1998 International Computer Symposium
Workshop on Image Processing and Character Recognition
December 17-19, 1998, N.C.K.U., Tainan, Taiwan, R.O.C.

A NEW SAMPLING METHOD FOR RANDOMIZED CURVE DETECTION*

Yu Chin Cheng†and *Yu-Sun Liu*‡

†Department of Electronics Engineering
National Taipei University of Technology, Taipei, Taiwan 106
email: yccheng@en.ntut.edu.tw
‡Department of Electrical Engineering
Chung-Yuan Christian University, Chungli, Taiwan 320
email: liu@dec.ee.cycu.edu.tw

ABSTRACT

In randomized curve detection, sampling is the process of collecting a set of edge pixels in a edge map from which hypotheses are generated and verified. In the present work, we propose to define a sample to be a set of edge pixels on the path of a random straight line thrown into the edge map. It is shown that with the proposed sampling method, the number of random samples to take can be determined easily and reliably. This is a significant improvement over some existing works where determining this number depends one or more edge map characteristics that may not be easily obtained. Experimental results show that the proposed sampling method is robust with respect to noise and incompleteness of the curve under detection.

Keywords: curve detection, random sampling, randomized iterative methods

1 Introduction

Randomized curve detection samples sets of edge pixels to form hypotheses of the curves under detection. In the existing works, the samples are either collected randomly on a pixel-by-pixel basis, or sought from the edge map heuristically. In both cases, deciding the number of random samples to take is an important design decision that must be made prior to executing the detection program. This design decision problem has been addressed in depth in several existing works utilizing theoretic or heuristic models [1, 2, 3, 4]. A common feature of nearly all of these works is that, in the computation of the number of random samples to take, they invariably use one or more edge map specific characteristics such as the number of true curves in the edge map, the amount of noise, the minimum number of pixels on a true curve, the maximum number of pixels on a false positive, and so on. The limitation of using edge map specific characteristics is obvious from a practical point of view: they are either not available or difficult to extract. It is also interesting to observe that, in some of these works, additional assumptions on certain edge map characteristics are made; they can impose severe limits on the types of edge maps that can be processed. For example, in [2], it is assumed that the minimum number of pixels on a true curve is several times the maximum number of pixels on a false positive. In our experience, this assumption may not be valid in a noisy edge map and the detection method can deliver unpredictable (and usually incorrect) results.

In this paper, we propose to define a random sample to be a set of collinear edge pixels which is retrieved from the edge map by throwing a random straight line and collecting the edge pixels on its path. Specifically, if the random straight line intersects a true curve with probability p, then a number of pixels on the true curve are inside the random sample with probability p. Depending on the detection method, pixels from one or more random samples are then enumerated systematically to generate hypotheses. With the proposed method, the number of random samples can be determined based on the probability p so that at least one hypothesis corresponding to a true curve is generated with high confidence; moreover, no edge map characteristics are used.

In the rest of this paper, Section 2 presents the proposed sampling method and its comparison to some existing methods. Section 3 presents experimental results to demonstrate its robustness with respect to incompleteness of the curve under detection and different noise levels. In Section 4, possible generalizations of the proposed method with nonlinear sampling agents are discussed.

2 The sampling method

We propose to define a random sample as follows: a random sample is a collinear set of edge pixels that are collected from the path of a random straight line thrown into the edge map. Let the probability that a random straight line intersects a degree-d curve under

*This research is supported by the NSC grant 87-2213-E-027-001.

†Corresponding author.

detection in the general position be p, then the random sample contains d edges of the curve with probability p. Notice that it may take edges from more than one sample to generate a hypothesis; in this case s random samples are taken in a batch, where s is minimum number so that sd is at least as large as the number of constraints required to generate a hypothesis. Thus, if ks samples are taken in k batches, a hypothesis corresponding to a true curve can be generated with a confidence no less than $1-(1-p^s)^k$. In what follows, we shall illustrate how the proposed method can be applied in randomized curve detection.

