

Deriving Stopping Rules for the Probabilistic Hough Transform by Sequential Analysis

D. Shaked, O. Yaron, and N. Kiryati

Department of Electrical Engineering
Technion-I.I.T., Haifa 32000, Israel

Abstract

It is known that Hough Transform computation can be significantly accelerated by polling instead of voting. A small part of the data set is selected at random and used as input to the algorithm. The performance of these Probabilistic Hough Transforms depends on the poll size. Most Probabilistic Hough algorithms use a fixed poll size, which is far from optimal since conservative design requires the fixed poll size to be much larger than necessary in average conditions. It has recently been experimentally demonstrated that adaptive termination of voting can lead to improved performance in terms of the error rate versus average poll size tradeoff. However, the lack of a solid theoretical foundation made general performance evaluation and optimal design of adaptive stopping rules nearly impossible.

In this paper it is shown that the statistical theory of sequential hypotheses testing can provide a useful theoretical framework for the analysis and development of adaptive stopping rules for the Probabilistic Hough Transform. The algorithm is restated in statistical terms and two novel rules for adaptive termination of the polling are developed.

The performance of the suggested stopping rules is verified using synthetic data as well as real images. It is shown that the extension suggested in this paper to Wald's one sided alternative sequential test performs better than previously available adaptive (or fixed) stopping rules.

1 Introduction

The Hough Transform [7, 10, 18] is a broad family of techniques for recognizing predefined curves and shapes in edge images. Hough algorithms consist of an accumulation stage in which edge points “vote” for the parameters of curves to which they possibly belong, followed by a search for peaks in the parameter space. In most Hough algorithms the parameter space is quantized and represented by an array of accumulators. Since it is always possible to replace exhaustive search by a search guided by the voting patterns [9], the computational complexity is essentially proportional to the number of edge points in the image.

It has recently been suggested to accelerate the Hough Transform by polling instead of voting, randomly selecting just a small part of the data set as input to the algorithm [5, 11, 14, 17, 25, 31, 32]. Reference [14] focuses on accelerating Duda and Hart’s standard Hough Transform [7], and references [5, 11, 17, 25, 31, 32] discuss similar ideas in the context of “many to one” variants of the Hough Transform, where several edge points (two in the case of straight lines) map into one point in parameter space. Fast Hough algorithms that are based on polling are referred to in this paper as *Probabilistic Hough Transforms*. The poll size must be large enough so that the detected objects will be, with high probability, similar to those that would have been obtained by full-scale voting. On the other hand, the computational advantage of Probabilistic Hough Transforms diminishes if the poll size is large. In most Probabilistic Hough Transforms, poll sizes are set *a priori* to fixed conservative values that are much larger than necessary in average conditions [32, 33].

It is possible to reduce poll size while maintaining a low error rate by adaptive setting of the poll size. A good stopping rule, i.e. a useful adaptive criterion for termination of voting, monitors the polling process and indicates whether the situation is clear and the maximum can be determined at reasonable certainty without additional votes, or else voting should continue. This allows to reduce error rate at a given average poll size, or to reduce the average poll size while maintaining a certain error rate [33]. With adaptive termination of voting, the poll size depends on the accumulated history of the incoming samples and is thus a random variable rather than a preset parameter.

Recently three suggestions to set the poll size adaptively [17, 33, 11] were made. Ylä-Jääski and Kiryati maintained [33] that setting the poll size adaptively may compensate for variations due to the randomness of polling and due to diversity

among images encountered in a given application. The stopping rule suggested in [33] is based on a measure of stability in the ranks of the highest peaks in the accumulator array. Instability in the ranks of the peaks indicates that the current highest peaks may yet change, and voting is therefore continued. If, however, the ranks seem stable, voting may be terminated. This stopping rule was experimentally shown [33] to use less votes, in average, than would be needed to achieve the same error rate with a fixed poll size. Or from another viewpoint, with the stopping rule tuned to use a given average poll size, the error rate was shown to be lower than would be achieved if the poll size was fixed to that value.

The stopping rule suggested by Leavers [17] uses an estimate of the full Hough Transform obtained after a short session of voting. This crude estimation allows for the determination of a poll size that would probably result in a poll whose highest peak will be close to a predetermined value. This stopping rule is, in principle, similar to a stopping rule that monitors the value of the highest peak in the accumulator array and stops when it reaches some predetermined value. Such a stopping rule was suggested for the Randomized Hough Transform, see [11]. In the statistical literature this stopping rule is known as the Inverse Sampling rule [6].

While relying on strong intuitive arguments and some empirical evidence, the adaptive stopping rules described above lack the theoretical foundation needed for a comprehensive analytic evaluation of their performance. In this paper we study stopping rules suggested for equivalent problems in other fields. We consider the statistical theory of sequential analysis, and show that it provides an analytic framework for adaptive termination of voting in probabilistic Hough algorithms. In particular, we use sequential analysis techniques to derive two novel stopping rules for the Probabilistic Hough Transform.

In Section 2 the adaptive stopping problem is formulated in statistical terms, and solutions proposed for it in the statistical literature are shortly reviewed. In Section 3 sequential analysis basics are presented. Two sequential analysis driven stopping rules are suggested in Section 4. In Section 5 various stopping rules are compared in both small scale simulations and experiments with real images. Technical considerations are discussed in Section 6 and the main conclusions are summarized in Section 7.

2 Statistical Framework

The Hough Transform is a common name for a broad family of algorithms. For a precise problem formulation we have to specify both the Hough technique used and the assigned goal. We use Hough techniques in which each voting element in the image supplies one vote in the parameter space, either by incorporating edge direction data [3, 22], by using “many to one” transformations, or by Gerig’s method [9].

The goal we specify for the Probabilistic Hough Transform is to obtain the same maximum as if the full Hough Transform had been accumulated and analyzed. Possible extensions of the adaptive stopping rules suggested in this paper to the case of multiple objects are discussed in the sequel.

Let us now formulate the problem in statistical terms. The collection of edge data obtained in the preprocessing stage may be considered as an ensemble inducing a probability measure on the parameter space. Since the parameter space is quantized to a finite number of accumulators, the probability is discrete. For each accumulator j in the parameter space, the probability p_j that a random choice from the edge ensemble would increment it, is equivalent to the normalized value of the full Hough Transform in accumulator j . The samples from the edge ensemble are chosen independently, hence, the induced samples in the parameter space are independent samples of a discrete source.

Hence, the task of the stopping mechanism is: Given a sequence of independent samples of an unknown discrete source, find a good stopping time to estimate the event having the maximal probability in the unknown source. Note that from now we shall usually use the statistical terms samples and events instead of votes and accumulators.

The problem of constructing a good stopping rule for the estimation of the maximum of a discrete distribution from its samples was addressed in the statistical literature [1, 4, 6, 23]. A few rules have been suggested, analyzed and compared empirically. The aim was to build up tables where one would be able to look up the problem in terms of size (in our case, the number of accumulators) and acceptable error probability. The tables would give the parameter one should apply to each stopping rule so as to assure the error rate is below the indexing error rate, and indicate an upper bound on the average number of samples the procedure should take. Having located the desired parameters in the tables one would then be able to decide on which stopping rule to use, possibly the rule promising the lowest upper bound on

the average number of samples.

The above description of the purpose of the tables is somewhat simplified. Additional complexity arises from the fact that the average number of samples a procedure may take is inherently dependent on the unknown distribution of the discrete source. It has, however, been proven that there exists a worst case discrete source for the problem of detecting its most probable event [12]. The figures in the tables relate only to this worst case distribution. Thus, for a specified parameter of a stopping rule, any source that may be encountered will require lower average number of samples and yield lower error probability than indicated in the tables. There still remains the theoretical question of whether a procedure that yields a lower upper bound on the average number of samples will also yield a lower average number of samples for other sources as well.

