}$. Fig. 4 gives a suggestive illustration.

B. Beamlet Arrays

Beamlets are planar line segments, while detectors must be based on pixel arrays. Let S be any line segment, and let ξ_S denote the $n \times n$ pixel array with entries

$$\xi_S(p) = |S \cap p|/N(S)$$

where $|\cdot|$ denotes length, and $p = (p_x, p_y)$ indexes a pixel $[p_x, p_x + 1/n) \times [p_y, p_y + 1/n)$, and $N(S)$ is a normalizing scale factor guaranteeing $\|\xi_S\|_{\ell^2} = 1$.

Using this correspondence, it is possible to translate the claim made above, about approximating arbitrary line segments by chains of a few beamlets from a claim about approximation in Hausdorff distance to a claim about approximation of pixel arrays in δ distance, which is the one relevant for detection.

To see how, let $C = \cup_i B_i$ denote a chain of beamlets—a connected set formed by concatenating beamlets with disjoint interiors, and consider the corresponding pixel array $\xi_C(p) = |C \cap p|/N(C)$, with normalization factor N, defining $\rho(S, C) = \langle \xi_S, \xi_C \rangle$ as usual. Appendix V proves the following approximation result.

Lemma 6.1: There is a constant $K > 0$ independent of n such that, if $\lambda < 1$, then with S a line segment in $\mathbb{S}_n$ and C a chain of $\mathbb{B}_{n,\lambda}$-beamlets having

$$\text{Hausdorff}(S, C) \leq h/n, \qquad 0 \leq h < 1/4$$

we have

$$\rho(S, C) \geq 1 - K \cdot h^{1/2}, \qquad \forall\, S, C.$$

C. Maximal Included Beamlet

A central observation for our study of interval detection was the fact that dyadic intervals form a $\sqrt{1/2}$-net for the space of all intervals. This fact is, in turn, based on the observation that every interval has a maximal dyadic interval having at least one-fourth the length. A parallel observation can be made in the line segment case.

Let $\mathbb{B}_{n,0}$ denote the continuous system consisting of *all* line segments connecting pairs of boundary points in some dyadic square of side length $\geq 1/n$. This is a natural limiting form of $\mathbb{B}_{n,\lambda}$ as $\lambda \to 0$. For this system, we can show that every line segment has a beamlet that coincides with it on a substantial portion of its length; the following is proved in Appendix VI.

Lemma 6.2: For every line segment $S \in \mathbb{S}_n$ there is a beamlet $B \in \mathbb{B}_{n,0}$ such that

$$B \subset S \quad \text{and} \quad \ell(B) \geq \ell(S)/8.$$

The implications emerge from the following result, proven in Appendix VII.

Lemma 6.3: For $S_1 \subset S_2$ two line segments joining pixel corners, we have

$$\rho^2(S_1, S_2) \geq \frac{\ell(S_1)}{2\ell(S_2)} \tag{6.1}$$

$$\rho^2(S_1, S_2) \geq 1 - 2\left(1 - \frac{\ell(S_1)}{\ell(S_2)}\right). \tag{6.2}$$

Lemma 6.2, together with (6.1), gives that $\mathbb{B}_{n,0}$ is an ϵ_0-net for $\mathbb{S}_n$ in the δ-metric, with $\epsilon_0 \leq \sqrt{3}/2$. Hence, by Lemma 6.1, we see that $\mathbb{B}_{n,\lambda}$ is an ϵ_0^λ-net for $\mathbb{S}_n$, with $\epsilon_0^\lambda \to \epsilon_0$ as $\lambda \to 0$.

In applying this, we will need the following.

Lemma 6.4:

$$|\rho(A, B) - \rho(B, C)| \leq \sqrt{2}\delta(A, C).$$

This follows from Cauchy–Schwartz

$$|\langle \xi_A, \xi_B \rangle - \langle \xi_B, \xi_C \rangle| = |\langle \xi_A - \xi_C, \xi_B \rangle|$$
$$\leq |\xi_A - \xi_C|$$
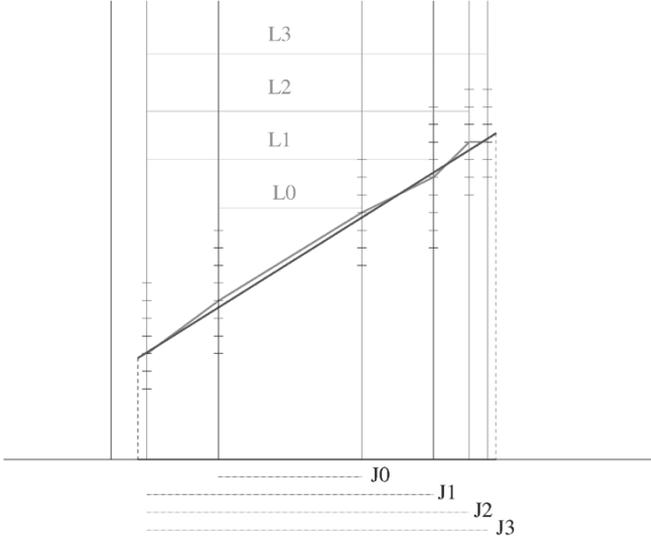$$= \sqrt{2}\delta(A, C).$$

Fig. 5.   Vertical strips, bases, and extensions.

D. Chains of Beamlets

We now construct chains of beamlets that provide reasonable approximation to any given line segment while having strictly controlled cardinality.

The chains in our construction come in two kinds, depending on whether they are meant to approximate line segments which make an angle less than $45°$ with the x-axis. If they do, we call them *basically horizontal*; otherwise, we call them *basically vertical*. We describe the construction of basically horizontal chains only, as the construction is similar in the other case.

The construction begins by defining a collection of special line segments and then approximating them by chains of beamlets. The construction of line segments works, as in the one-dimensional case, by extending a base.

A baseline segment L_0 is depicted in Fig. 5. It has for its projection on the x-axis an interval J_0 which is a base for the one-dimensional extension process discussed in Section II; namely, J_0 is either a dyadic interval $I_{j,k}$ (say) or the union of $I_{j,k}$ with its equal-length non-sibling-adjacent dyadic interval. L_0 itself joins vertices v_0 and v_1 on opposite sides of the vertical strip

$$\text{Strip}(J_0) = \{(x, y) : x \in J_0;\ y \in [1]\}$$

where $|v_{0,y} - v_{1,y}| \le |J_0|$.

To form an extension L_1, we first extend J_0, appending a dyadic interval of length $2^{-1}|I_{j,k}|$ on either side, or else appending nothing. Call the result J_1. We then project J_1 vertically onto the line spanned by L_0, producing L_1. We build higher level extensions L_2, L_3, etc., in a similar fashion.

Let $\mathbb{L}_{n,\lambda,\ell}$ denote the collection of all line segments reachable from this construction by ℓ stages of extension. Each such line segment can be approximated by a chain of beamlets with Hausdorff distance $\le \nu$; let $\mathbb{B}_{n,\lambda,\ell}$ denote the collection of such chains.

The collection $\mathbb{B}_{n,\lambda,\ell}$ has controlled cardinality and the sequence of such collections provides increasingly fine nets for the space $\mathbb{S}_n$. Appendix VIII proves the following.

Lemma 6.5:

$$\#\mathbb{B}_{n,\lambda,\ell} \le (n/\lambda)^2 \cdot 4^{\ell+1} \cdot (\log_2 N + 1) \qquad (6.3)$$

$$\rho^*_{n,\lambda,\ell} = \min_{s \in \mathbb{S}_n} \max_{c \in \mathbb{B}_{n,\lambda,\ell}} \rho(s,c) = 1 - O\left(\sqrt{\lambda} + 2^{-l/2}\right). \qquad (6.4)$$

Note particularly that these two results allow us to choose λ small and ℓ large, independently of n, in such a way that, for all large n, $\rho^*_{n,\lambda,\ell}$ is close to one and yet

$$\#\mathbb{B}_{n,\lambda,\ell} \le C_{\lambda,\ell} \cdot N^2 \log(n).$$

E. The Abstract Properties

We have in passing almost shown that the abstract properties [DT-1]–[DT-3] mentioned in Section IV hold in the present setting. We now complete the demonstration.

For [DT-1], with $\mathcal{S}_n = \mathbb{S}_n$, there are $\sim n^4/2$ line segments in $\mathbb{S}_n$, so [DT-1] holds with apparent dimension $a = 4$.

For [DT-2], pick a sequence $r_\ell \to \infty$ and $\lambda_\ell = 2^{-r_\ell}$. Define $\mathcal{N}_{n,\ell} = \mathbb{B}_{n,\lambda_\ell,\ell}$. Invoking Lemma 6.5, the level-ℓ extensions of the $\mathbb{B}_{n,\lambda_\ell}$ beamlets make up an $\epsilon_{\ell,\lambda_\ell}$-net for the space of line segments, with $\epsilon_{\ell,\lambda_\ell}$ as small as we like, for ℓ large enough; and the cardinality is strictly controlled $O\left(C_{\ell,\lambda_\ell} \cdot n^2 \log(n)\right)$ Hence [DT-2] holds with exponent of effective dimension $e = 2$.

Finally, turn to [DT-3]. Construct $n^2/2$ rectangles made by adjoining pairs of pixels which are horizontally adjacent and let $\mathbb{L}_n^2$ denote the collection of line segments connecting the southwest corner to the northeast corner of each such rectangle. Then the corresponding pixel arrays have disjoint supports and $\#\mathbb{L}_n^2 = n^2/4$. Hence, [DT-3] holds.