Example 1: *Circle detection where a hypothesis is generated from three non-collinear edges [5].* Two random samples are needed: one will provide a pair and the other will provide a singleton. The confidence is no less than $1-(1-p^2)^k$ if k pairs of random samples are taken. ∎

Example 2: *Circle detection where a hypothesis is generated by coaxal transform [6].* If coaxal transform is employed, it suffices to generate a hypothesis from one random sample because two edges can be used to construct a coaxal system. The confidence is no less than $1-(1-p)^k$ if k random samples are taken. ∎

Example 3: *Ellipse detection where a hypothesis is generated from five edges.* Three random samples are needed: two will each provide a pair and the other one will provide the singleton. The confidence is no less than $1-(1-p^3)^k$ if k triplets of random samples are taken. ∎

Example 4: *Ellipse detection where a hypothesis is generated from four edges with the pencil-of-conics construction [4].* Two random samples each providing a pair are needed. The confidence is no less than $1-(1-p^2)^k$ if k pairs of random samples are taken. ∎

In the above examples, if edge gradients are also used, it is possible to cut the number of random samples by half with appropriate change to the hypothesis generation method.

The probability p that a random line intersects a curve under detection depends on the geometry of the curve under detection and how the random line is thrown. We shall illustrates how p can be determined based on Example 2.

2.1 Applying the method for circle detection

Let the pair of random variables (X, Y) and the random variable R denote the center and the radius of a circle in the real plane, respectively. It is assumed that (X, Y) and R are independently distributed over $[0, c] \times [0, c]$ with $c > 0$ and $[r_1, r_2]$, with $r_2 > r_1 > 0$. respectively. Let $y = x \tan\Theta + B$ be a random straight line that intersects the edge map, where the random variable Θ denotes the angle between the line and the X-axis, and the random variable B is the Y-intercept. The random variables Θ and B are chosen according to uniform distributions over $[0, 2\pi]$ and $[0, c]$, respectively. We would like to calculate the probability that a random straight line $y = x\tan\Theta + B$ intersects a circle with center (X, Y) and radius R.

First, consider a fixed circle with $(X, Y) = (x_c, y_c)$ and $R = r$. The distance from (x_c, y_c) to the random line $y = x\tan\Theta + B$ is

$$|x_c \sin\Theta + (B - y_c)\cos\Theta|.$$

The probability that the circle is intersected by the random straight line is

$$p = \Pr\left(|x_c \sin\Theta + (B - y_c)\cos\Theta| \le r\right). \quad (1)$$

Thus, the probability that n independent random straight lines fail to intersect the circle is

$$(1-p)^n = \left(1 - \Pr\left(|x_c \sin\Theta + (B - y_c)\cos\Theta| \le r\right)\right)^n. \quad (2)$$

Finally, the probability that n independent random lines fail to intersect the random circle under detection is the weighted average of (2) over all possible (x_c, y_c) and r,

$$\int_{r_1}^{r_2}\int_0^c\int_0^c (1-p)^n f_{XY}(x,y) f_R(r)\,dx\,dy\,dr, \quad (3)$$

where $f_{XY}(x, y)$ and $f_R(r)$ are the probability density functions of the center (X, Y) and the radius R of the circle under detection, respectively.

In the following examples we shall consider three different models of the circle under detection.

Example 5: Both (X, Y) and R are uniformly distributed with $c = 256$ and $r1 = 12, r_2 = 64$. This uniformly distributed model can be used in the case where very little information of the location and size of the circle under detection is available except their limits.

Example 6: (X, Y) is uniformly distributed with $c = 256$ and R is a random variable with probability density $f_R(r) = \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{(r-\mu)^2}{2\sigma^2}\}/(2\Phi(3)-1)$ if $12 \le r \le 64$ and 0 otherwise, where $\mu = 36$, $\sigma = 24/3$, and $\Phi(x) = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\{\frac{-x^2}{2}\}dx$ [7]. This is the case where radius of the circle under detection is likely to be close to μ; the larger the radius deviates from μ, the less likely it is for such a circle to be present. This model corresponds to the case where some information about the size of the circle is available.

Example 7: X and Y are identical and independent random variables with probability density $f_X(z) = f_Y(z) = \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{(z-\mu)^2}{2\sigma^2}\}/(2\Phi(3)-1)$ if $0 \le z \le 256$ and 0 otherwise, where $\mu = 128$, and

$\sigma = 128/3$. R is a random variable with probability density function as in Example 6. This model is applicable to the case where some information about the location and the size of the circle under detection is available.