Bechhofer et al. [4] analyzed the fixed sample size case, usually considered as the default sequential stopping rule. The size of the poll is the parameter of the stopping rule. Indeed, aside of the fact that this specific stopping rule has a sample size whose variance is zero (i.e. the number of samples is deterministic), this stopping rule is just like any other sequential stopping rule. Cacoullos and Sobel [6] suggest the *Inverse Sampling* stopping rule, where the sampling is stopped once the number of samples for the most popular event reaches a certain number. This number is the parameter for the stopping rule. Alam [1] suggests a stopping rule that is based on the difference in the number of samples voting for the most popular and second most popular events. Sampling is stopped once the difference reaches a certain number, which is the parameter of the stopping rule. Empirical results indicate improvement in the upper bound of the average number of samples when using the Inverse Sampling rule rather than the fixed sample size. Further improvement is obtained by using the stopping rule suggested by Alam. Unlike the former stopping rules, the number of samples needed for a specific sampling sequence in the latter stopping rule is not upper bounded (although it is shown [1] to be finite with probability 1). In order to combine the advantages of the two adaptive stopping rules Ramey and Alam [23] suggested to combine them, stopping once any of the rules is satisfied. Naturally, this stopping rule has two parameters.

Although the above body of work is very thorough, it is mainly empirical, based on heavy Monte-Carlo simulations. For our purpose it is of limited relevance, since it lacks theoretical means to evaluate and compare the suggested stopping rules. Those are essential since in the Hough case the number of bins in the accumulator array representing the parameter space is no less than a few thousands, and usually much

higher. On the other hand, the interest of the statisticians who constructed the tables was in problems of considerably smaller scale, up to 50.

The roots of a possible theoretical basis to an optimal or a near to optimal stopping rule, lie in the fundamental statistical work done by A. Wald in the late forties. In his book *Sequential Analysis* [29], Wald lay the foundations to sequential decision theory.

In Section 3 we describe Wald's Sequential Probability Ratio Test (SPRT), the fixed sample size setting it relates to, and a few relevant later modifications [19]. In Section 4 we modify the SPRT to fit the statistical problem at hand.

3 Sequential Analysis

Consider the following two different problems in statistical hypotheses testing. The fixed size decision problem: Given a set of n observations decide which hypothesis to accept. The sequential stopping problem: Given a stream of observations decide when to stop observing and what hypothesis to accept. Although the problems seem different their solutions are surprisingly similar. Hence, before we describe sequential algorithms, we will describe the fixed size tests they correspond to.

3.1 The Fixed Sample Size Setting

Assume we have a source of independent and identically distributed samples, whose distribution f_θ is completely defined up to an unknown parameter θ (θ may also be a vector of unknown parameters). We examine a set X_1, X_2, \dots, X_n of n samples of the source and have to choose between two simple hypotheses concerning the unknown parameter θ of the source: The first hypothesis, denoted H_1 , is that $\theta = \theta_1$, and the second hypothesis, denoted H_2 , is that $\theta = \theta_2$, where θ_1, θ_2 are known constants. Denoting the decision we make by D , the two possible errors are the two mismatch errors called the errors of the first and second kind whose probabilities, $P(D=H_2 | H_1)$ and $P(D=H_1 | H_2)$ are denoted α and β respectively.

In general there is no way to minimize both error probabilities simultaneously, so in that sense there is no globally optimal decision rule. However, we may define an admissible test, as a test D whose error terms α and β are such that there exists

no other test D' whose error terms are $\alpha' \leq \alpha$ and $\beta' \leq \beta$ with at least one inequality strict.

Given a value α_0 for the error of the first kind we can order all the tests giving that error rate according to β , their error rate of the second kind. Admissible tests are the tests giving the minimum error rate of the second kind. It can be shown that admissible tests exist for each α_0 and that they are all equivalent to Neyman-Pearson tests, defined by

$$D = \begin{cases} H_1 & \text{if } \frac{P(X_1, X_2, \dots, X_n | H_1)}{P(X_1, X_2, \dots, X_n | H_2)} \geq \lambda \\ H_2 & \text{if } \frac{P(X_1, X_2, \dots, X_n | H_1)}{P(X_1, X_2, \dots, X_n | H_2)} < \lambda \end{cases} \quad (1)$$

where λ is set so that the test fits the requires error of the first kind α . Neyman-Pearson tests are Likelihood Ratio Tests (LRT), where the ratio of the probabilities to observe the actual samples given the two hypotheses, is compared to a parameter λ . Note that the admissible tests whose derivation was parameterized by α could have been equivalently parameterized by β .

We have described a family of fixed size tests for optimal decisions between two simple hypotheses. Before describing their generalization to the sequential setting, we describe two additional generalizations in the fixed size setting. The first generalization is to choose between $k > 2$ hypotheses. The second generalization is to decide between composite hypotheses. While simple hypotheses are of the kind $\theta = \theta_i$, composite hypotheses are of the kind $\theta \in \Theta_i$, where Θ_i is a set of possible values for the unknown parameter corresponding to hypothesis i .

Consider the multiple hypotheses generalization, where we have $k > 2$ hypotheses. Denote $\alpha_{ij} = P(D = H_j | H_i)$, so that α_{ij} for $i \neq j$ is the probability of mismatching j to i , and α_{ii} is the probability of correctly deciding H_i . Instead of minimizing error rates, we may maximize probabilities for correct decision. We cannot simultaneously maximize all correct decision probabilities, however, a linear combination $\sum_i C_i \alpha_{ii}$ can be maximized by

$$D = \arg \max_i C_i P(X_1, X_2, \dots, X_n | H_i) \quad (2)$$

The composite hypotheses case is much more complicated, and unlike the simple hypotheses case, has no known optimal solutions. The Generalized Likelihood Ratio Test (GLRT) for composite hypotheses selects the maximum likelihood parameter of an hypothesis to represent it in the test, so that the GLRT for two composite

hypotheses is

$$D = \begin{cases} H_1 & \text{if } \frac{\max_{\theta \in \Theta_1} P(X_1, X_2, \dots, X_n | \theta)}{\max_{\theta \in \Theta_2} P(X_1, X_2, \dots, X_n | \theta)} \geq \lambda \\ H_2 & \text{if } \frac{\max_{\theta \in \Theta_1} P(X_1, X_2, \dots, X_n | \theta)}{\max_{\theta \in \Theta_2} P(X_1, X_2, \dots, X_n | \theta)} < \lambda \end{cases} \quad (3)$$

The GLRT is only asymptotically optimal, i.e. it gets exponentially close to the optimality bound as n , the size of the set of samples, increases.

3.2 The Sequential Setting

In the sequential setting we do not have only a limited set of samples from the unknown source, we have the source itself, and can sample from it as much as we need. Therefore, we need a strategy to indicate when to stop sampling, and how to decide when we have stopped. Considering all possible strategies, we can, in contrary to the fixed sample size setting, get error probabilities as low as we want, however, we have to pay by increasing the average number of samples.

As in the fixed size setting, we define an optimal strategy with error probabilities α and β , as the strategy with the lowest average number of samples. Note that for most strategies the number of samples needed for a decision is a random variable. Therefore, we do not consider the number of samples, but rather the average number of samples. Note also that the random variable of the number of samples generally depends on the unknown parameter of the source, so that the average number of samples is also a function of the unknown parameter θ .

The sequential equivalent to the Neyman-Pearson LRT is the Sequential Probability Ratio Test (SPRT) of Wald [29]. Wald showed [29] [30] that optimal strategies exist, i.e. there are strategies that achieve the minimum average number of samples simultaneously for all θ . He has also determined that all optimal strategies are equivalent to an SPRT with certain parameters $A > B$:

$$\text{At the } n\text{th sample} \quad \begin{cases} \text{Stop and decide } H_1 & \text{if } \frac{P(X_1, X_2, \dots, X_n | H_1)}{P(X_1, X_2, \dots, X_n | H_2)} \geq A \\ \text{Stop and decide } H_2 & \text{if } \frac{P(X_1, X_2, \dots, X_n | H_1)}{P(X_1, X_2, \dots, X_n | H_2)} < B \\ \text{Continue sampling} & \text{otherwise} \end{cases} \quad (4)$$

It is somewhat surprising to note that whereas in the fixed size setting the correspondence between the desired error rates α , β and the LRT parameter λ was

unknown, in the sequential setting A and B are closely approximated by $\frac{(1-\beta)}{\alpha}$ and $\frac{\beta}{(1-\alpha)}$ respectively.