Theorem 1.2 of the Introduction follows; the optimal detection threshold $\sim \sqrt{2 \log(n^2)}$.

We now consider Theorem 1.3, concerning algorithmic complexity. Define our generating set $\mathcal{G}_{n,r}$ as $\mathbb{B}_{n,2^{-r},\ell}$, where $r \ge r_\ell$. Condition [FA-1] follows immediately from the definition of $\mathcal{N}_{n,\ell} = \mathbb{B}_{n,\lambda_\ell,\ell}$ as a concatenation of beamlets $\mathbb{B}_{n,\lambda_\ell}$ and the inclusion $\mathbb{B}_{n,\lambda_\ell} \subset \mathbb{B}_{n,2^{-r}} \equiv \mathcal{G}_{n,r}$, which follows from $r \ge r_\ell$.

For [FA-2], we invoke the Fast Beamlet Transform. This transform allows to calculate an accurate approximation to all statistics $X[B]$ in $O(n^2 \log(n))$ flops. The basic idea is the two-scale relation: each $B \in \mathbb{B}_{n,0}$ is decomposable into a union $B = \cup_i B_i$ of at most three B_i arising as beamlets at the next finer scale of dyadic subdivision. So, for $B \in \mathbb{B}_{n,0}$, we can write

$$X[B] = \sum_i w_i X[B_i].$$

Hence, if we know already the value of $X[B_i]$ for beamlets arising from dyadic squares at the finer scale, the two-scale relation gives us the value of $X[B]$ at the coarser scale. The recursive application of this principle costs $O(n^2 \log(n))$ flops. [FA-3] follows immediately from the construction of $\mathcal{N}_{n,\ell}$ as chains of beamlets, and the organizational properties of those chains in the previous subsection.

The principle of approximate recursive computation in beamlet-like systems has been proposed by Brandt and Dym [16] and by Götze and Druckenmiller [42]. Other effective approaches include fast Radon transforms in [10], [14]. If $B \in \mathbb{B}_{n,\lambda}$, then any such fast algorithm does not give perfect identity, but instead high correlation

$$\rho(B, \cup_i B_i) > \rho^+(\lambda)$$

where $\rho^+(\lambda) \to 1$ as $\lambda \to 0$. The point is that for $\rho^+(\lambda)$ sufficiently close to 1, an acceptable approximation to $X[B]$ can be computed by the recursive algorithm.

With [FA-1]–[FA-3] established, Theorem 1.3 of the Introduction follows.

VII. "TILTED" GEOMETRIC SHAPES

We now briefly consider detection of two-dimensional shapes—rectangles and ellipsoids—which can be highly anisotropic and can be "tilted" with respect to the coordinate axes. This requires different generators than used for axis-aligned rectangles and disks. Our approach requires that, for detecting an object class $\mathcal{S}_n$, we need a set of generators $\mathcal{G}_{n,r}$ of cardinality $O(n^e)$ allowing to construct an ϵ-net for $\mathcal{S}_n$ with $\epsilon < 1$ independent of n, where the method of construction "glues" together several generators in constructing each element of the net, and the total number glued together in constructing an element is uniformly bounded, independently of n. However, general ellipsoids and rectangles are not axis-aligned and cannot be generated in a limited-cardinality fashion by gluing together axis-aligned dyadic rectangles. Instead , we need a set $\mathcal{G}_{n,r}$ of "tilted" generators for which there is a fast transform. We will build such generators essentially by thickening beamlets.

A. Fast Axoid Transform

Architectural drawings based on so-called *axonometric* drawing principles contain polygons where certain line segments are constrained to be horizontal or vertical, and others may be drawn in a possibly nonvertical, nonhorizontal direction. Consider the class $\mathbb{A}$ of *axoids*: parallelograms having two sides vertical or two sides horizontal. This class contains ordinary two-dimensional axis-aligned rectangles, but also sheared rectangles. Consider the subclass $\mathbb{A}_{n,\lambda}$ of axoids with corners on the subpixel-resolved grid. In detail, let again $V_{n,r}$ denote the vertices in the subpixel-resolved grid, with subpixel resolution parameter r and spacing $\lambda/n = 2^{-r}/n$. $\mathbb{A}_{n,\lambda}$ consists of axoids with vertices in $V_{n,r}$.

As long as we do not consider axoids which are too "small," $\mathbb{A}_{n,\lambda}$ provides a good uniform approximation. To formalize this statement, let $\mathbb{A}_n^*$ denote the collection of general axoids in $\mathbb{A}$ with both side lengths at least $1/2n$.

Lemma 7.1: $(\mathbb{A}_{n,2^{-r}})_r$ offers a sequence of ϵ_r-nets for $\mathbb{A}_n^*$, with ϵ_r independent of $n \geq n_0$, and

$$\epsilon_r \to 0, \qquad \text{as } r \to \infty.$$

Of course, the constraint $1/2n$ in defining the family $\mathbb{A}_n^*$ is somewhat arbitrary; the same result holds if it is replaced by $1/cn$ for any $c > 1$. We omit the proof of the lemma. The idea is simply that, although the axoids in $\mathbb{A}_n^*$ may have general endpoints not belonging to any grid $V_{n,r}$, because the endpoints are $\geq 1/2n$ apart, using endpoints in $V_{n,r}$ gives a positioning error $\leq 2^{-r}/n$, which can be much smaller than $1/2n$ for large r.

Also, consider the class $\mathbb{X}_{n,\lambda}$ of *dyadic axoids*: parallelograms for which a) the corners belong to $V_{n,r}$; b) one pair of sides is either perfectly vertical or perfectly horizontal, and of side length 2^{-j}, $j = 0, 1, 2, \ldots, \log_2(n) + r$; and c) the other pair of sides projects onto the horizontal (resp., vertical) axis in a dyadic interval. There are $O(\lambda^{-2} n^2 \log^2(n))$ dyadic axoids.

The following is essentially obvious, and we omit the proof.

Lemma 7.2: Let $\mathbb{X}_{n,\lambda,\ell}$ denote the collection of axoids obtained by ℓ-level extensions of dyadic axoids in $\mathbb{X}_{n,\lambda}$ by gluing along sides in common. The sequence $(\mathbb{X}_{n,\lambda,\ell})_\ell$ provides a sequence of ϵ_ℓ-nets for $\mathbb{A}_{n,\lambda}$ in Δ-distance, and $\epsilon_\ell \to 0$ as $\ell \to \infty$.

We also need the following algorithmic assertion.

Lemma 7.3: Given an n-by-n pixel array X, this transform approximately computes all inner products with dyadic axoids

$$X[A] : A \in \mathbb{X}_{n,\lambda}$$

in $O(C_\lambda n^2 \log^2(n))$ flops, and with error bounded by $\epsilon_\lambda \to 0$ as $\lambda \to 0$.

We sketch the argument. The idea is to create a stack of $\approx 2(\log_2(n) + r)$ arrays obtained by either vertically integrating or horizontally integrating the raw data, as follows. Let $X^{m,v}(p)$ be the sum over all p' in the same vertical column, obeying $p_x = p'_x$ but $p_y \geq p'_y > p_y - 2^m/n$. Similarly, let $X^{m,h}(p)$ in the same vertical column, obeying $p_y = p'_y$ but $p_x \geq p'_x > p_x - 2^m/n$. In this way, we form $2(\log_2(n) + r)$ arrays $X^{m,\cdot}$ with $m = 0, \ldots, \log(n) + r$. We then apply the fast beamlet transform on each of these arrays, and obtain in effect a broad collection of integrals over dyadic axoids, with dyadic length and dyadic thickness. The accuracy estimate follows from the accuracy of the fast beamlet transform.

B. Tilted Rectangles

Consider now the class of tilted rectangles, whose sides need not be parallel to the axes. $\mathbb{R}_n^2$ denotes the collection of all rectangles where

- the center of the rectangle is at a pixel corner;
- some corner of the rectangle is at a pixel corner;
- the interior of the rectangle does not intersect the boundary of $[0,1]^2$.

We note that every such rectangle contains at least five grid points from $P_{n,0}$: its center and the four nearest neighbors. We have the following results.

Theorem 7.4: For detection of tilted rectangles in $\mathbb{R}_n^2$, the optimal detection threshold $\sim \sqrt{2 \log(n^2)}$, and GLRT is near-optimal.

Theorem 7.5: For detection of tilted rectangles in $\mathbb{R}_n^2$, there is a near-optimal detector which runs in $O(n^{2+\eta})$ flops for each $\eta > 0$.

The proofs are analogous to those in the disk case, with the following changes:

- dyadic axoids replace subpixel-resolved dyadic rectangles in the role of generators;
- minimal circumscribing rectangles replace minimal circumscribing disks in the role of associating generators to net elements.

1) Detection Threshold: The main work in checking the abstract [DT] conditions comes with [DT-2]. Indeed, we easily see that [DT-1] holds with $a = 4$, as enumerating the centers and corners of all possible tilted rectangle in $\mathbb{R}_n^2$ gives an $O(n^4)$ upper bound for the number of different such rectangles. Also, [DT-3] automatically holds for tilted rectangles by inheritance from the special case of axis-aligned rectangles, which was already discussed in Sections III and IV.

