Multiscale Significance Run: Realizing the 'Most Powerful' Detection in Noisy Images *

Xiaoming Huo, Jihong Chen

School of ISyE
Georgia Institute of Technology
Atlanta, GA 30332

David L. Donoho

Department of Statistics
Stanford University
Stanford, CA 94305

Abstract

Detection is a fundamental problem in many applications. In many cases, knowing the presence of underlying objects is of significant importance. Multiscale methods have been demonstrated to be advantageous in solving this problem. Besides theoretical results that have been achieved, this paper discusses how the 'most powerful' detection can be realized, for a set of specifically organized underlying objects. We focus on the design of the detection procedure. Multiscale Significance Run Algorithm—MSRA—serves as a general framework. It is shown that by assigning an hierarchy to the alternatives, one can nearly realize the most powerful detection under certain conditions.

1 Introduction

Detection is a fundamental problem in many applications. There are two levels of detection problems. At the first level, one may simply determine whether or not an underlying object exists. At the second level, given that an underlying object is present, one wants to estimate the location, sizes, strength, orientation, etc., of the underlying object. The second level detection problem usually is also called an *inference* problem. In this paper, we consider the detection problem at the first level. The improvement is that by organizing the potential underlying objects, a 'most powerful' hypothesis testing can be (in some sense approximately) realized.

Recently, it has been shown that a multiscale analysis can reveal the true 'dimensionality' of detecting geometric objects in a Gaussian random field [1]. Due to the space, we refer to that paper for more details. The multiscale tools that are utilized in the paper [1] have been applied in detection problem, and scattered in several papers [4, 3, 6]. Most interestingly, while considering another type of detection problem—i.e. detecting a function in a point cloud—an approach called Multiscale Significance Run Algorithm (MSRA) is developed in [2]. A synopsis of this approach will be provided later. The key result is that by applying MSRA, the fundamental threshold of the detectability—above this threshold, an object is detectable; below this threshold, no one should be able to distinguish it from pure noises—is obtained. The paper [2] solves a problem for random point clouds. Later on, in [5], the same method is extended for detecting objects in imagery data.

Despite the successes in determining the asymptotic rates of the detectability under various situations, the original MSRA serves as a starting point. In the sense that (1) there is no discussion on how to choose the optimal values for the procedural parameters, (2) the discussion on the power of the hypothesis testing procedure can be strengthened. In this paper, by assigning a hierarchical structure to the possible underlying objects, we discuss how to choose the optimal values of the procedural parameters in MSRA. We will argue that our choices of these parameters will lead to (in an approximate sense) the 'most powerful' test (i.e., detection).

The rest of the paper is organized as follows. In Section 2, three exemplary detection problems are presented. In Section 3, the MSRA is summarized. Section 4 describes the effects of the values of the parameters in MSRA. Section 5 describes selection of the parameters, which will lead to a statistically 'most powerful' test. Some discussions, summary and conclusion are provided in Section 6.

2 Exemplary Detection Problems

We consider the detection problem, in which the underlying objects are barely above the possible level of being detectable. Three examples are given.

 Published in the Proceedings of Asilomar Conference on Signals, Systems, and Computers, November, 9-12, 2003, Pacific Grove, CA.

Example 1 (Ship Wake Detection) *An example of ship detection is presented in Figure 1. In the first figure, there are two small boats (with ship wakes) in the northwest corner. The second image only contains the ocean—no ship. We are interested in whether the two images are distinguishable. If they are, in what sense?*

Both images are taken by satellite, and are downloadable on the web:

- *The image (a) is taken near golden gate bridge, San Francisco, CA. Besides the two small boats, a large one left the picture, and an incoming one is near the bottom. It was taken on July 10, 1993, and covers roughly a region of* 800 *meters by* 600 *meters.*
- *Image (b) is an ocean image with no boat in it. It is taken about* 17 *km South of St. Petersburg, Florida, United States.*