Multihypothesis sequential tests are surveyed in [8]. When the number of hypotheses is larger than 2, it is generally impossible to find strategies that minimize the expected sampling size simultaneously for all valid parameters θ_i . To check whether a suggested strategy is good, one may, if mathematically tractable, check it against a set of lower bounds given by Simons [26]. Often one simply checks it against alternative strategies in a relevant benchmark, as done, for example, for a specific simple 3-hypotheses case in [8].

A generalization of the SPRT to $k > 2$ hypotheses was first suggested by Sobel and Wald [27]. They introduced it for $k = 3$ simple hypotheses about the unknown mean of a normal distribution. It can, however, be immediately generalized for arbitrary k hypotheses about a single scalar unknown parameter. Because of the limitation to scalar parameters, one can always order the hypotheses according to the order of the parameter θ_i they stand for. Sobel and Wald suggested to use $(k-1)$ SPRTs testing H_i against H_{i-1} for each $2 \leq i \leq k$. The result of the different SPRTs is not interpreted as a choice between the two alternatives, but rather as a pointer pointing up or down. Sobel and Wald showed that for certain error probabilities it can be assured that when all SPRTs have ended they will all point consistently towards a specific hypothesis which will then be decided upon.

The Armitage test [2] is an improvement to the Sobel Wald procedure. There possible $k(k-1)$ probability ratios are monitored. A probability ratio

$$\lambda_{ij} = \frac{P(X_1, X_2, \dots, X_n | H_i)}{P(X_1, X_2, \dots, X_n | H_j)} \quad (5)$$

for all $i \neq j$, may either point at H_i if $\lambda_{ij} \geq A_{ij}$ for a given A_{ij} , or else be inconclusive. The test ends as soon as all the $(k-1)$ ratios that may point at an hypothesis are conclusive. Naturally the hypothesis that has won all the $(k-1)$ alternatives is decided upon. Note that there may be only one such hypothesis. As opposed to the Sobel Wald test, in the Armitage test one may control each error term α_{ij} separately. Furthermore, in the Sobel Wald procedure one could not make sure that each time the test stops, its decision equals the decision of the fixed size test of the same samples, as can be shown to be true for the Armitage test.

Sequential tests for composite hypotheses are surveyed by Lai [16]. While in the fixed sample size setting analysis of the composite hypotheses case was difficult, in the sequential setting it is intractable. However, in several specific sequential testing

problems involving composite hypotheses it can be shown that certain generalized SPRTs yield asymptotically optimal results. In particular, many modifications have been suggested to the one sided alternative of Wald described below. Others, e.g. [20], are generalizations in the spirit of the GLRT.

The *One Sided Alternative Test* is a specific composite hypotheses test for which Wald [29] has proven an optimal generalization of his SPRT. We introduce that test and through it the important concept of *indifference zone*. We first note that the definition of the error rates for two composite hypotheses is: $\alpha = \max_{\theta \in \Theta_1} P(D = H_2 | \theta)$ and $\beta = \max_{\theta \in \Theta_2} P(D = H_1 | \theta)$. The one sided alternative test is defined as follows: H_1 is the composite single parameter hypothesis that $\theta < \theta_0$ and H_2 is the hypothesis that $\theta > \theta_0$. It is clear that if the source distribution f_θ is continuous in θ around θ_0 , decision rules cannot do better than $\alpha + \beta = 1$. To overcome this fundamental problem an indifference zone (θ_1, θ_2) is defined around θ_0 , so that $\theta_1 < \theta_0 < \theta_2$. We assume that the unknown parameter θ of the source is never in the indifference zone. (In the indifference zone the error rates and the average number of samples may be large.) We redefine the hypotheses to be $H_1 : \theta < \theta_1$ and $H_2 : \theta > \theta_2$. Now the corresponding error rates $\alpha = \max_{\theta < \theta_1} P(D = H_2 | \theta)$ and $\beta = \max_{\theta > \theta_2} P(D = H_1 | \theta)$ can get arbitrarily low. The sequential hypotheses testing problem is to find a strategy whose average number of samples is optimal for every θ outside the indifference zone. The optimal sequential test for this problem is the SPRT between the simple hypotheses $H_1 : \theta = \theta_1$ and $H_2 : \theta = \theta_2$.

It is important to note that all composite hypotheses tests (not only the one sided alternative test) have to incorporate an indifference zone, otherwise no meaningful error definition is possible. This results from the fact that no apriori distribution of the unknown parameter is available. Thus, in the composite hypotheses case the only error figure that can be defined is the worst case error, i.e. $\alpha_{ij} = P(D = H_i | H_j)$ can only be interpreted as $\alpha_{ij} = \max_{\theta \in \Theta_j} P(D = H_i | \theta)$. Therefore, as in the one sided alternative test, if a composite hypothesis H_i bounds with a composite hypothesis H_j in the space of distribution parameters, then possible distribution parameters arbitrarily close to the boundary (from either the H_i or the H_j side of the boundary), will cause the sum of α_{ij} and α_{ji} , the confusion errors between H_i and H_j to be arbitrarily close to 1.

In the next section we formulate the problem at hand and locate it in the setting described in this section, as a problem requiring a sequential composite multihypothesis test. We then suggest two tests for the problem, the first is a sequential GLRT and the second involves one sided alternative type tests.

4 Development of Stopping Rules

In the Probabilistic Hough Transform the unknown source is a discrete source, i.e. the source produces a sample $X_j \in \{1, 2, 3 \dots k\}$ with unknown probabilities $P(X_j = i) = p_i$. X_j is the index of the chosen accumulator, and p_i is the normalized value of accumulator i in the unknown full Hough Transform. Our goal is to detect the most probable event of the distribution, $\arg \max_i p_i$, which is the maximal accumulator of the full Hough Transform. Of course we want to complete the task as well as possible based on as few samples as possible. Therefore, the problem at hand is the sequential problem of detecting the most probable event of an unknown discrete distribution.

We define hypotheses: H_i is the hypothesis that the most probable event is event i . Since the number of possible events k is (much) larger than 2, we need a multihypothesis test. The family of all possible distributions has k parameters, and there are many distributions corresponding to each hypothesis (i.e. their most probable event is event i), so that we have a k -parameter composite k -hypotheses decision problem. In fact, the k parameters are restricted by $\sum p_i = 1$. The resulting possible parameter space is thus a $(k-1)$ -dimensional subset of the k -dimensional parameter space. Note also that in our case all the k hypotheses are symmetric.

The samples of the source are independent, hence the empiric distribution is a sufficient statistics at each step. We will denote the empiric distribution by $Q = (q_1, q_2, \dots q_k)$. Each empiric probability q_i is the ratio of n_i , the number of samples $X_j = i$ to the total number of samples, so that after n samples

$$n_i = |\{X_j | X_j = i\}| \quad ; \quad q_i = \frac{n_i}{n} \quad (6)$$

where $|\Psi|$ stands for the cardinality of the set Ψ . Note that $n = \sum_i n_i$.

The way we denote the indices of the events can have no importance in the procedure itself, we therefore adopt a convention that at each stage, the indices are permuted so as to order the current empiric probabilities: $q_1 \geq q_2 \geq \dots \geq q_k$.

Note that according to any reasonable fixed sample size decision test (certainly also the GLRT), the best decision to make at the end of a sampling process is to indicate the most popular event of the sample set as the most probable event. In the convention we adopted, the most popular event is always denoted by index 1. Therefore, whenever the sequential procedure stops, the hypothesis denoted by convention as H_1 is chosen.