So consider [DT-2]. We first remark that we can associate each axoid A with the collection $R^*(A)$ of rectangles of minimal area that contains it (while there is typically a single such minimal rectangle in $R^*(A)$, the example of a rhombus pitched $45°$ shows that there can be two). We can also associate each rectangle with the collection of maximal inscribed axoids (again there can be two when the pitch of the rectangle is $45°$); let $A^*(R)$ denote the set-valued association. We next remark that each rectangle $R \in \mathbb{R}_n^2$ has a maximal inscribed axoid with each side length at least $1/2n$. Indeed, each such R contains five standard gridpoints in a "cross" configuration, at least $1/n$ apart. Therefore, it contains a square of side $1/2n$. It follows that $\text{image} A^*\left(\mathbb{R}_n^2\right) \subset \mathbb{A}_n^*$ and so

$$\mathbb{R}_n^2 \subset \text{image } R^*(\mathbb{A}_n^*). \tag{7.1}$$

As a set-valued mapping, R^* has the partial continuity property that, for every Δ-convergent sequence A_i

$$\lim_i R^*(A_i) \subset R^*\left(\lim_i A_i\right). \tag{7.2}$$

It follows that the image of R^* maps a sequence of increasingly fine nets for axoids into a sequence of increasingly fine nets for rectangles.

We can get a sequence of increasingly fine nets for axoids from the extensions $\mathbb{X}_{n,\lambda,\ell}$. Indeed, Lemma 7.2 says that these make an ϵ_ℓ-net for $\mathbb{A}_{n,\lambda}$, with $\epsilon_\ell \to 0$ as $\ell \to \infty$, and Lemma 7.1 tells us that $\mathbb{A}_{n,2^{-r}}$ makes an ϵ_r-net for $\mathbb{A}_n^*$, with $\epsilon_r \to 0$ as $r \to \infty$. Combine these, picking $r_\ell \to \infty$ as $\ell \to \infty$, and setting $\lambda_\ell = 2^{-r_\ell}$, to get that for $n \geq n_0$, $(\mathbb{X}_{n,\lambda_\ell,\ell})_\ell$ furnishes a sequence of ϵ_ℓ-nets in Δ-metric, with $\epsilon_\ell \to 0$ as $\ell \to \infty$. Define now the pointset $\mathcal{N}_{n,\ell} = R^*(\mathbb{X}_{n,\lambda_\ell,\ell})$. From the partial continuity property (7.2), this defines a sequence of increasingly fine nets for rectangles in $\mathbb{R}_n^2$.

Finally, for each ℓ, the cardinality of the ℓth net $\mathbb{X}_{n,\lambda_\ell,\ell}$ obeys $\#\mathbb{X}_{n,\lambda_\ell,\ell} \leq C_\ell n^2 \log^2(n)$, so from $\#\mathcal{N}_{n,\ell} \leq \#\mathbb{X}_{n,\lambda_\ell,\ell}$ we see that constraint [DT-2] holds with $e = 2$. Theorem 7.4 follows.

2) Algorithmic Complexity: We obtain [FA-1] by checking that every element of our net $\mathcal{N}_{n,\ell}$ can be well approximated by a finite list of dyadic axoids.
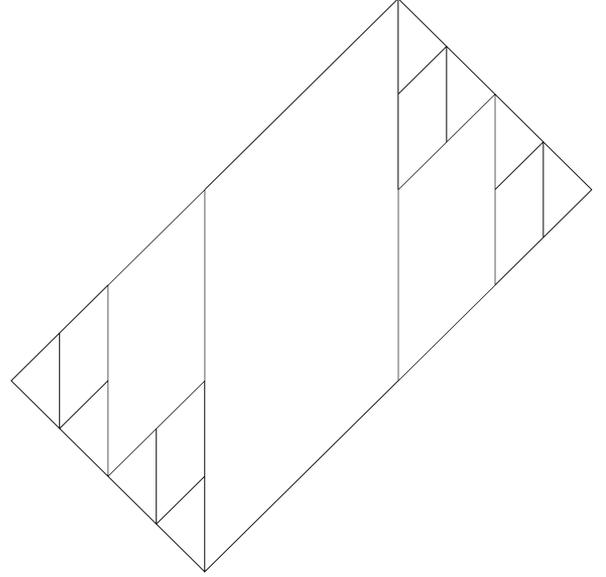

Fig. 6. Illustration of an axoid decomposition of a tilted rectangle.

Any given tilted rectangle can be dissected into a countable list of general axoids, as illustrated in Fig. 6. Let $\gamma_j(R)$ denote the maximal Δ-distance between a tilted rectangle R and the union of the first j axoids in such a dissection; we have the following.

Lemma 7.6:

$$\gamma_j^* \equiv \sup_R \gamma_j(R) \to 0, \qquad \text{as } j \to \infty.$$

To prove this, consider the dissection procedure in more detail. At stage one, a maximal inscribed axoid is extracted, and two triangles remain. At stage two, a maximal inscribed axoid is extracted from each of the two triangles, leaving four triangles, and so on. The maximal inscribed axoid in a triangle has exactly half the area in the triangle, and so, the residual area $|R \backslash (\cup_{i=1}^j A_i)| \to 0$ as $j \to \infty$. In fact, for j of the form $2^{k+1} + 1$, the above halving propery yields the estimate

$$\left| R \backslash \left(\cup_{i=1}^j A_i \right) \right| \leq C|R|2^{-\ell}$$

where C does not depend on the specific rectangle. We also note that

$$\Delta \left(R, \cup_{i=1}^j A_i \right) \leq \sqrt{1 - 2|R \backslash (\cup_i A_i)|/|R|}$$

is valid as soon as $|R \backslash (\cup_i A_i)| < |R|/10$. Putting these observations together, $\Delta \left(R, \cup_{i=1}^j A_i \right) \to 0$ with increasing j in a fashion independent of R. $\qquad \square$

The axoids in our dissection are, in general, nondyadic. However, each one can be accurately approximated using finite unions of subpixel-resolved dyadic axoids, derived by the same sort of extension procedure developed for intervals and rectangles.

In detail, this works as follows. Given an ultimate error tolerance η, we can choose j such that $\gamma_j^* \leq \eta$. Then for any tilted rectangle R, the j-term approximation by general axoids in $\mathbb{A}$ yields approximation error in Δ metric $\leq \eta$. The rectangles in

$\mathbb{R}_n^2$ have side lengths which are bounded below by $1/n$, and so the general axoids arising in the first j terms of each such dissection have a size bounded below by $1/(c_j n)$. Using this bound, and remembering the comment immediately following Lemma 7.1, we can choose both a subpixel resolution level λ_j and an extension level ℓ_j, independently of n, so that every general axoid arising in this construction can be approximated within accuracy η as a union of ℓ_j-level extensions of dyadic axoids, i.e., using $\mathbb{A}_{n,\lambda_j,\ell_j}$. Invoking again Lemma 5.4, we get [FA-1], for $\mathcal{G}_{n,r} \equiv \mathbb{X}_{n,\lambda_j}$, and $\epsilon_{\ell,\lambda_j} \leq (1+\sqrt{3})\eta$.

Invoking the fast axoid transform gives [FA-2].

For [FA-3], just use the fact that the approximation of axoids by ℓ-level extensions of dyadic axoids is essentially gotten by a (slanted) product of ℓ-level extensions of intervals.

C. Tilted Ellipsoids

Consider now the class of tilted ellipsoids, whose axes need not be parallel to the coordinate axes. We consider the collection of all ellipsoids where

- the center of the ellipsoid is at a pixel corner,
- both the width and length of the ellipsoid are an integral multiple of $1/n$.
- the major axis is oriented at an integral multiple of $2\pi/n^2$,
- the interior of the rectangle does not intersect the boundary of $[0,1]^2$.

We have the following results.

Theorem 7.7: For detecting ellipsoids, the optimal detection threshold $\sim \sqrt{2\log(n^2)}$, and GLRT is near-optimal.

Theorem 7.8: For detecting ellipsoids, there is a near-optimal detector which runs in $O(n^{2+\eta})$ flops for each $\eta > 0$.

The proofs are analogous to those in the tilted rectangle case:

- the dyadic axoids again play the role of generators;
- minimal circumscribing ellipsoids play the role of minimal circumscribing rectangles.

The underlying idea is that ellipses can be obtained from disks by rotation and dilation. While for disks we were able to use axis-aligned rectangles as generators, this would suggest using tilted rectangles as generators. However, as tilted rectangles can themselves be dissected into axoids, we simply use dyadic axoids for our generators in [FA-1].

VIII. DETECTION OF OTHER GEOMETRIC SHAPES

Various extensions of these results seem within reach. We focus on those in the imaging setting of two-dimensional, noisy data.

Continuous Classes: We have focused in this paper on detecting members of finite classes $\mathcal{S}_n$—e.g., $O(n^4)$ lines, or $O(n^4)$ rectangles. However, similar results should hold for certain continuum classes; e.g., let $\mathcal{S}_n$ be the collection of all rectangles with endpoints in the continuum square $[0,1]^2$ and with endpoints separated by at least $1/n$ distance; or let $\mathcal{S}_n$ be the collection of all disks with centers in $[0,1]^2$ and radius at least $1/n$. Indeed, arguing exactly as in Section VII-B1, where the class $\mathbb{A}_n^*$ was mentioned, shows that the corresponding classes of detector arrays ξ_S are compact for the metric δ, and

there are ϵ_ℓ-nets for such classes with cardinality $\leq C_\ell \cdot n^a$, where, crucially, C_ℓ does not depend on n. If so, the optimal detection threshold in the continuum variant of the rectangle and disk problems will be $\sim \sqrt{2\log(n^2)}$, and algorithmic complexity $O(n^{2+\eta})$ for each $\eta > 0$.