Notice that a random line intersects the circle at two points that are independent of each other. Thus, if the circle is $100w\%$ complete, $0 \leq w \leq 1$, the probability that none of the n independent random straight lines intersect the incomplete circle in two points is

$$\int_{r_1}^{r_2} \int_0^c \int_0^c \{1 - pw^2\}^n f_{XY}(x,y) f_R(r) dx\, dy\, dr. \quad (4)$$

Fig. 1 shows this probability for complete circles and semi-complete circles in the three models above. It is evident from the figure that for a given confidence, fewer random lines are thrown if more information about the circle under detection is available. Notice that the failure probabilities can be computed off-line before executing the detection program.

2.2 A comparison with existing method

The way that the number of random samples to take is determined in the proposed method is a significant improvement over several existing sampling methods. To see this, we shall review three existing methods.

Probabilistic Hough transform [1]. In order to approximate the Hough transform with confidence $(1-\delta)$, Bergen and Shvaytser propose to set the number of random samples to take to be

$$n = \min\{3\frac{\ln \frac{1}{\delta}}{(\epsilon^2\mu)^r}, \frac{1-\mu^r}{(\mu\epsilon^2)^r\delta}\}, \quad (5)$$

where r is the sample size, ϵ is the error bound, and

$$h_s \leq \mu \leq \frac{1-\epsilon}{1+\epsilon} h_b.$$

Let m be the number of edge pixels in the edge map, mh_b is the minimum number of pixels on a true curve and mh_s is the maximum number of pixels on a false positive, and $0 \leq h_s < h_b \leq 1$.

Randomized Hough transform (RHT) [2]. RHT conducts detection in *epochs*, and at most one curve is detected in each epoch. With the assumption that the minimum number of pixels on a true curve n_{min} is much larger than the maximum number of pixels on a false positive n^{ps}_{max}, Xu and Oja propose to set the number of random samples to take in each epoch to be

$$K_{max} \approx (10 \sim 100)\frac{N^d}{n^d_{min}} \quad (6)$$

where d the the degree of freedom of the curve under detection and N is the number of edge pixels in the edge map.

K random sample consensus [4]. To achieve a confidence $1 - \frac{1}{w^x}$ that a hypothesis corresponding to a true curve is generated, the number of random samples to take is set to be

$$\frac{k^{(d-1)}(1+w)}{\mu^d} \quad (7)$$

where d is the sample size, k is the number of curves under detection, and μ is the ratio of signal edge pixels to the total number of edge pixels in the edge map.

Clearly, each of the three methods above relies on edge map characteristics that may not be easy to obtain in general. In contrast, the proposed method does not rely on such information.

3 An experiment on circle detection

We use edge maps with uniformly distributed noise to test the robustness of the proposed method. The following definition of signal-to-noise ratio (SNR) is used [8]: given an edge map E1 of resolution $W \times H$, an edge map E2 is said to be E1's x-dB edge map if

$$\text{SNR} = x = 20 \log \frac{W \times H}{R}, \quad (8)$$

where R is the number of pixel reversals in obtaining E2 from E1. Figure 2 shows a 256 × 256 edge map at various SNR levels. The input edge maps are created as follows. The 256 × 256 edge map contains a single (complete or semi-complete) circle which has the distributions of Example 5, Section 2. The circle is generated with the Bresenham's algorithm[9], and the semi-complete circle is obtained by randomly removing one half of edge pixels from the full circle. Noisy edge maps at SNR levels of 40 dB, 30 dB, and 20 dB are then generated from the noise-free edge map. The number of random samples taken are set to be $5, 10, \cdots, 100$ for the complete circle and $5, 10, \cdots, 200$ for the semi-complete circle. The robustness of the proposed method is measured by $PE(n)$, the probability that none of n random samples contain at least two edge pixels of a true curve. The failure rate PE is obtained by taking the average over 1000 experiments. Figure 3 shows the curves of the failure rates PE against the number of random samples taken at the given SNR levels. It is evident that the probabilities PE are both behaviorally and numerically similar to the theoretical fail-to-intersect probabilities PT.