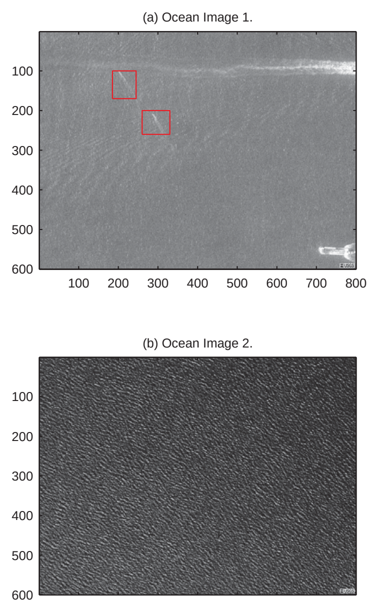

Figure 1: A ship wake image and an oceanic image. In (a), two small boats are in two boxes.

Example 2 (Cryo-EM Images) *Figure 2 presents an image that is generated by cryo-electron microscopy (cryo-EM). It is an example in Microtubule Studies (refer to http://cryoem.berkeley.edu/). Since the image is taken at molecular level, significant amount of noises are present. The question is whether at the current resolution, enough information is available for confident inference of the underlying structures.*

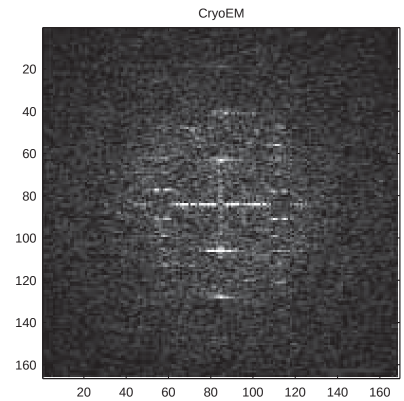

Figure 2: A cryo-electron microscopy (cryo-EM) image.

Example 3 (Features in Texture Images) *Figure 3 illustrates some features embedded in texture images. Again, we would like to detect the presence and locations of the embedded objects.*

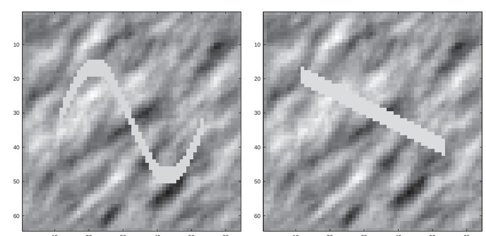

Figure 3: Features embedded in texture images.

3 MSRA: Multiscale Significance Run Algorithm

MSRA is introduced in [2] to derive an asymptotic rate of the level of detectability for an underlying curve. The basic idea can be summarized as follows.

1. *Basic elements.* Some regions, which satisfy specific geometric conditions, are chosen as *basic elements.* In [2], the basic elements are *axoids.* In [5], the basic elements are *digital axoids.* In multiscale approach, basic elements are the set of multiscale features (such as dyadic intervals, or dyadic rectangles, or beamlets, etc as in [1]).

2. *Significant elements.* For a given basic element, one would like to decide whether or not the content inside this element is likely to be 'abnormal': likely to overlap with an underlying structure. An abnormal basic element is *significant.* Let "p" de-

note the probability that a basic element is significant.

3. *Significant graph and a Bernoulli table.* If two basic elements are next to each other, they are 'connected'. Let all the significant elements be the nodes in a graph. There is an edge between two nodes iff the two corresponding basic elements are connected. Hence we have a graph, which is called *significant graph.* In the cases that are described in [2, 5], the basic elements are equally spaced at x and y coordinates and orientations, the significance graph can be abstractized as a table; each entry has a fixed chance to be significant. For obvious reasons, this table is called a Bernoulli net (or table). An illustration of a Bernoulli net is given in Figure 4.