Pose/Perspective Families: We have also focused in this paper on detecting members of specific classes of objects, i.e., disks, rectangles, and so on. In these cases, the objects are "directly observed." However, for imaging applications, one is dealing with objects in the three-dimensional world which are rendered into flat two-dimensional imagery according to their pose and to the details of perspective projection. Consequently, the data reflects an intervening variable, the pose and perspective, which are not ordinarily known to the viewer. However, it seems clear that the ideas of this paper carry over to this setting. In effect, we are attempting to detect a class

$$\mathcal{S}_n = \{S_{\theta,K} : K \in \mathcal{K}_n, \theta \in \Theta\}$$

in which the objects of interest K run through a certain set and the pose and perspectives run through another set, and $S_{\theta,K}$ is the net effect on the imaging array of imaging an object at given pose and perspective. Results completely parallel to the existing results should hold in this setting as well. All that is needed is that $\mathcal{S}_n$ be compact for the δ metric, with appropriate cardinality estimates as discussed earlier. It will then follow that for simple object classes (rectangular solids, ellipsoidal solids) and standard pose/perspective models, the optimal detection threshold from two-dimensional imaging will be $\sim \sqrt{2\log(n^2)}$, and the algorithmic complexity will be $O(n^{2+\eta})$ for each $\eta > 0$.

Articulation Families of Composite Objects: We have also focused in this paper on detecting members of specific classes of simple objects, i.e., disks, rectangles, and so on. What about composite objects made out of simple objects? For example, consider a model "hand" with five "fingers" formed by attaching five ellipsoids to a circle, and which is allowed to articulate the fingers by maneuvering them more or less independently. To each articulation of this composite "hand" there is a set of five parameters, specifying the positioning of the fingers. Consequently, an image of the hand reflects an intervening variable, the articulation, not ordinarily known to the viewer. However, it seems clear that the ideas of this paper carry over to this setting. In effect, we are attempting to detect a class

$$\mathcal{S}_n = \{S_\theta : \theta \in \Theta\}$$

in which the objects of interest are articulated according to the parameter θ as it runs through a certain set. Results completely parallel to the above results ought to hold in this setting as well. All that is needed is that $\mathcal{S}_n$ be compact for the δ metric, with appropriate cardinality estimates as discussed earlier. It will then follow that for simple parameterized objects such as the hand model, the optimal detection threshold is asymptotically $\sim \sqrt{2\log(n^2)}$, and the algorithmic complexity of a near-optimal detector will be $O(n^{2+\eta})$ for each $\eta > 0$.

Nonparametric Object Classes: We have also focused in this paper on detecting members of specific classes of simple objects, i.e., disks, rectangles, and so on, which are parameterized by a few parameters. It is possible to consider nonparametric
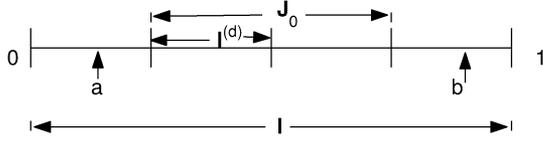
Fig. 7. A case where $I^{(d)}$ and its nonsibling neighbor together form the base J_0.

Fig. 8. The supposed case where a and b bracket more than two consecutive equal-length maximal dyadic intervals. Note that two of the middle three intervals must together form a larger dyadic interval, which contradicts the assumption of maximality.

classes as well, and obtain similar results. For example consider "blobs," which are formally defined as objects with C^2 boundaries with inscribed and circumscribed disks differing by a factor of at most 4, say, and with curvature bounded by B, say. There is no finite-dimensional parametrization of this class. However, the induced class of pixel arrays ξ_S is compact for the δ-distance, and obeys the familiar cardinality estimates. One can check that for the simple blobs model, the optimal detection threshold is asymptotically $\sim \sqrt{2\log(n^2)}$, and the algorithmic complexity of a near-optimal detector will be $O(n^{2+\eta})$ for each $\eta > 0$.

Other Data Types: Obviously, one can ask all the same questions for non-Gaussian data; e.g., a Poisson random field with elevated mean in an subset, or a Poisson random scatter with elevated intensity in a subset. Presumably, very similar results are possible for the limit of increasing total intensity.

APPENDIX I
PROOF OF LEMMA 2.6

Proof: To see (2.2), note that a given dyadic interval can either form a base alone, or perhaps together with its nonsibling dyadic neighbor. The base can be extended four ways at each generation (extending or not at each end). Hence,

$$\#\mathcal{J}_\ell[I] \leq 2 \cdot 4^\ell.$$

There are $2n$ dyadic intervals in $\{0, \dots, n-1\}$, so

$$\#\mathcal{J}_{n,\ell} \leq n4^{\ell+1}.$$

To see (2.3) requires a more complex argument. Given $I = \{a, \dots, b\}$, let $I^{(d)} \subset I$ be a maximal inscribed dyadic interval; $|I^{(d)}| > 1/4 \, |I|$. We now assume $a, b \notin I^{(d)}$, the other cases being handled similarly. Now we claim that the set $\mathcal{J}_\ell[I^{(d)}]$ contains an extension of $I^{(d)}$ satisfying the required inequality (2.3) for ρ. To build this extension, we first ask whether the neighboring dyadic intervals of $I^{(d)}$ having the same length contain a and b. If this is the case, let $J_0 = I^{(d)}$. If this is not the case, the situation is as in Fig. 7. There are four consecutive dyadic intervals of equal length, the first and last contain a and b, respectively. $I^{(d)}$ is one of the two middle ones. In that case, the base J_0 consists of the two middle most intervals, i.e., $I^{(d)}$ and its (nonfraternal) neighbor (nonfraternal since $I^{(d)}$ is maximal). To see that no more than four intervals can occur in a row, inspect Fig. 8.

Let J_0^a denote the leftmost of the dyadic intervals in the group of three or four equal-length dyadic intervals mentioned in the previous paragraph, and let J_0^b denote the rightmost; thus, $a \in J_0^a$ and $b \in J_0^b$. We now show how to construct an extension $J \in \mathcal{J}_\ell[I^{(d)}]$ which misses at most $2^{-\ell}|I^{(d)}|$ of the points in I at each end.
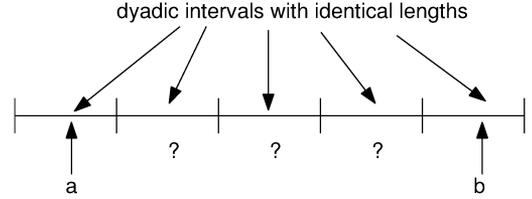
At stage 1, if a lies in the left half of J_0^a, append the right half of J_0^a to J_0; otherwise, do not append. Also, if b lies in the right half of J_0^b, append the left half of J_0^b; otherwise, do not append. Call the result of the (possible) left and right extensions J_1.

Define J_1^a to be whichever half of J_0^a contains a, J_1^b to be whichever half of J_0^b contains b, and now continue with J_0 replaced by J_1, etc.

Now, after each step $j = 0, \dots, l$

$$J_j^a \cup J_j \cup J_j^b \subset I$$

and evidently

$$\left|J_j^a\right| = 2^{-j}|J_0^a| = 2^{-j}|I^{(d)}|;$$

likewise for $\left|J_j^b\right|$. Hence,

$$|I| \leq |J_j| + 2 \cdot 2^{-j}|I^{(d)}|.$$

Now, as $J_\ell \subset I$

$$\rho(I, J_\ell) = \sqrt{|J_\ell|/|I|}$$

and $|J_\ell| \geq |I^{(d)}|$. Hence,

$$\rho(I, J_\ell) \geq \sqrt{\frac{|J_\ell|}{|J_\ell| + 2 \cdot 2^{-l}|J_\ell|}} = \sqrt{\frac{1}{1 + 2^{-\ell+1}}}. \qquad \square$$

APPENDIX II
PROOF OF THEOREM 4.1

Upper Bound: Fix $\eta > 0$. We want to prove that

$$P\left\{ X^*[\mathcal{S}_n] > \sqrt{2(1+\eta)\log(n^a)} \right\} \to 0, \qquad n \to \infty.$$

Start with

$$X^*[\mathcal{S}_n] \leq X^*[\mathcal{N}_{n,\ell}] + \Delta_{n,\ell},$$

where

$$\Delta_{n,\ell} = \max\{X[S] - X[T] : S, T \in \mathcal{S}_n \text{ and } \delta(S,T) \leq \epsilon_\ell\}.$$

Apply Lemma 2.9 to get

$$P\left\{ X^*[\mathcal{N}_{n,\ell}] > \sqrt{2\log(\#\mathcal{N}_{n,\ell})} \right\} \to 0 \quad \text{and}$$

$$P\left\{ \Delta_{n,\ell} > \sqrt{2}\epsilon_\ell \sqrt{2\log((\#\mathcal{S}_n)^2)} \right\} \to 0$$

as $n \to \infty$, uniformly in ℓ. Furthermore

$$\log(\#\mathcal{N}_{n,\ell}) = \log(n^a) + \log(C_\ell) + c \quad \text{and}$$

$$\log(\#\Delta_{n,\ell}) = \log(n^{2b}) + c.$$

Conclude by making $\ell = \ell_n \to \infty$, growing slowly enough so that $\log(C_\ell) \ll \log(n)$.

Lower Bound: Fix $\eta > 0$. We want to prove that

$$P\left\{X^*[\mathcal{S}_n] > \sqrt{2(1 - \eta)\log(n^a)}\right\} \to 1, \qquad n \to \infty.$$

We now employ a converse to Lemma 2.9.