It is interesting to note that noise does not seem to have an effect on the failure probabilities. However, it should be pointed that noise does have an definite effect on the hypothesis generation and verification; high levels of noise implies that more hypotheses are generated and that verification is more difficult. This effect has been observed from the experiments in reference [6].

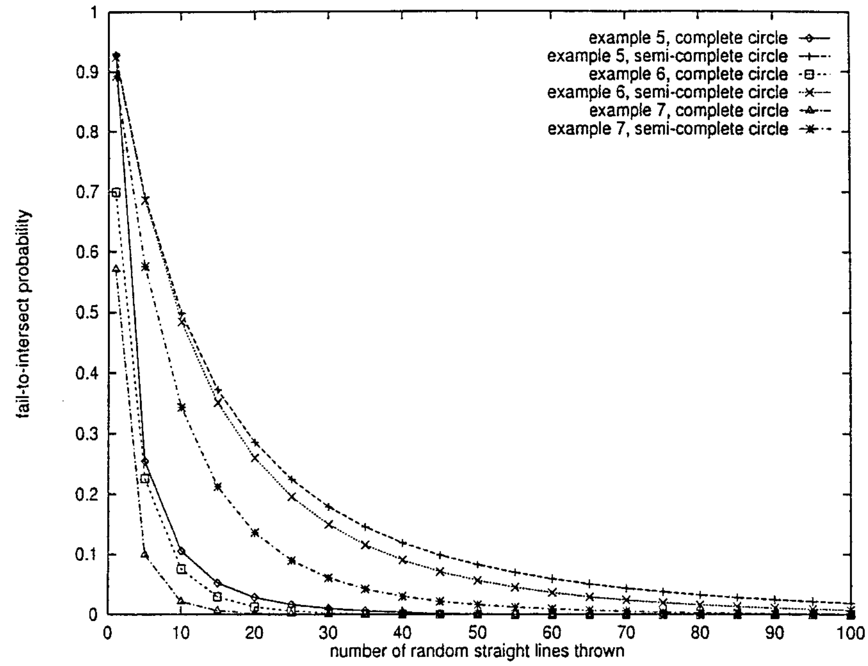


Figure 1: The probabilities that n random lines fail to intersect the complete circles ($w = 1.0$) and the semi-complete circles ($w = 0.5$) in Examples 5, 6, and 7.

4 Concluding remarks

The present paper proposes a new sampling method for randomized curve detection. By defining a random sample to be a set of collinear edge pixels, the number of random samples to take can be determined easily and reliably and without using edge map specific characteristics. This is a significant improvement over the way number of random samples are determined in some existing works, where edge map characteristics are required. Robustness of the sampling method is demonstrated using circle detection as an example.

It should be noticed that the proposed sampling method is only responsible for providing edges for hypothesis generation. Thus, although it is robust, the processes of hypothesis generation and verification must also be robust for the overall detection to be robust. The proposed method can be extended to use an alternative type of sampling set (for example, random samples in which the edge pixels are cocyclic) provided that the probability can be calculated.

References

[1] J. R. Bergen and H. Shvaytser. "A Probabilistic Algorithm for Computing Hough Transforms," *Journal of Algorithms*, 12:639–656, 1991.

[2] L. Xu and E. Oja. "Randomized Hough Transform (RHT): Basic Mechanisms, Algorithms, and Computational Complexities," *CVGIP: Image Understanding*, 57(2):131–154, 1993.

[3] V. F. Leavers. "The Dynamic Generalized Hough Transform: its Relationship to the Probabilistic Hough Transforms and an Application to the Concurrent Detection of Circles and Ellipses," *CVGIP:Image Understanding*, 56(3):381–398, November 1992.

[4] Y. C. Cheng and S. C. Lee. "A New Method for Quadratic Curve Detection Using K-RANSAC with Acceleration Techniques," *Pattern Recognition*, 28(5):663–682, 1995.

[5] L. Xu, E. Oja, and P. Kultanen. "A New Curve Detection Method: Randomized Hough Transform (RHT)," *Pattern Recognition Letters*, 11:331–338, May 1990.

[6] Y. C. Cheng and Y.-S. Liu. "Circle Detection with Randomized Linear Enumeration and Coaxal Transform," *Technical Report NTUT-EN-98-01*, February 1998.