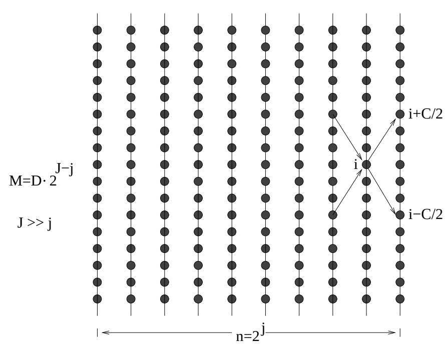

Figure 4: A Bernoulli net (significance graph).

4. *Longest runs.* A *significance run* is a chain of connected significant nodes. The *longest run* is the longest significance run, i.e. the longest chain of connected significant basic elements in a Bernoulli net.

5. *Test.* The test statistic is the length of the longest run(s). The motivation is that if there is no underlying structure, the length of the longest run is statistically upper bounded. If the *observed* length of the longest run is much bigger than the just mentioned upper bound, then there is an evidence that there is an underlying object.

Remark 4 *We solve a first level detection problem.*

Figure 5 presents a showcase of MSRA, for imagery data.

4 Parameters in MSRA

As illustrated in Figure 4, there are four parameters in a Bernoulli net. They are

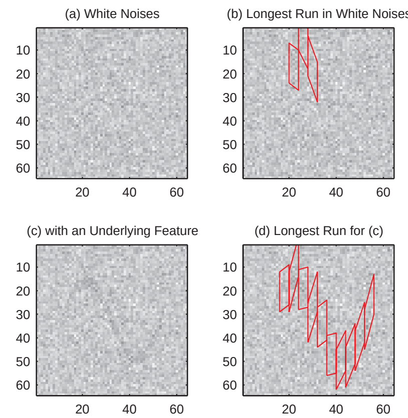

Figure 5: Two showcases of MSRA. In each case, (a) is a pure noisy image; (b) is a longest run in the noisy image (a); (c) is a noisy image with an embedded feature; (d) is a longest run for image (c). The longest runs are dramatically different when there is a feature embedded.

- the size of the Bernoulli net, which are denoted by "M" (height) and "n" (width);
- the success probability, which is equal to "p"—the probability of being significant;
- the number of nodes that are connected in the next level, which is denoted by "C".

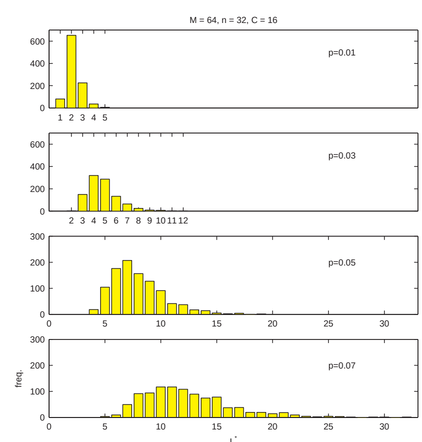

Figure 6: L^* versus p: effects of significance probability p. When the value of p is increased, the histogram of L^* is shifted to the right. No. of simulations is 1000 per histogram.

Let L^* denote the length of the longest run in the Bernoulli net. The distribution of L^* depends on these

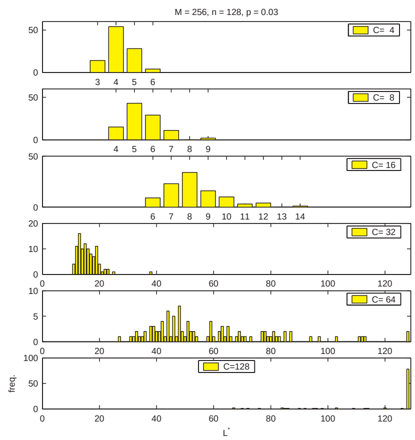

Figure 7: L^* versus C: effects of connectivity. Every time when the value of C is doubled, the histogram of the lengths of the longest run L^* is shifted to the right significantly. No. of simulations is 100 per histogram.