Let $w_1, \ldots, w_k$ be independent $\mathcal{N}\left(0, \sigma_i^2\right)$ with $\sigma_i^2 \geq \tau^2 > 0$. For each $\eta > 0$

$$P\left\{\max_i w_i > (1 - \eta)\tau\sqrt{2\log(k)}\right\} \to 1, \qquad k \to \infty.$$

Apply this result together with [DT-3] to $w_i = X^*[S_i]$ where S_i is an enumeration of $\mathcal{P}_n$.

Near-Optimality: We want to prove that the GLRT is near-optimal, in the sense that, for $A_n \leq \sqrt{2(1 - \eta)\log(n^a)}$ for some $\eta > 0$, every sequence of tests is asymptotically powerless. The proof is similar to that of Theorem 2.3—consider a uniform prior on $\mathcal{P}_n$ and we are again in the context of finding a needle-in-a-haystack.

APPENDIX III
PROOF OF THEOREM 4.2

Fix $\eta > 0$. Choose ℓ, λ such that

$$(1 - (\epsilon_\ell + \epsilon_{\ell,\lambda})^2)\sqrt{\frac{1 + \eta}{1 + \eta/2}} \geq 1.$$

The algorithm has three stages.

1) *Fast $\mathcal{G}_{n,r}$-Transform.* Compute $X[G]$ for every $G \in \mathcal{G}_{n,r}$.
2) *Approximation to $\mathcal{N}_{n,\ell}$.* For each $E \in \mathcal{N}_{n,\ell}$, enumerate $\{G_i : i = 1, \ldots, k_{\ell,r}\}$ elements of $\mathcal{G}_{n,r}$ such that $\delta(E, \cup_i G_i) \leq \epsilon_{\ell,r}$. Approximate $X[E]$ by

$$X[\cup_i G_i] = \sum_i \frac{N(G_i)}{N(E)} X[G_i].$$

3) *Decide.* If the maximum among $X[E]$, $E \in \mathcal{N}_{n,\ell}$ at stage 2 exceeds $\sqrt{2(1 + \eta/3)\log(n^e)}$, reject H_0.

In the following we establish in turn each of the main claims about this procedure.

The algorithm runs in $O(n^{e+\eta})$ flops. By [FA-2], stage 1 requires only $C_{\eta,r}n^{e+\eta} = O(n^{e+\eta})$ flops. There are $C_{\ell,\eta/2}n^{e+\eta/2}$ elements in $\mathcal{N}_{n,\ell}$. For $E \in \mathcal{N}_{n,\ell}$, the enumeration takes $C_{\ell,r} = O(1)$ flops and the sum contains at most $k_{\ell,r}$ terms, which have been computed in stage 1. Hence, stage 2 requires only $O(n^{e+\eta/2})$ flops. As we take the maximum over $O(n^{e+\eta/2})$ numbers, stage 3 can be done in $O(n^{e+\eta/2}\log n) = O(n^{e+\eta})$ flops.

The probability of Type I error tends to zero as n increases. Suppose H_0 holds, so the data are pure noise. Just notice that the maximum over $X[E]$, $E \in \mathcal{N}_{n,\ell}$ is bounded by the GLRT and conclude by invoking Theorem 4.1.

The probability of Type II error tends to zero as n increases. Let the data be distributed as $\mathcal{N}(A\xi_S, I)$, with

$$A = \sqrt{2(1 + \eta)\log(n^e)} \quad \text{and} \quad S \in \mathcal{S}_n.$$

The algorithm accepts H_0 only if the maximum over $X[E]$, $E \in \mathcal{N}_{n,\ell}$ does not exceed $\sqrt{2(1 + \eta/3)\log(n^e)}$.

By [DT-2], there is $E \in \mathcal{N}_{n,\ell}$ such that $\delta(S, E) \leq \epsilon_\ell$. By [FA-1], there are $G_1, \ldots, G_k$ such that $\delta(E, \cup_i G_i) \leq \epsilon_{\ell,r}$. Applying the triangle inequality for the δ-metric, we have

$$\delta(S, \cup_i G_i) \leq \epsilon_\ell + \epsilon_{\ell,r}$$

so

$$\rho(S, \cup_i G_i) \geq 1 - (\epsilon_\ell + \epsilon_{\ell,r})^2.$$

Hence,

$$\begin{aligned} E(X[\cup_i G_i]) &= A\rho(S, \cup_i G_i) \\ &\geq (1 - (\epsilon_\ell + \epsilon_{\ell,r})^2)\sqrt{2(1 + \eta)\log(n^e)} \\ &\geq \sqrt{2(1 + \eta/2)\log(n^e)} \end{aligned}$$

and in the same way, we conclude that

$$P\left\{X[\cup_i G_i] > \sqrt{2(1 + \eta/3)\log(n^e)}\right\} \to 1, \qquad n \to \infty.$$

No near-optimal algorithm runs in $O(n^{e-\eta})$ flops for any $\eta > 0$. Since the sets in $\mathcal{P}_n$ are disjoint, it will be necessary to test every one of them in order to detect at optimal sensitivity. Otherwise, a clever opponent knowing the algorithm can place all prior probability on the untested intervals. But if you test every one paying even one flop per test, you have spent $\geq C_\eta n^{e-\eta}$ flops.

APPENDIX IV
PROOF OF LEMMA 5.4

We will need the following result.

Lemma 4.1: For R and S two sets of equal Hausdorff dimension, with $\Delta(R, S) \leq \eta < 1$, we have

$$(1 - \eta^2)^2 \leq \frac{|R|}{|S|} \leq \frac{1}{(1 - \eta^2)^2}.$$

Proof: Call $R = (R \setminus S) \cup (R \cap S) = R_0 \cup T$ and $S = (S \setminus R) \cup (R \cap S) = S_0 \cup T$, say. Then

$$\begin{aligned} \eta^2 &\geq \Delta(R, S)^2 \\ &= 1 - \frac{|T|}{\sqrt{|R_0| + |T|}\sqrt{|S_0| + |T|}} \\ &\geq 1 - \sqrt{\frac{1}{\sqrt{|R_0|/|T| + 1}}} \end{aligned}$$

using $|S_0| \geq 0$. Hence,

$$\frac{|R_0|}{|T|} + 1 \leq \frac{1}{(1 - \eta^2)^2}.$$

Hence,

$$\frac{|R|}{|S|} = \frac{|R_0| + |T|}{|S_0| + |T|} \leq \frac{|R_0|}{|T|} + 1 \leq \frac{1}{(1 - \eta^2)^2}.$$

The lower bound follows by symmetry. $\qquad\square$

We can now proceed with the proof of Lemma 5.4. By definition

$$\Delta(\cup_i R_i, \cup_i S_i)^2 = 1 - \frac{|(\cup_i R_i) \cap (\cup_i S_i)|}{\sqrt{|\cup_i R_i|}\sqrt{|\cup_i S_i|}}.$$

By Property 1, $|\cup_i R_i| = \sum_i |R_i|$ and $|\cup_i S_i| = \sum_i |S_i|$, so the denominator in the fraction above is equal to

$$\sqrt{\sum_i |R_i|}\sqrt{\sum_i |S_i|}.$$

Using Lemma 4.1, we upper-bound this by $\sum_i |R_i|/(1 - \eta^2)$.

The numerator in the same fraction is bounded below by $\sum_i |R_i \cap S_i|$. By Property 2

$$|R_i \cap S_i| \geq (1 - \eta^2)\sqrt{|R_i||S_i|}.$$

Using the Lemma 4.1, we lower-bound this by $(1 - \eta^2)^2 |R_i|$.

Using the upper bound on the denominator and the lower bound on the numerator, we get

$$\Delta^2(\cup_i R_i, \cup_i S_i) \leq 1 - (1 - \eta^2)^3.$$

The term on the right is smaller than $3\eta^2$ for η small enough. $\square$

APPENDIX V
PROOF OF LEMMA 6.1

Without loss of generality, suppose that S makes an angle between $0°$ and $45°$ with the x-axis, as other cases may be reduced to this by symmetry. For clarity, rescale the pixels to side length 1. The new Hausdorff distance is h. Assume $h < 1/4$. Consider two subcases. In the first, the slope s of the line is less than $\sqrt{h}$. In the second, $1 \geq s \geq \sqrt{h}$.

Case 1: $0 < s < \sqrt{h}$. Because S joins pixel corners, S must cross at least $1/s$ columns $\mathrm{col}(k) = \{(x, y) : k \leq x < k+1\}$; in the process it will increase in altitude by at least 1. In traversing these columns, two sorts of things can happen: S can be near an integral value of y—i.e., within distance $2h$—or not near to such a value.

When $S \cap \mathrm{col}(k)$ does not lie within $2h$ of an integral value of y, it also does not cross an integral value of y, and neither does the approximating chain C, since the endpoints can be at most $h(1 + h)^{1/2} < 2h$ apart. Let G be the set of such columns which intersect S, and B the remaining ones. (Mnemonics for G and B: "good" and "bad.")

Now as S connects pixel corners and is not horizontal, it traverses $\Delta y/s$ columns, where Δy is the integral difference in altitude between beginning and end. The number of columns in B is at most

$$\Delta y \cdot (4h/s + 2)$$

since in traversing a column, the integral coordinate increases by s, and this can happen only $4h/s$ times while s stays within $2h$ distance of each integral value. Thus,

$$\frac{\#G}{\#(G \cup B)} \geq \frac{\Delta y \cdot (1/s - 4h/s - 2)}{\Delta y(1/s)}$$
$$= 1 - 4h - 2s$$
$$\geq 1 - 4h - 2\sqrt{h}.$$

We now argue that throughout G, the arc lengths $l_S(p) = |S \cap p|$ and $l_C(p) = |C \cap p|$ are close. Let γ run through $G \cup B$.