[7] A. Papoulis. *Probabilities, Random Variables, and Stochastic Processes*. McCraw-Hill, New York, 2nd edition, 1984.

[8] A. Khotanzad and J.-H. Lu. "Classification of Invariant Image Representations using a Neural Network," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(6):1028–1038, 1990.

[9] D. Hearn and M. P. Baker. *Computer Graphics*. Prentice Hall, New Jersey, 2nd edition, 1994.

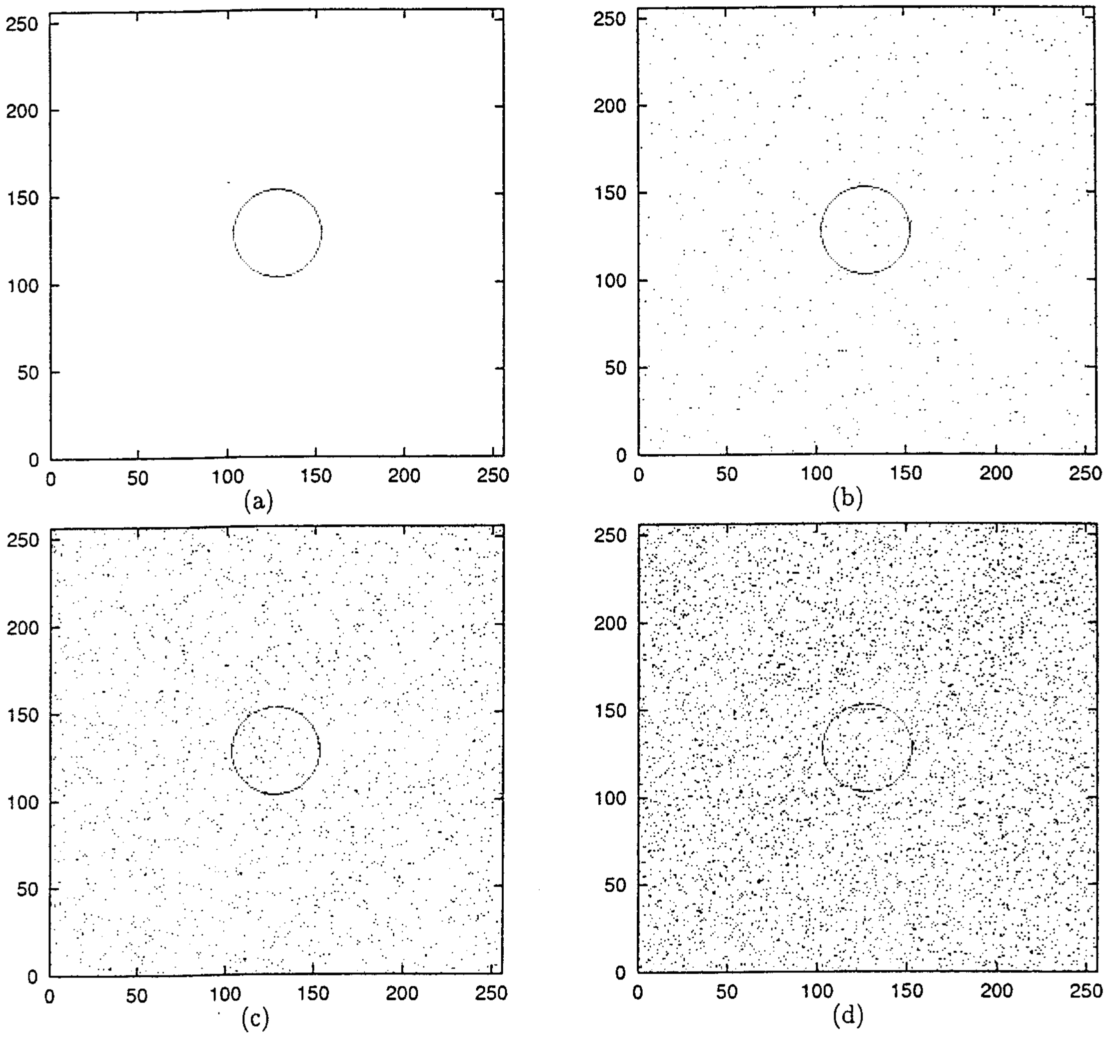