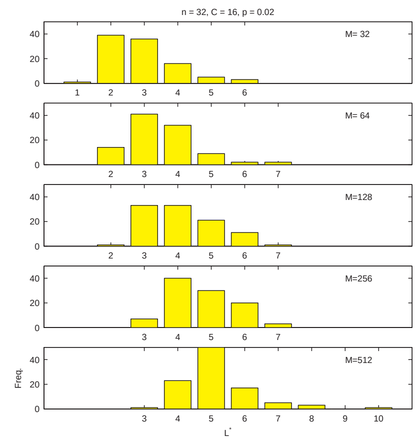

Figure 8: L^* versus M: effects of heights. When the value of M is doubled, the histogram of L^* does not change dramatically. No. of simulations is 100 per histogram.

parameters: M, n, p, and C. Simulations are carried out to study their relation at finite samples. The following presents a summary of the results.

- For fixed values of M and n, when the value of C or p is increased, the distribution of the L^* changes dramatically. These can be seen in Figure 6 and Figure 7.
- For fixed C and p, if the value of M or n is doubled, the change of distribution of L^* is *insignificant*. These can be observed in Figure 8 and 9.

Normally in designing an MSRA, one assumes that the value of C is fixed. Values of M and n may vary. However because they have small effects on L^*, their effects are negligible. Hence an important designing factor is the probability p. In the next section, we will focus on the choice of probability p.

Remark 5 *For a fixed scale, there is a Bernoulli net. When the scale is increased (resp. decreased) by* 1*, the value of* n *is doubled (resp. halfened). Because it is shown that the value of* n *has small effect on the maximum length* L^**; we can assume to be working at one scale. Such a simplification for analysis should* **not** *be interpreted as working on a mono-scale methodology.*

5 'Most Powerful' Detection

In this section, we will prove that if there exists a test (i.e. detection) with power close to 1, then there is a selection approach for MSRA, such that the power of the derived test is close to one. In this sense, we claim that the derived test is nearly 'most powerful'. In the rest of this section, the formulation of the hypothesis testing problem is described in Section 5.1; a lower bound of the power of a type of MSRA is derived in Section 5.2; the *nearly most powerfulness* is proved in Section 5.3.

5.1 Hypothesis Testing with Composite Alternative

The detection problem is equivalent to a hypothesis testing problem, with a simple null hypothesis (no underlying structure) and composite alternative hypotheses (determined by the shape, thickness, length, orientation, amplitude, smoothness, etc., of the underlying structures).

As a preparation, we introduce some notations. For a fixed scale s, let $r_a(s)$ denote a chain of basic elements that 'cover' the underlying object (also known as structure). Let $L(r_a)$ denote the length of this chain. Note $L(r_a)$ is also the length of the corresponding run in the Bernoulli net. Let θ_a denote the parameters associated with the underlying object. It may include in the information on amplitudes, etc. Let $T(p, s)$ denote the parameter by following which, the success probability in the Bernoulli net is equal to p for scale s. Note that in the axoids cases in [2, 5], $T(p, s)$ is the threshold for the X-statistics. However, $T(p, s)$ may stand for parameters in a more general sense. Let $p'(\theta_a, s, T(p, s))$ stand for the probability of being significant for an element in run r_a. Sometimes, $p'(\theta_a, s, T(p, s))$ is also written as $p'(p; \theta_a, s)$, to emphasize the fact that p' is a function of p.

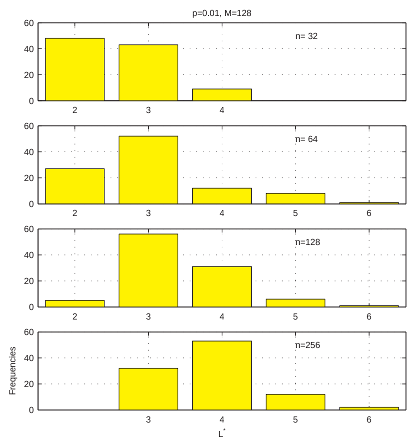

Figure 9: L^* versus n: effects of the width of the Bernoulli table. Every time when the value of n is doubled, the histogram of L^* does not change dramatically. No. of simulations is 100 per histogram.