To deal with the multiplicity of the hypotheses we adopt the approach of Armitage [2]. Therefore, in both tests we contest k likelihoods in $k(k-1)$ ratios and compare them to a threshold A . Note that since the hypotheses are symmetric, the threshold A does not need to depend on the two hypotheses it arbitrates. Naturally, when the algorithm stops the hypothesis whose $(k-1)$ ratios are conclusive can be no other than H_1 , the most likely hypothesis. Furthermore, the $(k-1)$ ratios that are expected to indicate H_1 will all be conclusive if and only if the ratio confronting it with H_2 , the second most likely hypothesis, is conclusive. Therefore, although adopting the Armitage approach we do not have to check $k(k-1)$ ratios in each iteration but only a single one.

$$\lambda = \frac{P(X_1, X_2, \dots, X_n | H_1)}{P(X_1, X_2, \dots, X_n | H_2)} \quad (7)$$

Note that this significant complexity reduction in the Armitage test is not always possible. To be able to compare only the two most popular hypotheses one has to identify them. In applications where it is as difficult to determine if one hypothesis is more likely than another as it is to calculate their probability ratio, the reduction in complexity is not as drastic. In our case the reduction is possible, since the hypotheses can be easily ranked while voting, with just small additional computing time.

To deal with the compositeness of the hypotheses we adopt an indifference zone at distributions with no definite maximum. We redefine the hypotheses as the following sets of distributions

$$H_i : \Theta_i = \{P | p_i \geq (1 + \delta)p_j, \forall j \neq i\} \quad (8)$$

where δ is the indifference zone parameter. Such an indifference zone is in line with the indifference zone adopted in previous works [1, 4, 6, 23].

As already mentioned before, in order to properly define and solve a composite hypotheses problem in which different hypotheses may bound in the distribution parameter space, an indifference zone (δ in our case) has to be specified. Note that δ is not another parameter of the stopping rule which has to be “tuned”, but a part of the definition of the desired result. The outcome of the algorithm will thus be considered correct even if it points at an accumulator whose value is (relatively) δ -close to the global maximum. See Section 6 for an additional discussion of the indifference zone determination.

In the following subsections we suggest two possible tests for the problem, the

first is a sequential GLRT and the second involves one sided alternative type tests.

4.1 Sequential GLRT

Let us first solve the following two general convex optimization problems: Given k numbers $q_1 \geq q_2 \geq \dots \geq q_k \geq 0$ such that $\sum_i q_i = 1$,

Problem1: Find k numbers $\{\hat{p}_i\}_{i=1}^k$, that minimize $\sum_i q_i \log \hat{p}_i$ subject to:

1. $\sum_i \hat{p}_i = 1$
2. $\hat{p}_1 / \hat{p}_j \geq 1 + \delta, \quad \forall j \neq 1$

Problem2: Find k numbers $\{\tilde{p}_i\}_{i=1}^k$, that minimize $\sum_i q_i \log \tilde{p}_i$ subject to:

1. $\sum_i \tilde{p}_i = 1$
2. $\tilde{p}_2 / \tilde{p}_j \geq 1 + \delta, \quad \forall j \neq 2$

To solve the above convex optimization problems, one can either use Lagrange multipliers or check that the following solutions conform with the Kuhn-Tucker lemma [24]. Given the set $q_1 \geq q_2 \geq \dots \geq q_k$ we find the smallest $m \geq 1$ such that

$$\frac{\sum_{i=1}^m q_i}{m + \delta} \geq q_{m+1} \quad (9)$$

If no such m exists, let $m = k$. The solution to the first problem is:

$$\begin{cases} \hat{p}_1 = (1 + \delta) \sum_{i=1}^m q_i / (m + \delta) \\ \hat{p}_2 = \hat{p}_3 = \dots = \hat{p}_m = \sum_{i=1}^m q_i / (m + \delta) \\ \hat{p}_j = q_j \quad \forall j > m \end{cases} \quad (10)$$

And the solution to the second problem is:

$$\begin{cases} \left. \begin{array}{l} \tilde{p}_1 = \hat{p}_2 \\ \tilde{p}_2 = \hat{p}_1 \\ \tilde{p}_j = \hat{p}_j \quad \forall j > 2 \end{array} \right\} \text{if } m > 1 \\ \left. \begin{array}{l} \tilde{p}_1 = \frac{q_1 + q_2}{2 + \delta} \\ \tilde{p}_2 = \frac{(1 + \delta)(q_1 + q_2)}{2 + \delta} \\ \tilde{p}_j = \hat{p}_j \quad \forall j > 2 \end{array} \right\} \text{if } m = 1 \end{cases} \quad (11)$$

Note that from (9), we can formulate the condition $m = 1$ in terms of q_i and δ

$$m = 1 \iff q_1 \geq (1 + \delta)q_2 \quad (12)$$

We now return to the main problem at hand. To deal with composite hypotheses we use the generalized likelihood ratio scheme, i.e. we use the maximum likelihood parameter of an hypothesis to represent it in the ratio. The test is therefore

$$\frac{\max_{P \in \Theta_1} P(X_1, X_2, \dots, X_n)}{\max_{P \in \Theta_2} P(X_1, X_2, \dots, X_n)} > A \quad (13)$$

To see what the above expression amounts to, let us elaborate on the expression of the likelihood. Since X_j are independent samples, the probability of n samples is the product of the n individual probabilities.

$$P(X_1, X_2, \dots, X_n) = \prod_{j=1}^n P(X_j) \quad (14)$$

Since $P(X_j = i) = p_i$, and from (6), we can accumulate all events $X_j = i$ together. The product of their probabilities will be $p_i^{nq_i}$. Therefore

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^k p_i^{nq_i} = 2^{n \sum_i q_i \log p_i} \quad (15)$$

Note that here and elsewhere in this paper the log function is of base 2. Since the exponent is a monotone function

$$\max_{P \in \Theta_i} P(X_1, X_2, \dots, X_n) = 2^{n \max_{P \in \Theta_i} \sum_i q_i \log p_i} \quad (16)$$

Observe that in the solution to Problems 1 and 2 in the beginning of this subsection, we found the probabilities yielding the maximal value in the exponent of (16) for hypotheses 1 and 2 respectively (8). Incorporating those solutions, we get a new formulation for the test

$$\frac{\max_{P \in \Theta_1} P(X_1, X_2, \dots, X_n)}{\max_{P \in \Theta_2} P(X_1, X_2, \dots, X_n)} = \frac{2^{n \sum_i q_i \log \hat{p}_i}}{2^{n \sum_i q_i \log \tilde{p}_i}} = 2^{n \sum_i q_i \log \frac{\hat{p}_i}{\tilde{p}_i}} > A \quad (17)$$

Or, since the exponent is a monotone function

$$n \sum_i q_i \log \frac{\hat{p}_i}{\tilde{p}_i} > \log A \quad (18)$$

In substituting (10), (11) and (12) into the above, note that $\hat{p}_i = \tilde{p}_i \quad \forall i > 2$, therefore, the test is

$$\begin{cases} n(q_1 - q_2) \log(1 + \delta) & \text{if } q_1 \geq (1 + \delta)q_2 \\ nq_1 \log \frac{(2 + \delta)q_1}{q_1 + q_2} + nq_2 \log \frac{(2 + \delta)q_2}{(1 + \delta)(q_1 + q_2)} & \text{otherwise} \end{cases} > T_1 \quad (19)$$

where $T_1 = \log A$. The test can be formulated in accumulator count terms instead of empiric probability terms (6):

$$\begin{cases} (n_1 - n_2) \log(1 + \delta) & \text{if } n_1 \geq (1 + \delta)n_2 \\ n_1 \log \frac{(2 + \delta)n_1}{n_1 + n_2} + n_2 \log \frac{(2 + \delta)n_2}{(1 + \delta)(n_1 + n_2)} & \text{otherwise} \end{cases} > T_1 \quad (20)$$

In the Hough algorithm this means that whenever the stopping condition is evaluated (after one or a few edge elements completed their votes), one should compare the two highest peaks in the accumulator array according to (20). If the test succeeds, the sampling is stopped and the highest accumulator is decided upon, otherwise sampling is continued. Note that exhaustive scans of the accumulator array are not needed in order to determine the highest peaks; they can be identified while voting [9, 33]. Note also that no initial period of sample gathering is necessary, i.e. condition (20) could be checked right from the first samples.