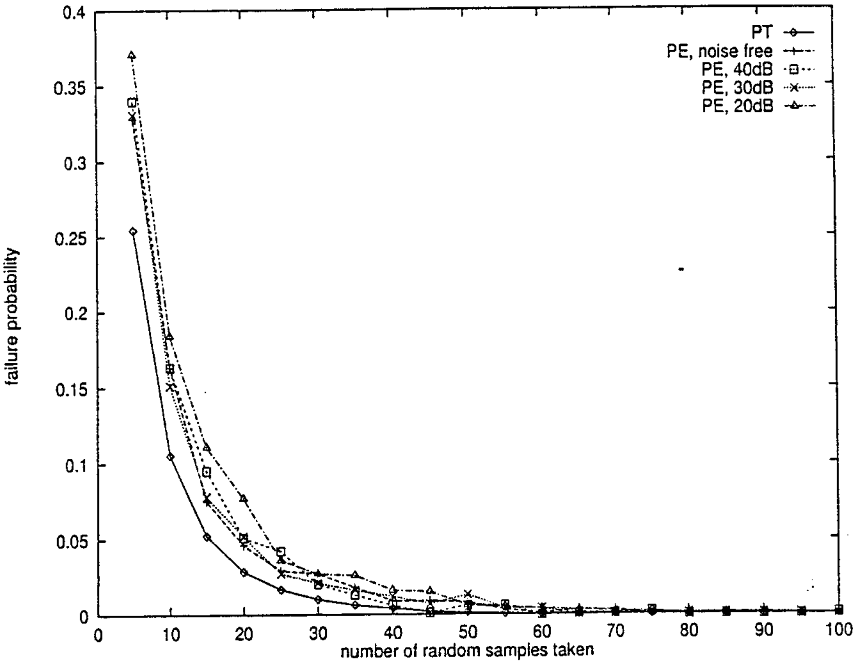
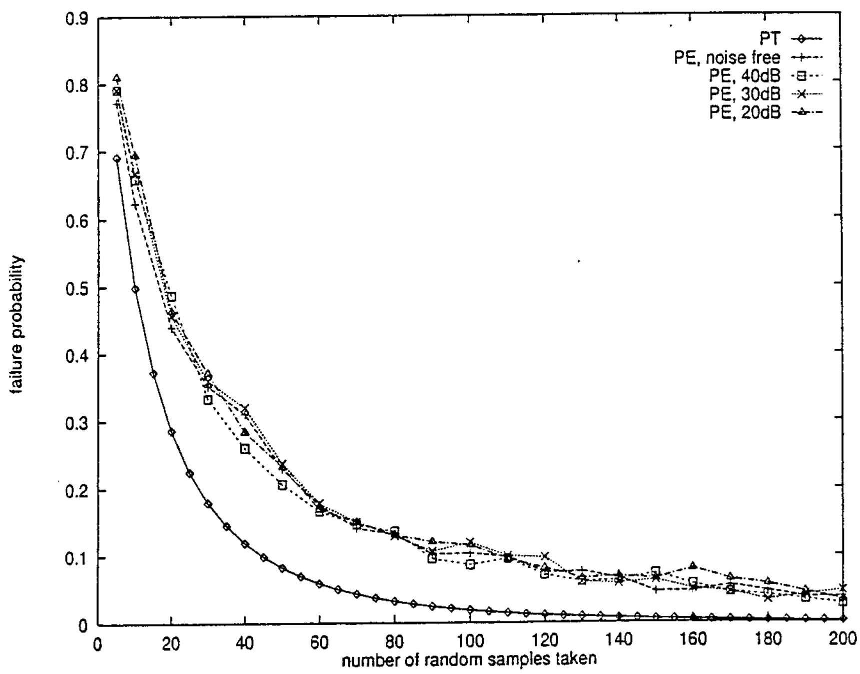


Figure 2: A 256 × 256 edge map containing a circle with center (127, 127) and radius 25 (a) with no noise, (b) at SNR noise levels of 40 dB, (c) 30 dB, and (d) 20 dB.



(a) Complete circle



(b) Semi-complete circle

Figure 3: Experimental results on the robustness of the proposed method applied to circle detection under various noise levels.