Write $p \frown \gamma$ if pixel p belongs to column γ and intersects S. If $\gamma \in G$, then

$$l_C(p) \geq l_S(p) - 2h(1 + h)^{1/2} \geq l_S(p) - 3h$$

because $C \cap p$ and $S \cap p$ have endpoints within $h \cdot (1 + h)^{1/2}$ of each other, and the triangle inequality. Hence,

$$\sum_{\gamma \in G} \sum_{p \frown \gamma} l_S(p)l_C(p) \geq \sum_{\gamma \in G} \sum_{p \frown \gamma} l_S^2(p) - 3hl_S(p).$$

As $p \frown \gamma \in G$, $l_S(p) = \sqrt{1 + s^2} \geq 1$, so $l_S^2(p) \geq l_S(p)$, so $l_S^2(p) - 3hl_S(p) \geq l_S^2(p)(1 - 3h)$. Hence,

$$\sum_{\gamma \in G} \sum_{p \frown \gamma} l_S(p)l_C(p) \geq (1 + s^2) \cdot \#G \cdot (1 - 3h).$$

Finally, a similar inequality holds for $\gamma \in G \cup B$

$$\sum_{\gamma \in G \cup B} \sum_{p \frown \gamma} l_S(p)l_C(p) \geq (1 + s^2) \cdot \#(G \cup B) \cdot (1 - 3h).$$

On the other hand

$$\sum_{\gamma \in G \cup B} \sum_{p \frown \gamma} l_S^2(p) \leq \#(G \cup B) \cdot (1 + s^2)$$

while, if $\gamma \in G \cup B$

$$\sum_{p \frown \gamma} l_C(p) \leq [(1 + s^2)^{1/2} + 2h(1 + s^2)^{1/2}]$$
$$\leq (1 + h)^{1/2}(1 + 2h).$$

Hence, using $(a^2 + b^2) \leq (a + b)^2$

$$\sum_{p \frown \gamma} l_C^2(p) \leq (1 + h)(1 + 2h)^2.$$

Conclude that

$$\left(\sum_{p \cap S \neq \emptyset} l_S^2(p)\right)\left(\sum_{p \cap C \neq \emptyset} l_C^2(p)\right) \leq [\#(G \cup B)(1+h)(1+2h)]^2.$$

Hence,

$$\frac{\sum_p l_S(p)l_C(p)}{\left(\sum_p l_S^2(p) \cdot \sum_p l_C^2(p)\right)^{1/2}} \geq \frac{\#G \cdot (1 - 3h)}{\#(G \cup B)(1 + h)(1 + 2h)}$$
$$\geq \frac{(1 - 4h - 2\sqrt{h})(1 - 2h)}{1 + 2h}$$
$$\geq 1 - C_1 \cdot \sqrt{h}, \qquad \forall\, h < 1/4;$$

here C_1 is a constant which can be made explicit.

Case 2: $\sqrt{h} \leq s < 1$. The subcase $\sqrt{h} \leq s < 1/2$ being slightly simpler, we begin there, indicating later the modifications for $1/2 \leq s \leq 1$.

When $1/2 > s$, as S crosses $\mathrm{col}(k)$ it intersects at most two pixels, and the same is true of the approximant C; moreover there is in each column a single pixel or pair of pixels in $\mathrm{col}(k)$ that covers $S \cap \mathrm{col}(k)$ and $C \cap \mathrm{col}(k)$. We let p_1 and p_2 be the
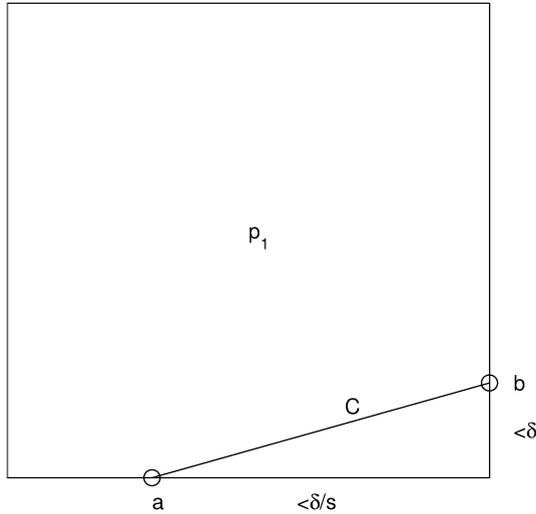
Fig. 9. The case in which $l_C(p_1) = |ab|$ is bounded by $\sqrt{h^2 + (h/s)^2}$. Note that S does not intersect with pixel p_1.
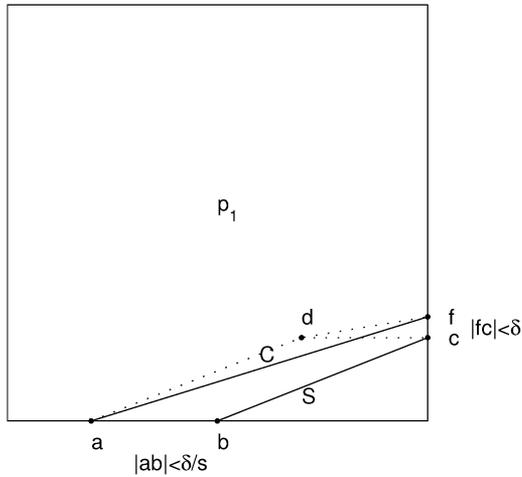

Fig. 10. The case in which $|l_C(p_1) - l_S(p_1)| \leq |df| < \sqrt{h^2 + (h/s)^2}$. Note that $abcd$ forms a parallelogram.

pixels in question, and let $l_S(p_i)$ be the arclength in the indicated pixels (with $l_S(p_i) = 0$ if $S \cap p_i = \emptyset$ etc.).

The key insight is expressible as

$$|l_S(p_i) - l_C(p_i)| < 2\sqrt{h}, \qquad i = 1, 2. \qquad (V.1)$$

This crucial inequality is not what one would naively expect; the $2\sqrt{h}$ would seem, at first glance, improvable to $2h$, since S and C are Hausdorff distance h apart. To see (V.1), note that it is possible for a given pixel p_1, say, to intersect C but not S. In that case $l_S(p_1) = 0$ while $l_C(p_1) > 0$. From Fig. 9, it is easy to see that if C is at Hausdorff distance h from S while $S \cap p_1 = \emptyset$, then we can have

$$l_C(p_1) = \sqrt{h^2(1 + s^2) + h^2(1 + s^2)/s^2}$$

whence, from $s \geq \sqrt{h}$, we get, for h small enough

$$l_C(p_1) < h\sqrt{2 + s^2 + 1/s^2} \leq 2\sqrt{h}.$$

In case p_1 intersects both C and S, although the two line segments inside p_1 are at Hausdorff distance h, their endpoints are controlled within distance $h\sqrt{1 + s^2}/s \leq 2\sqrt{h}$; see Fig. 10. Hence, (V.1) follows.

Arguing from (V.1), we obtain

$$\sum_{i=1}^{2} l_C(p_i) l_S(p_i) \geq \sum_{i=1}^{2} l_S(p_i)(l_S(p_i) - 2\sqrt{h})_+$$

$$\geq \sum_{i=1}^{2} l_S^2(p_i) - 2\sqrt{h} l_S(p_i)$$

$$\geq \sum_{i=1}^{2} l_S^2(p_i)(1 - 4\sqrt{h})$$

where we used

$$\sum_{i=1}^{2} l_S(p_i) < \left(\sum_{i=1}^{2} l_S(p_i) \right)^2 \leq 2 \sum_{i=1}^{2} l_S^2(p_i).$$

Similarly, for h small enough

$$\sum_{i=1}^{2} l_C^2(p_i) - \sum_{i=1}^{2} l_S^2(p_i)$$

$$= \sum_{i=1}^{2} (l_C(p_i) - l_S(p_i))(l_C(p_i) + l_S(p_i))$$

$$\leq 2\sqrt{2} \cdot 2 \cdot 2\sqrt{h} < 12\sqrt{h}$$

using $l_C(p_i) + l_S(p_i) \leq 2\sqrt{2}$ and (V.1). Therefore, since $\sum_{i=1}^{2} l_S^2(p_i) > 1/2$

$$\sum_{i=1}^{2} l_C^2(p_i) \leq \sum_{i=1}^{2} l_S^2(p_i) \cdot (1 + 24\sqrt{h}).$$

Hence, we conclude that, with γ running over columns $\mathrm{col}(k)$ meeting S

$$\frac{\sum_p l_S(p) l_C(p)}{\left(\sum_p l_S^2(p) \cdot \sum_p l_C^2(p) \right)^{1/2}}$$

$$\geq \frac{\sum_\gamma \sum_{p \rightsquigarrow \gamma} l_S^2(p)(1 - 4\sqrt{h})}{\left(\sum_\gamma \sum_{p \rightsquigarrow \gamma} l_S^2(p) \right)^{1/2} \left(\sum_\gamma \sum_{p \rightsquigarrow \gamma} l_S^2(p)(1 + 24\sqrt{h}) \right)^{1/2}}$$

$$= \frac{1 - 4\sqrt{h}}{(1 + 24\sqrt{h})^{1/2}} \geq 1 - C_2 \cdot \sqrt{h}, \qquad h < 1/4;$$

here C_2 is again a constant which can be made explicit.