4.2 One Sided Alternative Type Test

Suppose the problem were to select the maximum of a binomial distribution with $P(X=1) = p$, and $P(X=2) = (1-p)$. The problem would be equivalent to the composite hypotheses testing problem with $H_1 : p > 0.5$ and $H_2 : p < 0.5$. This problem can be solved via a one sided alternative test. Let us define an indifference zone which is a reduction of the indifference zone in (8) to the binomial case. Hypothesis H_1 is the set of distributions for which $p \geq (1 + \delta)(1 - p)$ and hypothesis H_2 is the set of distributions for which $(1 - p) \geq (1 + \delta)p$. The resulting indifference zone is $\left[\frac{1}{2 + \delta}, \frac{1 + \delta}{2 + \delta} \right]$.

If in the binomial maximum selection problem we adopt the convention about the indices of the empiric distribution so that $q_1 \geq q_2$, the SPRT corresponding to the one sided alternative test is

$$\frac{\left(\frac{1 + \delta}{2 + \delta}\right)^{nq_1} \left(\frac{1}{2 + \delta}\right)^{nq_2}}{\left(\frac{1 + \delta}{2 + \delta}\right)^{nq_2} \left(\frac{1}{2 + \delta}\right)^{nq_1}} = (1 + \delta)^{n(q_1 - q_2)} > A \quad (21)$$

Or

$$n(q_1 - q_2) \log(1 + \delta) > \log A \quad (22)$$

Let us return to our problem, which is the multihypothesis equivalent of the binomial problem solved above. Viewing each of the $k(k-1)$ likelihood ratios as a decision problem between two hypotheses, we can treat it as the binomial maximum selection problem described above. Indeed, if we want to contest two specific events, we are not interested in the occurrences of any other event. Ignoring all other events we can view the source as a binomial source.

In this case too we do not check all the $k(k-1)$ ratios. It is sufficient to check only the ratio contesting the most popular, and second most popular hypotheses. Hence, when q_1 and q_2 are the empirical probabilities of the most and second most popular events respectively, the test has the same formulation as (22).

The sequential test derived above has an interesting property. Generally, as in the case of the test derived in the previous subsection, sequential tests of composite hypotheses may depend on the indifference zone parameter (δ in our notation) in a complex manner. δ must be small enough since the good features of the test are valid only outside the indifference zone. On the other hand, a smaller indifference zone causes larger average number of samples in the sequential tests. Here we do not specify the indifference zone explicitly. Since A is an unknown parameter anyway, we define $T_2 = \frac{\log A}{\log(1+\delta)}$. In this way, the stopping rule is based simply on the difference in the counts (6) of the most and second most popular events.

$$n_1 - n_2 > T_2 \quad (23)$$

For the above OSAT stopping rule as for the GLRT stopping rule no initial period of sample gathering is necessary. (Note that a similar test was considered by Alam [1], without theoretical justification and not in the context of Wald's theory).

5 Experimental Results

In this section we contest the two stopping rules suggested in the previous section against each other, and against other stopping rules suggested in the literature for the problem of detecting the most probable event of a discrete distribution. In the first subsection we present small scale simulations, in which we tested the stopping

rules on several discrete distributions having 10 events. In the second subsection we present results from applying the different stopping rules to real images. To allow comparison in the context of an actual application in which the Probabilistic Hough Transform has already proved itself [33], the images are of dishes on cafeteria trays. The goal was to reliably detect the location of plates in the images, using a minimal poll size.

5.1 Small Scale Simulations

In this subsection we present results from small scale simulations. The term small scale relates to the number of events (10 in this case). We contested five different stopping rules. The task on which the stopping rules were tested was to detect the most probable event of an unknown discrete distribution. An outcome is considered correct if it is either the most probable event or a different event which is nearly as probable. If p is the probability of the most probable event of the distribution, then any event with probability q , such that $p < (1 + \delta)q$, is considered a correct outcome of the test. This interpretation of correct outcome is consistent with the definition of the indifference zone (8). Three different types of distributions were used as sources for the tests:

1. An “almost uniform” source in which all but one event have the same probability q . This one event has a probability $p = (1 + \delta)q$. This distribution is the least favorable distribution for the task of determining the most probable event of a discrete distribution with an indifference zone determined by δ [12].
2. An “exponentially decaying” distribution in which the probability p_i of event i satisfies $p_{i+1} = \varepsilon p_i$ for $\varepsilon < \frac{1}{1 + \delta}$. As in the “almost uniform” distribution, the “exponentially decaying” distribution has only one correct outcome, however, in the “exponentially decaying” distribution there are only few good candidate events.

The “exponentially decaying” distribution is much more realistic than the least favorable, “almost uniform” distribution. We hope that this distribution provides a good model for vote distributions with a single dominant peak in the Hough space. Those vote distributions are interesting since they model distributions for which the stopping rules were optimized, i.e. those distributions are within an hypothesis and not in the indifference zone.

3. A “step like” distribution in which a few events (three) have high probabilities, and the rest (seven) have exponentially decaying probabilities. The events with

the high probabilities differ in their probabilities only slightly so that they will all be accepted as correct hypotheses. All other events will be rejected.

We hope that the “step like” distribution provides a good model for the common case of vote distributions with multiple dominant peaks. In that sense, the “step like” distribution seems to be the most realistic distribution we examine in the simulations. Note that since several events are acceptable, the distribution of this source is in the indifference zone.

The probabilities of the specific sources used to obtain the graphs appear in Table 1. Also presented are the possible correct outcomes, relating to the indifference parameter, which was selected to be $\delta = 0.11$.

	Figure 1a	Figure 2b	Figure 3c
p_0	0.109793	0.17297	0.175
p_1	0.0989119	0.150484	0.166667
p_2	0.0989119	0.130921	0.158333
p_3	0.0989119	0.113901	0.140774
p_4	0.0989119	0.0990939	0.106917
p_5	0.0989119	0.0862117	0.0812026
p_6	0.0989119	0.0750042	0.0616729
p_7	0.0989119	0.0652537	0.0468402
p_8	0.0989119	0.0567707	0.0355748
p_9	0.0989119	0.0493905	0.0270188
$p_0/(1+\delta)$	0.0989126	0.155829	0.157658
Correct Answers	0	0	0,1,2

Table 1: Source probabilities for Figure 1.

Figure 1 consists of three sets of graphs. Each set corresponds to tests performed on samples taken independently from a source having one of the above distributions and contains five graphs corresponding to the five stopping rules described below. Each graph describes the error rate in an interval around 5%, versus the average number of samples needed to obtain the performance using the corresponding stopping rule. Every graph point is the outcome of 12,000 full sequential decision processes performed using statistically independent data. The error rate is the number

of erroneous decisions divided by the total number of decisions (12,000). Each full sequential decision process requires many independent samples. The number of samples for each graph point amounts therefore, to 12,000 times the respective average number of samples.