5.2 Power of the Test for a Single Alternative

Given a simple alternative hypothesis: $H_a : r_a, \theta_a$. Let $L^*_\alpha(p,s)$ denote the $(1-\alpha)$-percentile of the maximum lengths (L^*) under the null hypothesis (H_0). Let I stand for an image generated according to the alternative. The power of an MSRA-based-test with level α, for a simple null and a simple alternative, is

$$\text{power} = P\{L^*(I) > L^*_\alpha(p,s)\},$$

where $L^*(I)$ is the length of the longest run in image I.

We only consider the cases when $p' \sim 1$. Because if there is a counterexample against the null hypothesis, the counterexample should be in the neighborhood of the underlying object. A lower bound of the power function is

$$\text{power} \geq [p'(p;\theta_a,s)]^{L^*_\alpha(p,s)+1}. \quad (1)$$

Notice that both $p'(p;\theta_a,s)$ and $L^*_\alpha(p,s)$ are monotonically increasing functions for variable p. Recall a condition:

$$L^*_\alpha(p,s) < L(r_a);$$

otherwise the underlying object will *not* be detectable. To maximize the lower bound of the power in (1), the following algorithm can be applied:

---

1. **For** $t = L(r_a)-1, L(r_a)-2, \ldots, 1$,
   - choose the maximal value of p, conditioning on
     $$L^*_\alpha(p,s) \leq t;$$
   - For the above p, compute $p'(p;\theta_a,s)$;
   - compute $\text{power}(t) = [p'(p;\theta_a,s)]^{t+1}$.
2. the maximum value of the lower bound in (1) is equal to
   $$\max_{t<L(r_a)} \quad \text{power}(t). \quad (2)$$

Remark 6 (New Detection Procedure) *The above can be viewed as a procedure including $[L(r_a)-1]$ times of hypothesis tests: the null is rejected iff the null is rejected at least for one time; Instead of computing $p'(p;\theta_a,s)$, we find $L^*(I)$ and reject the null iff $L^*(I) > t$. The upper bound of the Type-I error (level of the overall test) is increased to $[L(r_a)-1]\alpha$; while the power of the test is the maximum in the right hand side of (1). If the lower bound in (1) is close to one, the power of this test is close to* 1.

Remark 7 *It is shown that value of s has little effect on the value of $L^*_\alpha(p,s)$. Hence the values of $L^*_\alpha(p,s)$'s may be computed across scales. Finding the maximum value of p under condition "$L^*_\alpha(p,s) \leq t$" may not be an easy task. It can be done numerically.*

Remark 8 *The above is a simplification of a much more complex dependence relationship. For the case of filament detection in [5], Figure 10 illustrate the interdependences. The notation N^* is the threshold of significance, which is the same as $T(p,s)$ that was defined earlier.*

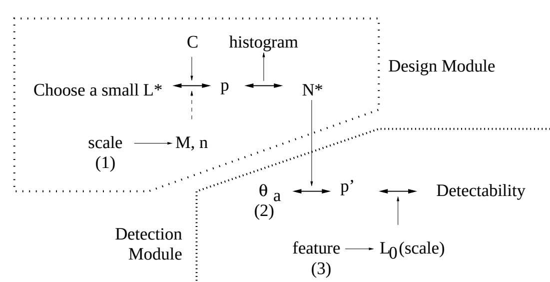

Figure 10: Design strategy for MSRA.