For $1/2 < s < 1$, there is the possibility of S touching pixels p_1 and p_2 and C touching p_2 and p_3, where p_1, p_2, p_3 are distinct pixels belonging to the same column. Equation (V.1) needs to be modified accordingly, but holds in a similar form. The rest of the argument then proceeds in the same way, yielding a constant C_3. The proof goes through with $K = \max_i C_i$. □

APPENDIX VI
PROOF OF LEMMA 6.2

Proof: Consider a line segment S. Suppose it is basically horizontal (angle with x-axis not exceeding $45°$). Let I denote its projection onto the x-axis. There is a dyadic interval J in the projection I such that $\ell(J)/\ell(I) > 1/4$. Project interval J back onto S, and denote it by T. Obviously, $\ell(T)/\ell(S) > 1/4$. Note that line segment T is either one (or the union of two) continuous beamlet(s). If $T = B_1 \cup B_2$, where B_1 and B_2 are two beamlets, then either $\ell(B_1)/\ell(S)$ or $\ell(B_2)/\ell(S)$ exceeds $1/8$. □

APPENDIX VII
PROOF OF LEMMA 6.3

A. Inequality (6.1)

Let S_1 be a line segment joining two pixel corners, and s its slope. Without loss of generality, we assume $0 \leq s \leq 1$. We first prove that

$$\frac{1}{2} \frac{\sqrt{1+s^2}}{n} \leq \frac{N(S_1)^2}{\ell(S_1)} \leq \frac{\sqrt{1+s^2}}{n}.$$

Start with $N(S_1)^2 = \sum_p \ell(S_1 \cap p)^2$. Now, S_1 crosses exactly $n\,\ell(S_1)/\sqrt{1+s^2}$ columns. In each column, S_1 touches one or two pixels. In the first case, $\ell(S_1 \cap p)^2 = (1+s^2)/n^2$; in the second case, $(\ell(S_1 \cap p_1) + \ell(S_1 \cap p_2))^2 = (1+s^2)/n^2$, and so $\ell(S_1 \cap p_1)^2 + \ell(S_1 \cap p_2)^2 \leq (1+s^2)/n^2$. Hence, summing over columns gives

$$N(S_1)^2 \leq (n\,\ell(S_1)/\sqrt{1+s^2}) \times (1+s^2)/n^2$$

which is the upper bound. For the lower bound, use the Cauchy–Schwartz inequality

$$\ell(S_1) = \sum_p \ell(S_1 \cap p) \leq \left(2\,n\,\ell(S_1)/\sqrt{1+s^2}\right)^{1/2} N(S_1).$$

Now, if $S_1 \subset S_2$ are two line segments joining pixel columns

$$\rho^2(S_1, S_2) = \frac{N^2(S_1)}{N^2(S_2)} = \frac{(n/\sqrt{1+s^2})\,N^2(S_1)}{(n/\sqrt{1+s^2})\,N^2(S_2)} \geq \frac{\ell(S_1)/2}{\ell(S_2)}.$$

B. Inequality (6.2)

Again, consider $S_1 \subset S_2$ are two line segments, joining pixel corners. Then, $S_2 = S_1 \cup A_1 \cup A_2$, a disjoint union of line segments with endpoints at pixel corners. Let s be their common slope, and we assume $0 \leq s \leq 1$. Compute

$$\begin{aligned}
N(S_2)^2 &= N(S_1)^2 + N(A_1)^2 + N(A_2)^2 \\
&\leq N(S_1)^2 + \ell(A_1)\sqrt{1+s^2}/n + \ell(A_2)\sqrt{1+s^2}/n \\
&= N(S_1)^2 + (\ell(S_2) - \ell(S_1))\sqrt{1+s^2}/n.
\end{aligned}$$

Further

$$\begin{aligned}
\rho(S_1, S_2) &= \frac{N(S_1)}{N(S_2)} \\
&\geq \frac{N(S_2)^2 - (\ell(S_2) - \ell(S_1))\sqrt{1+s^2}/n}{N(S_2)} \\
&= \sqrt{1 - \frac{\sqrt{1+s^2}}{n} \frac{\ell(S_2)}{N(S_2)^2} \left(1 - \frac{\ell(S_1)}{\ell(S_2)}\right)} \\
&\geq \sqrt{1 - 2\left(1 - \frac{\ell(S_1)}{\ell(S_2)}\right)}.
\end{aligned}$$

APPENDIX VIII
PROOF OF LEMMA 6.5

Start with (6.3). The cardinality of the set of (v_0, v_1) arising in the construction, with $|v_{0,x} - v_{1,x}| = 2^{-j}$ and $|v_{0,y} - v_{1,y}| \leq 2^{-j}$, is bounded by $2\nu^{-2}$. The number of extensions of each line segment with endpoints (v_0, v_1) through level ℓ is at most 4^ℓ.

Now, consider (6.4). We again suppose that S makes an angle between $0°$ and $45°$ with the x-axis, as other cases may be reduced to this by symmetry.

Let I be the projection of S on the x-axis, and let $I^{(d)}$ denote the maximal dyadic subinterval. Let $I_\ell^{(d)}$ denote the best ℓ-level extension of $I^{(d)}$, and project vertically $I_\ell^{(d)}$ onto S. There is a line segment $L_\ell \in \mathbb{L}_{n,\lambda,\ell}$ which lies within Hausdorff distance $\nu = \lambda/n$ from $\tilde{S} = S \cap \mathrm{Strip}(I_\ell^{(d)})$. Let $C \in \mathbb{B}_{n,\lambda,\ell}$ be a beamlet chain which lies within Hausdorff distance $\nu = \lambda/n$ from L_ℓ. We have

$$\mathrm{Hausdorff}(C, \tilde{S}) \leq 2\nu.$$

By Lemma 6.3, $\rho(S, \tilde{S}) \geq \sqrt{1 - 2^{-\ell+2}}$. This, together with Lemma 6.4, yields

$$\rho(S, C) \geq \rho(\tilde{S}, C) + O(2^{-\ell/2}).$$

Now apply Lemma 6.1, noting that $\mathrm{Hausdorff}(C, \tilde{S}) \leq 2\nu$, and obtain

$$\rho(\tilde{S}, C) \geq 1 - K'\sqrt{\nu}$$

for a constant K', and the proof is complete.

REFERENCES

[1] Beamlet Software. [Online]. Available: http://www.beamlab.org
[2] "Expanding the vision of sensor materials," in *National Research Council, Committee on New Sensor Technologies, Materials, and Applications*. Washington, DC: National Academies Press, 1995.
[3] Multidisciplinary Workshop on Automatic Particle Selection for Cryo-Electron Microscopy. National Resource for Automated Molecular Microscopy (NRAMM), The Scripps Research Institute (TSRI), La Jolla, CA, Apr. 2003.
[4] H. Abramowicz, D. Horn, U. Naftali, and C. Sahar-Pikielny, "An orientation-selective neural network for pattern identification in particle detectors," in *Advances in Neural Information Processing Systems*, M. C. Mozer, M. J. Jordan, and T. Petsche, Eds. Cambridge, MA: MIT Press, 1997, vol. 9, pp. 925–931.
[5] P. Agouris, A. Stefanidis, and S. Gyftakis, "Differential snakes for change detection in road segments," *Photogrammetric Eng. Remote Sensing*, vol. 67, no. 12, pp. 1391–1399, 2001.
[6] D. Amaratunga and J. Cabrera, "Mining data to find subsets of high activity," *J. Statist. Plann. Inference*, vol. 122, no. 1–2, pp. 23–41, 2004.
[7] Y. Amit, D. Geman, and X. Fan, "A coarse-to-fine strategy for multiclass shape detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 12, pp. 1606–1621, Dec. 2004.
[8] Y. Amit, U. Grenander, and M. Piccioni, "Structural image restoration through deformable templates," *J. Amer. Statist. Assoc.*, vol. 86, no. 414, pp. 376–387, Jun. 1991.
[9] R. Arratia and M. S. Waterman, "The Erdös-Rényi strong law for pattern matching with a given proportion of mismatches," *Ann. Probab.*, vol. 17, no. 3, pp. 1152–1169, 1989.
[10] A. Averbuch, R. R. Coifman, D. L. Donoho, M. Israeli, and J. Waldén, "Fast slant stack: A notion of radon transform for data in a cartesian grid which is rapidly computible, algebraically exact, geometrically faithful and invertible," Dept. Statistics, Stanford Univ., Stanford, CA, 2001.
[11] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. Ser. B.*, vol. 48, no. 3, pp. 259–302, 1986.
[12] J. Besag, P. Green, D. Higdon, and K. Mengersen, "Bayesian computation and stochastic systems (with discussion)," *Statist. Sci.*, vol. 10, pp. 3–66, 1995.
[13] A. Blake and M. Isard, *Active Contours*. Berlin, Germany: Springer-Verlag, 1998.
[14] M. L. Brady, "A fast discrete approximation algorithm for the Radon transform," *SIAM J. Comput.*, vol. 27, no. 1, pp. 107–119, Feb. 1998.

[15] A. Brandt, "Multiscale scientific computation: Review," in *Multiscale and Multiresolution Methods: Theory and Applications*, T. J. Barth, T. F. Chan, and R. Haimes, Eds. Heidelberg, Germany: Springer-Verlag, 2001.

[16] A. Brandt and J. Dym, "Fast calculation of multiple line integrals," *SIAM J. Sci. Comput.*, vol. 20, no. 4, pp. 1417–1429, 1999.