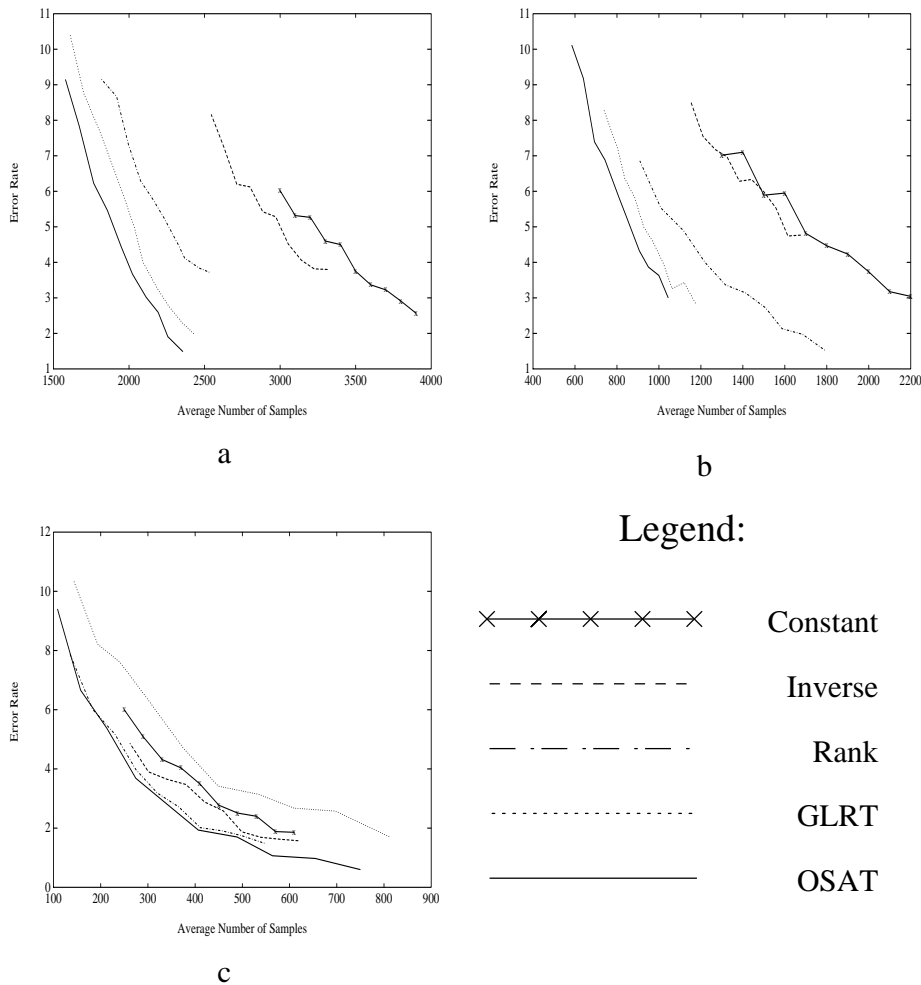


Figure 1: Small scale simulation results for the a) “Almost Uniform” source, b) “Exponentially Decaying” source, c) “Step Like” source.

The five stopping rules are:

1. Constant or fixed sample size test (which can be viewed as a degenerate sequential test). Sampling stops when the number of samples reaches a certain number, which is the parameter of the stopping rule. Graphs for this test are denoted in the figures by a solid line with x marks.
2. Inverse Sampling [6] is a test that stops when the most popular event reaches a certain number, which is the parameter of the stopping rule. Graphs for this

test are denoted in the figures by a dashed line.

3. Rank stability test [33], is a test that stops when the time since the last change in the location of the most popular event, reaches a certain number. This number is the parameter of the stopping rule. Graphs for this test are denoted in the figures by the dot dash pattern.
4. GLRT is the stopping rule presented in Section 4.1. It stops when condition (20), dependent on the votes for the most popular and the second most popular events, is satisfied. Threshold T_1 in (20) serves as the parameter of the stopping rule. Graphs for this test are denoted in the figures by a dotted line.
5. OSAT is the stopping rule presented in Section 4.2 (also applied in [1]). The test stops when the sample difference between the most popular and the second most popular events reaches a certain number, which is the parameter of the stopping rule T_2 in (23). Graphs for this test are denoted in the figures by a solid line.

Each of the five stopping rules has a parameter that determines the average number of samples and the error rate. A higher value of that parameter causes the algorithm to run longer, thereby also lowering the error rate. Varying the parameter one can tune a stopping rule to yield an error rate within a predefined interval.

In the presented results, as well as in all other results we obtained, the consistent feature is that the OSAT strategy yields the best results. The GLRT strategy is second best for the “almost uniform” and the “exponentially decreasing” distributions, but is the worst strategy for the “step like” distribution. This could be related to the fact that the “step like” distribution is in the indifference zone. As already mentioned, the performance of stopping rules is not specified for distributions in the indifference zone.

5.2 Results on Images

In this subsection we present results obtained by applying the different stopping rules for the Probabilistic Hough Transform to real images. The images are of dishes on cafeteria trays. The goal is to detect plates of a specified size in the images. Two examples of the 16 (512×512 pixel) image database are presented in Figure 2,

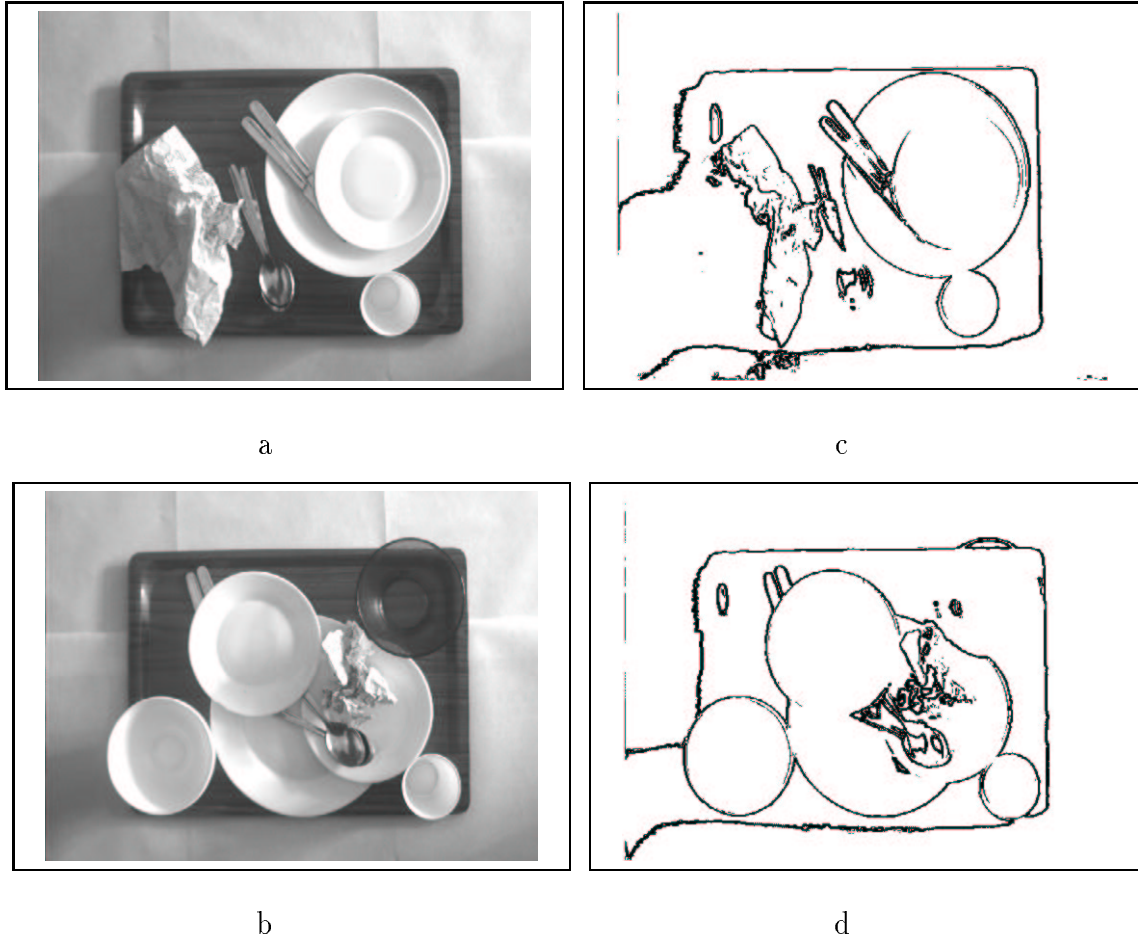


Figure 2: a,b) Two examples from the image database, c,d) Corresponding edge images. The large plates are to be detected.

together with corresponding edge images. It should be noted that 3 of the 16 images in the database included *two* plates of the specified size.

In the preprocessing stage of the algorithm, both the gradient size and its direction are estimated via Sobel operators. Pixels with a large gradient are considered as edge pixels. The average number of edge pixels in the 16 images is 11,000. Since the radius of the plates is known, the parameter space spans possible plate center locations.