5.3 Nearly 'Most Powerful' Test

Now we prove the statement at the beginning of this section. Consider the following cases:

$$L(r_a) = 2^{j_0}, 2^{j_0+1}, 2^{j_0+2}, \ldots, 2^{j_1}; j_0 < j_1.$$

Assume there is an integer $j_2, j_0 \leq j_2 \leq j_1$, such that $p' \sim 1$ and there is a test with power close to one—more specifically, we have $(p')^{L^*+1} \sim 1$. By choosing

$$\begin{aligned} \alpha &= \alpha_0 \frac{1}{2^k}(j_1 - j_0 + 1)^{-1}, \quad \text{and} \\ L(r_a) &= 2^k, k = j_0, \ldots, j_1, \end{aligned}$$

applying the procedure that was described in the previous subsection, similar to the arguments in Remark 6, the overall test has level α_0 and power being close to one. For obvious reasons, this test is called a *nearly most powerful* test.

Remark 9 *The above test requires*

$$2^{j_0} + \cdots + 2^{j_1} = 2^{j_1+1} - 2^{j_0}$$

times of using Significance Run Algorithms (SRA).

Remark 10 *The above test can be simplified without loss of the power of the test, by taking advantages of the interdependence between the SRAs. The necessary efforts that are required in implementation may be significantly less than it appears. The details are very involved and lengthy. They are skipped.*

Remark 11 *For nondyadic values of $L(r_a)$, because it can be approximated by a dyadic integer, the same result still holds. The reason we do* **not** *consider arbitrary valued integer is that we do not want the value of α being too small.*

6 Summary and Conclusions

We considered the problem of detecting the presence of embedded objects. MSRA serves as a foundation to organize the underlying objects. When there is an object, with which there are powerful detections—the powers of the tests are close to "1" for a range of levels of the test—then a combination of MSRA with specially chosen procedural parameters is nearly most powerful—its power will be close to "1" too. The essence of the approach is to decompose a test with composite alternative to a combination of multiple hypothesis testings with each of them having simple alternative. This method provides a unified and most powerful detection for underlying objects.

The current paper presents a theoretical analysis, which points the feasibility of the proposed approach. Details in the numerical algorithms will be refined and improved. More importantly, finding the maximum value for probability p when the value of L^* is given could be a challenging numerical problem. We will also explore how to use the interdependence in the Bernoulli net to improve the power of the test. It is conjectured that the current estimate has the right asymptotic order; this needs to be proved in a rigorous way.

Acknowledgements

This work has been partially supported by National Science Foundation grants DMS 00-77261, DMS 01-40587 and DMS 01-40698.

References

[1] E. Arias, D. L. Donoho and X. Huo (2003). Asymptotically optimal detection of geometric objects by fast multiscale methods. *Tech. Report.* Stanford University. http://www-stat.stanford.edu/∼donoho/reports.html.

[2] E. Arias, D. L. Donoho and X. Huo (2003). Adaptive multiscale detection of filamentary structures embedded in a background of Uniform random points. *Tech. Report.* Stanford University. http://www-stat.stanford.edu/∼donoho/reports.html.

[3] D. Donoho and X. Huo (2001). Applications of beamlets to detection and extraction of lines, curves and objects in very noisy images. *Nonlinear Signal and Image Processing (NSIP)*, Baltimore, June.

[4] D. Donoho and X. Huo (2000). Beamlet pyramids: a new form of multiresolution analysis, suited for extracting lines, curves and objects from very noisy image data. Published in *Wavelet applications in signal and image processing VIII.* Presented in *SPIE*, San Diego, CA.

[5] X. Huo, J. Chen, and D.L. Donoho (2003). Multiscale Detection of Filamentary Features in Image Data. *SPIE Wavelet-X,* San Diego, CA, August.

[6] X. Huo and D. Donoho (2002). Recovering filamentary objects in severely degraded binary images using beamlet-decorated partitioning. *International Conference on Acoustic Speech and Signal Processing (ICASSP)*, May, Orlando, FL.