[17] M. V. Burnashev and I. A. Begmatov, "A problem of signal-detection leading to stable-distributions," *Theory Prob. its Applic.*, vol. 35, no. 3, pp. 556–560, 1990.

[18] D. Casasent and A. Q. Ye, "Detection filters and algorithm fusion for ATR," *IEEE Trans. Image Process.*, vol. 6, no. 1, pp. 114–125, Jan. 1997.

[19] H. P. Chan and T. L. Lai, "Asymptotic approximations for error probabilities of sequential or fixed sample size tests in exponential families," *Ann. Statist.*, vol. 28, no. 6, pp. 1638–1669, 2000.

[20] ——, "Saddlepoint approximations and nonlinear boundary crossing probabilities of markov random walks," *Ann. Appl. Prob.*, vol. 13, no. 2, pp. 395–429, 2003.

[21] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, "Deformable templates using large deformation kinematics," *IEEE Trans. Image Process.*, vol. 5, no. 10, pp. 1435–1447, Oct. 1996.

[22] A. C. Copeland, G. Ravichandran, and M. H. Trivedi, "Localized Radon transform-based detection of ship wakes in SAR image," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 1, pp. 35–45, Jan. 1995.

[23] R. W. R. Darling and M. S. Waterman, "Matching rectangles in d dimensions: Algorithms and laws of large numbers," *Adv. Math.*, vol. 55, no. 1, pp. 1–12, 1985.

[24] ——, "Extreme value distribution for the largest cube in a random lattice," *SIAM J. Appl. Math.*, vol. 46, no. 1, pp. 118–132, 1986.

[25] G. David and S. Semmes, *Analysis of and on Uniformly Rectifiable Sets*. Providence, RI: Amer. Math. Soc., 1993, vol. 38, Mathematical Surveys and Monographs.

[26] A. Desolneux, L. Moisan, and J.-M. Morel, "Meaningful alignments," *Int. J. Comp. Vision*, vol. 40, no. 1, pp. 7–23, 2000.

[27] D. L. Donoho, "Wedgelets: Nearly minimax estimation of edges," *Ann. Statist.*, vol. 27, no. 3, pp. 859–897, 1999.

[28] D. L. Donoho and X. Huo, "Beamlet pyramids: A new form of multiresolution analysis, suited for extracting lines, curves, and objects from very noisy image data," *Proc. SPIE*, pt. 1–2, vol. 4119, pp. 434–444, Jul. 2000.

[29] ——, *Beamlets and Multiscale Image Analysis (Lecture Notes in Computational Science and Engineering)*. Berlin, Germany: Springer-Verlag, 2001, vol. 20, , pp. 149–196.

[30] D. L. Donoho and J. Jin, "Higher criticism for detecting sparse heterogeneous mixtures," *Ann. Statist.*, vol. 32, no. 3, pp. 962–994, 2004.

[31] D. L. Donoho and I. M. Johnstone, "Minimax estimation via wavelet shrinkage," *Ann. Statist.*, vol. 26, no. 3, pp. 879–921, 1998.

[32] D. L. Donoho and O. Levi, "Fast X-ray and beamlet transforms for three-dimensional data," in *Modern Signal Processing*, D. Rockmore and D. Healy, Eds. Cambridge, U.K.: Math. Sci. Res. Ins. Publications, Cambridge Univ. Press, 2003.

[33] G. Dror, H. Abramowicz, and D. Horn, "Vertex identification in high energy physics experiments," in *Neural Information Processing Systems (NIPS)*, 1998.

[34] P. Felzenszwalb, "Representation and detection of deformable shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 208–220, Feb. 2005.

[35] F. Fleuret and D. Geman, "Coarse-to-fine face detection," *Int. J. Comp. Vision*, vol. 41, pp. 85–107, 2001.

[36] J. Frank, "Single-particle imaging of macromolecules by cryo-electron microscopy," *Ann. Rev. Biophys. Biomol. Struct.*, vol. 31, pp. 303–319, Oct. 2002.

[37] J. H. Friedman and N. I. Fisher, "Bump hunting in high-dimensional data," *Statist. and Comp.*, vol. 9, no. 2, pp. 123–143, Apr. 1999.

[38] D. Geman and B. Jedynak, "Shape recognition and twenty questions," in *Proc. Reconnaissance des Formes et Intelligence Artificielle (RFIA)*, 1994. (in French).

[39] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[40] J. Glaz and N. Balakrishnan, *Scan Statistics and Applications*. Boston, MA: Birkhauser, 1999.

[41] J. Glaz, J. Naus, and S. Wallenstein, *Scan Statistics*. New York: Springer-Verlag, 2001.

[42] W. A. Götze and H. J. Druckmüller, "A fast digital Radon transform—An efficient means for evaluating the Hough transform," *Pattern Recogn.*, vol. 28, no. 12, pp. 1985–1992, 1995.

[43] U. Grenander, *General Pattern Theory: A Mathematical Study of Regular Structures*. New York: Oxford Univ. Press, 1993.

[44] ——, *Elements of Pattern Theory*. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[45] U. Grenander and M. I. Miller, "Representations of knowledge in complex systems. With discussion and a reply from the authors," *J. Roy. Statist. Soc. Ser. B*, vol. 56, no. 4, pp. 549–603, 1994.

[46] X. Huo, "Multiscale approximation methods (MAME) to locate embedded consecutive subsequences—Its applications in statistical data mining and spatial statistics," *Comp. Industr. Eng.*, vol. 43, no. 4, pp. 703–720, Sep. 2002.

[47] I. A. Ibragimov and R. Z. Khas'minskii, *Statistical Estimation: Asymptotic Theory*. New York: Springer-Verlag, 1981.

[48] I. A. Ibragimov, A. S. Nemirovskii, and R. Z. Khas'minskiii, "Some problems of nonparametric estimation in Gaussian white noise," *Theory Probab. its Applic.*, vol. 31, no. 3, pp. 391–406, 1986.

[49] Y. I. Ingster, "Some problems of hypothesis testing leading to infinitely divisible distributions," *Math. Methods Statist.*, vol. 6, no. 1, pp. 47–69, 1997.

[50] ——, "Minimax detection of a signal for l_n^p-balls," *Math. Methods Statist.*, vol. 7, no. 4, pp. 401–428, 1999.

[51] P. W. Jones, "Rectifiable sets and the traveling salesman problem," *Inventiones Mathematicae*, vol. 102, pp. 1–15, 1990.

[52] M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer-Verlag, 1983.

[53] G. Lerman, "Geometric transcriptions of sets and their applications to data analysis," Ph.D. dissertation, Dept. Math., Yale Univ., New Haven, CT, 2000.

[54] S. Mahadevan and D. P. Casasent, "Detection of triple junction parameters in microscope images," *Proc. SPIE*, vol. 4387, pp. 204–214, 2001.

[55] D. Marr and E. Hildreth, "Theory of edge detection," in *Proc. Roy. Soc. London, Ser. B, Biological Sci.*, vol. 207, Feb. 1980, pp. 187–217.

[56] A. Pievatolo and P. Green, "Boundary detection through dynamic polygons," *J. Roy. Statist. Soc. Ser. B, Stat. Methodol.*, vol. 60, no. 3, pp. 609–626, 1998.

[57] D. Pollard, *Convergence of Stochastic Processes*. Berlin, Germany: Springer-Verlag, 1984.

[58] H. Rue and M. A. Hurn, "Bayesian object identification," *Biometrika*, vol. 86, no. 3, pp. 649–660, 1999.

[59] H. Rue and O. K. Husby, "Identification of partly destroyed objects using deformable templates," *Statist. Comput.*, vol. 8, pp. 221–228, 1998.

[60] H. Rue and A. R. Syversveen, "Bayesian object recognition with Baddeley's delta loss," *Adv. Appl. Probab.*, vol. 30, pp. 64–84, 1998.

[61] E. Sharon, A. Brandt, and R. Basri, "Fast multiscale image segmentation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, 2000, pp. 70–77.

[62] G. R. Shorack and J. A. Wellner, *Empirical Processes With Applications to Statistics*, ser. Series in Probability and Mathematical Statistics. New York: Wiley, 1986.

[63] D. Siegmund and E. S. Venkatraman, "Using the generalized likelihood ratio statistic for sequential detection of a change-point," *Ann. Statist.*, vol. 23, no. 1, pp. 255–271, 1995.

[64] R. Stoica, X. Descombes, and J. Zerubia, "A Gibbs point process for road extraction in remotely sensed images," *Int. J. Comp. Vision*, vol. 57, no. 2, pp. 121–136, 2004.

[65] P. Thuman-Commike and W. Chiu, "Automatic detection of spherical particles from spot-scan electron cryomicroscopy images," *J. Micros. Soc. Amer.*, vol. 1, pp. 191–201, 1995.

[66] V. Torre and T. A. Poggio, "On edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 2, pp. 147–163, Mar. 1986.

[67] F. Tupin, H. Maitre, J. F. Mangin, J. M. Nicolas, and E. Pechersky, "Detection of linear features in SAR images: Application to road network extraction," *IEEE Trans. Geosc. Remote Sens.*, vol. 36, no. 2, pp. 434–453, Mar. 1998.

[68] M. N. M. van Lieshout, "Stochastic Geometry Models in Image Analysis and Spatial Statistics," CWI tract 108, 1995.

[69] P. Viola and M. Jones, "Robust real-time object detection," *Int. J. Comp. Vision*, vol. 57, pp. 137–154, 2004.

[70] Z. Yu. Particle Picking. [Online]. Available: http://www.ices.utexas.edu/~zeyun/CryoEM.htm