The locus of centers of all the circles with a predefined radius that pass through a given point in the plane, is a circle of the same radius, centered at the given point. In a “one to many” Hough Transform, each point would therefore vote for all accumulators the circle passes through. However, incorporating edge direction data, the direction to the center can be estimated. For bright plates on a dark background

this is the direction opposite to the gradient. Hence, each edge pixel votes for the accumulator positioned in the direction opposite to the gradient, at a distance equal to the predetermined radius of the expected circles. This voting scheme provides a sequence of independent samples as required by the above analysis.

In the algorithm used in these experiments each edge pixel supplies a single vote. It is well known that sophisticated voting patterns that provide compensation for edge location and direction errors can improve the performance of the Hough Transform. But Probabilistic Hough algorithms are used in extremely time sensitive applications. It is not obvious whether time that can be used for increasing the poll size should have been spent on complicated voting schemes instead.

A 64×64 accumulator array was used in these experiments. Selecting the “right” size for accumulator arrays is a fundamental problem in the design of Hough algorithms. For best performance, significant issues beside the obvious complexity vs. resolution tradeoff should be taken into consideration. See, e.g. [28, 21, 13, 15]. These aspects of the design seem however not to be directly related to the comparative evaluation of stopping rules for the Probabilistic Hough Transform.

The five graphs in Figure 3 are of the error rate versus average number of samples for each of the five stopping rules described in the previous subsection. In these experiments the indifference zone parameter was taken to be approximately $\delta = 0.89$. Unlike the small scale simulations described in the previous subsection, here there is no unique unknown source. The source for the algorithms whose results are presented in Figure 3 was alternating between the 16 images of the database. This fact has been found to cause a large variance in the number of samples.

Each point in the graphs of Figure 3 is the average of 32,000 independent tests. We performed 2000 independent tests on each of the 16 images, and the graphs average over all the images of the database. Please note that each test we performed was a fresh test, polling its samples randomly from all the edge pixels, totally independent of any previous test performed on the same image in the past.

The graphs in Figure 3 indicate an advantage for the OSAT and GLRT based stopping rules over the other stopping rules. All the adaptive stopping rules are significantly better than the fixed sample size policy.

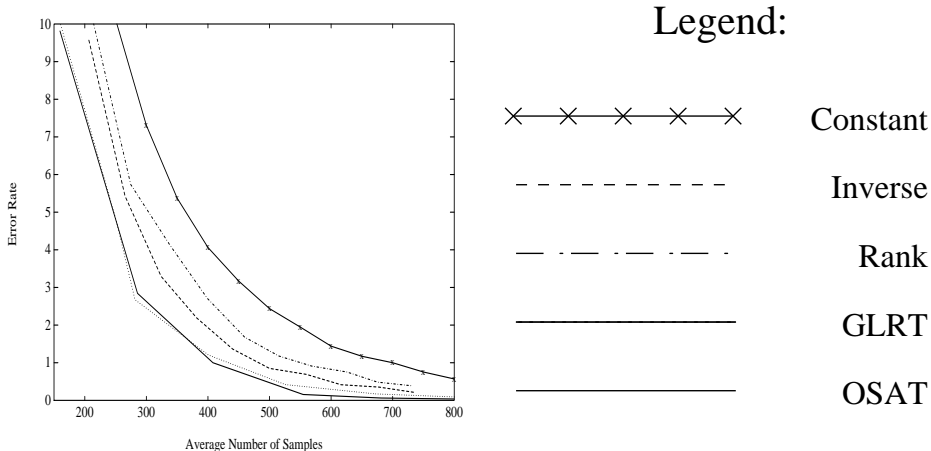


Figure 3: Results for applying the stopping rules to Real Images.

6 Technical Considerations

Computational Overheads: The experiments show that the suggested stopping rules save 30-50% of the samples needed for a decision having a specified error rate with respect to using a fixed sample size. In order to obtain the actual computation time saving, the overhead of evaluating the stopping condition has to be considered. This overhead is relatively small due to the following reasons.

The suggested stopping rules depend only on the counts in the two most popular accumulators. These can be easily updated while voting at little additional cost. Note that since the poll size is usually small, this operation is overall cheaper than scanning the accumulator array for the maximum after the voting stage, and should therefore be preferred in all probabilistic Hough Transforms, regardless of whether the poll size is fixed or adaptive.

After each vote, the value of the incremented accumulator is compared to the value of the second most popular accumulator. If the currently updated count is lower, as is usually the case, the data for the stopping condition has not changed, nothing has to be done and sampling continues. Otherwise, the two top ranks of the accumulators have to be updated. It is however necessary to actually evaluate the stopping condition only if the top rank accumulator received the current vote.

We conclude that in most sample cycles the overhead of adaptive stopping is a single register comparison operation. From time to time few additional comparisons have to be performed (updating the two top ranks). The number of times the

stopping condition itself has to be evaluated is equal to the value of the most popular accumulator at the stopping time, which was about 1% of the total number of points sampled *up to that time* in the experiments with real images. Evaluation of the OSAT stopping condition is then a mere subtraction. Straightforward evaluation of the GLRT stopping condition is more difficult, but can be easily done using a look up table.

Multiple Objects: The goal we specify for the Probabilistic Hough Transform is to obtain the same maximum as if the full Hough Transform had been accumulated and analyzed. In practice the goal may sometimes be to obtain the same *set* of dominant maxima. An attempt to devise adaptive stopping rules for detecting several objects (i.e., dominant maxima) is described in [33] for a predefined as well as for an unknown number of objects. However, since the definition of dominant maxima must be application dependent, using this concept in theoretical analysis is problematic.

Various strategies for adaptive stopping can be suggested when several objects should be detected. In the spirit of [33], one might observe the dominance of a set of maxima (whose cardinality may be predefined or unknown) with respect to other maxima. This parallel approach leads to simultaneous detection of multiple objects. An alternative serial approach is to recognize objects in the image one at a time. The possible extension of the adaptive stopping rules suggested in this paper to the detection of multiple objects is intuitively appealing and straightforward, but seems to be theoretically nontrivial. It is an interesting topic for future research at the statistical and algorithmic levels.

Tuning Parameters: As mentioned in the development of the stopping rules, it is fundamentally difficult for any algorithm to detect the global maximum in the parameter space in the presence of other maxima that are nearly as high. This observation is the basis for the concept of an indifference zone, and its parameter δ (8). The performance of the stopping rules is optimized for detecting δ -distinct maxima. As long as the detection of a single maximum is required, the presence of δ -close maxima will lead to larger polls on one hand, and to possible mismatches on the other¹.

Only the system designer using this algorithm in an application can specify to what extent a maximum that is slightly lower than the global maximum is an accept-

¹In order to reduce computation time in the presence of δ -close maxima, one must be willing to abandon the strict requirement for a single maximum and allow multiple maxima as suggested in [33].

able outcome of the algorithm. Thus, δ is not a parameter that should be tuned, but a technical specification that must be derived from the requirements of the application. Nevertheless, the determination of δ is not a trivial task. Hence, it is pleasing to note that in the OSAT stopping rule the explicit specification of δ is not necessary. There δ is part of the threshold T_2 that has to be tuned to the application anyway.

Consider the determination of the threshold in the stopping rules. This refers to the fixed number of samples in the fixed sample size rule, as well as to the thresholds T_1 and T_2 of the GLRT or OSAT based rules in (20) and (23) respectively. In the statistical literature [1, 4, 6, 23], the thresholds are tuned by Monte Carlo simulations to the worst case distribution. This allows to establish lower bounds on the performance, but generally leads to poll sizes that are too large with respect to error rate in actual applications. From an engineering point of view, the threshold value has to be tuned in the application according to the characteristics of the images and the required detection error rate.

7 Conclusions

We have addressed the main open theoretical issue concerning the Probabilistic Hough Transform, namely the determination of an adaptive stopping rule for the voting process. Up to the present paper most of the rules were fixed size rules, and the few adaptive rules were essentially heuristic, supported by experimental evidence. In this paper we revisited the statistical theory of sequential analysis, and applied it to the stopping rule problem.

Sequential analysis gives optimal solutions to a large family of problems, however, the determination of an optimal stopping rule for the Probabilistic Hough Transform belongs to a different family of problems, for which there is no known optimal solution. Still, even for this problematic family of tasks, there are techniques which are considered to give relatively good solutions, either because they can sometimes be shown to be asymptotically optimal, or because they are immediate generalizations of techniques that are optimal for related problems. We have applied two sequential techniques to obtain two novel stopping rules for the Probabilistic Hough Transform.

In Section 5, we contested the two proposed stopping rules against each other and against a set of alternative stopping rules, suggested by either statistical or Probabilistic Hough literature. The stopping rules were contested in two different setups:

The first was a Monte-Carlo simulation similar in both scale and spirit to simulations performed in the statistical literature. The second set of experiments was in the context of an actual application with data coming from images of dishes on a cafeteria tray. In this realistic task the input images yield a varying large scale source (4096 accumulators as opposed to 10).

It has been shown that adaptive termination of voting considerably reduces the average number of votes needed to achieve a specific error rate. The results of the simulations and experiments clearly demonstrate that the one sided alternative type test (OSAT) stopping rule and in most cases also the generalized likelihood ratio test (GLRT) stopping rule perform significantly better than previously known stopping rules. The OSAT rule performed best in the simulations. The sequential GLRT rule performed nearly as well as the OSAT rule in the experiments with real images.

Acknowledgment: We thank Gidi Goren and Yoav Sadot for performing the experiments with images. We also thank Prof. Yosi Rinott for introducing us to some of the statistical literature. This research was supported in part by the Ministry of Science and Arts and by the Ollendorf center of the Technion department of Electrical Engineering.

References

- [1] K. Alam, “On Selecting the Most Probable Category”, *Technometrics*, Vol. 13, pp. 843–850, 1971.
- [2] P. Armitage, “Sequential Analysis with More than Two Alternative Hypotheses, and its Relation to Discriminant Function Analysis”, *J. Roy. Stat. Soc. B*, Vol. 12, pp. 137–144, 1950.
- [3] D. H. Ballard, “Generalizing the Hough Transform to Detect Arbitrary Shapes”, *Pattern Recognition*, Vol. 13, pp. 111–122, 1981.
- [4] R. E. Bechhofer, S. Elmaghraby, and N. Morse, “A Single-Sample Multiple-Decision Procedure for Selecting the Multinomial Event Which has the Largest Probability”, *Ann. Math. Stat.*, Vol. 30, pp. 102–119, 1959.
- [5] J. R. Bergen and H. Shvayster, “A Probabilistic Algorithm for Computing Hough Transforms”, *J. of Algorithms*, Vol. 12, pp. 639–656, 1991.
- [6] T. Cacoullos, and M. Sobel, “An Inverse-Sampling Procedure for Selecting the Most Probable Event in a Multinomial Distribution”, in *Multivariate Analysis*, Ed: P. R. Krishnaiah, Academic Press, New York, 1966.
- [7] R. O. Duda, and P. E. Hart, “Use of the Hough Transformation to Detect Lines and Curves in Pictures”, *Comm. ACM*, Vol. 15, pp. 11–15, 1972.
- [8] B. Eisenberg, “Multihypothesis Problems”, in *Handbook of Sequential Analysis*, Eds: B. K. Ghosh, and P. K. Sen, Marcel Dekker, New York, 1991.
- [9] G. Gerig, “Linking Image-Space and Accumulator-Space: A New Approach for Object Recognition”, *Proc. IEEE 1st International Conf. on Computer Vision*, pp. 112–117, London, 1987.
- [10] J. Illingworth and J. Kittler, “A Survey of the Hough Transform”, *Computer Vision Graphics and Image Processing*, Vol. 44, pp. 87–116, 1988.
- [11] H. Kälviäinen, P. Hirvonen, L. Xu, and E. Oja, “Probabilistic and Non-probabilistic Hough Transforms: Overview and Comparisons”, to appear in *Image and Vision Computing*.
- [12] H. Kesten, and N. Morse, “A Property of the Multinomial Distribution”, *Ann. Math. Stat.*, Vol. 30, pp. 120–127, 1972.
- [13] N. Kiryati, and A. M. Bruckstein, “Anti-aliasing the Hough Transform”, *CVGIP: Graphical Models and Image Processing*, Vol. 53, pp. 213–222, 1991.
- [14] N. Kiryati, Y. Eldar, and A. M. Bruckstein, “Probabilistic Hough Transform”, *Pattern Recognition*, Vol. 24, pp. 303–316, 1991.

- [15] N. Kiryati, and M. Soffer, “Guaranteed Convergence of the Hough Transform”, in *Proc. NSF/ARPA Workshop on Performance vs. Methodology in Computer Vision*, pp. 104–110, Seattle, June 1994.
- [16] T. L. Lai, “Asymptotic Optimality of Generalized Sequential Likelihood Ratio Tests in Some Classical Sequential Testing Problems”, in *Handbook of Sequential Analysis*, Eds: B. K. Ghosh, and P. K. Sen, Marcel Dekker Inc., New York, 1991.
- [17] V.F. Leavers, “Dynamic Generalized Hough Transform: Its Relationship to the Probabilistic Hough Transform and an Application to the Concurrent Detection of Circles and Ellipses”, *CVGIP: Image Understanding*, Vol. 56, pp. 381–398, 1992.
- [18] V. F. Leavers, “Which Hough Transform?”, *CVGIP: Image Understanding*, Vol. 58, pp. 250–264, 1993.
- [19] E. L. Lehmann, *Testing Statistical Hypotheses*, John Wiley & Sons, 1959.
- [20] G. Lorden, “Likelihood Ratio Tests for Sequential k-Decision Problems”, *Ann. Math. Stat.*, Vol. 43, pp. 1412–1427, 1972.
- [21] W. Niblack, and D. Petkovic, “On Improving the Accuracy of the Hough Transform”, *Machine Vision and Applications*, Vol. 3, pp. 87–106, 1990.
- [22] F. O’Gorman, and M. B. Clowes, “Finding Picture Edges Through Collinearity of Feature Points”, *IEEE Trans. on Computers*, Vol. 25, pp. 449–456, 1976.
- [23] J. T. Ramey, and K. Alam, “A Sequential Procedure for Selecting the Most Probable Multinomial Event”, *Biometrika*, Vol. 66, pp. 171–173, 1979.
- [24] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, 1972.
- [25] G. Roth and M.D. Levine, “Random Sampling for Primitive Extraction”, *Proc. Robust Computer Vision Workshop*, pp. 352–366, Seattle, 1990.
- [26] G. Simons, “Lower Bounds for Average Sample Number of sequential Multihypothesis Tests”, *Ann. Math. Stat.*, Vol. 38, pp. 1343–1364, 1967.
- [27] M. Sobel, and A. Wald, “A Sequential Decision Procedure for Choosing One of Three Hypotheses Concerning the Unknown Mean of a Normal Distribution”, *Ann. Math. Stat.*, Vol. 20, pp. 502–522, 1949.
- [28] T. M. Van Veen, and F. C. A. Groen, “Discretization Errors in the Hough Transform”, *Pattern Recognition*, Vol. 14, pp. 137–145, 1981.
- [29] A. Wald, *Sequential Analysis*, Wiley, New York, 1947.
- [30] A. Wald, and J. Wolfowitz, “Optimal Character of the Sequential Probability Ratio Test”, *Ann. Math. Stat.* Vol. 19, pp. 326–339, 1948.
- [31] L. Xu, E. Oja and P. Kultanen, “A New Curve Detection Method: Randomized Hough Transform (RHT)”, *Pattern Recognition Letters*, Vol. 11, pp. 331–338, 1990.

- [32] L. Xu, and E. Oja, “Randomized Hough Transform (RHT): Basic Mechanisms, Algorithms, and Computational Complexities”, *CVGIP: Image Understanding*, Vol. 57, pp. 131–154, 1993.
- [33] A. Ylä-Jääski, and N. Kiryati, “Adaptive Termination of Voting in the Probabilistic Circular Hough Transform”, *IEEE PAMI*, Vol. 16, pp. 911–915, 1994.