PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

A probability of error-constrained sequential decision algorithm for data-rich automatic target recognition

Irwin O. Reyes, Michael D. DeVore, Peter A. Beling, Barry M. Horowitz

Irwin O. Reyes, Michael D. DeVore, Peter A. Beling, Barry M. Horowitz, "A probability of error-constrained sequential decision algorithm for data-rich automatic target recognition," Proc. SPIE 7696, Automatic Target Recognition XX; Acquisition, Tracking, Pointing, and Laser Systems Technologies XXIV; and Optical Pattern Recognition XXI, 769615 (13 May 2010); doi: 10.1117/12.858293

SPIE.

Event: SPIE Defense, Security, and Sensing, 2010, Orlando, Florida, United States

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 12/5/2017 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

A Probability of Error-Constrained Sequential Decision Algorithm for Data-Rich Automatic Target Recognition

Irwin O. Reyes[a1], Michael D. DeVore[b], Peter A. Beling[a], Barry M. Horowitz[a]
[a]Dept. of Systems and Information Engineering, University of Virginia, Charlottesville, VA 22903;
[b]Barron Associates, Inc., 1410 Sachem Place, Suite 202, Charlottesville, VA 22901

ABSTRACT

This paper illustrates an approach to sequential hypothesis testing designed not to minimize the amount of data collected but to reduce the overall amount of processing required, while still guaranteeing pre-specified conditional probabilities of error. The approach is potentially useful when sensor data are plentiful but time and processing capability are constrained. The approach gradually reduces the number of target hypotheses under consideration as more sensor data are processed, proportionally allocating time and processing resources to the most likely target classes. The approach is demonstrated on a multi-class ladar-based target recognition problem and compared with uniform-computation tests.

Keywords: sequential hypothesis testing, LADAR, target recognition

1. INTRODUCTION

Given the capability of modern high-resolution sensors to quickly generate very large collections of observations[1], there is much ongoing research in developing methods to automatically derive useful inferences from these raw data sets. In the realm of 3D measurement data, and particularly those collected from fine-resolution sensors like laser-radar (LADAR) platforms, the literature reveals numerous thrusts to make practical use of that raw data; for instance, determining an object's pose relative to the sensor, detecting targets of interest occluded by foliage, distilling a scene down to only certain selected items, and recognizing observed subjects as members of known classes.[2-5]

Research efforts into systems that automatically classify and recognize objects from 3D measurement data have produced, among many other methods, statistical approaches to this particular problem. Rather than treating data sets as the outcome of deterministic measurement processes, statistical algorithms treat measurements as random variables under particular distributions in order to model observation uncertainty.[4] One such approach revolves around the use of a minimum probability of error decision rule to compute the likelihood that a set of 3D observations (collectively known as a point cloud) arose from any one of a number of known possible targets. Under this decision rule, however, likelihoods must be calculated for every component point in the cloud against each candidate target hypothesis.[5] Applying this scheme to the practical application of recognizing real-world vehicles, an individual point-wise likelihood calculation necessitates a surface integral over a non-trivial figure. The number of these costly computations scale linearly with each data point processed and each target hypothesis tested. Thus, it is very desirable to reduce the number of times new likelihoods are evaluated.

To this end, one way of reducing the number of likelihoods computed is to sequence evaluations in a certain order such that unlikely target hypotheses are discounted as quickly as possible, thereby narrowing the field of plausible candidate hypotheses to further investigate. Employing offline-calculated 2-hypothesis likelihood vs. accuracy-of-recognition performance estimates, significant computational savings may be achieved by reducing an M-hypothesis uniform-calculation strategy into a series of 2-hypothesis tests arranged in a progressing single-elimination tournament tree.[6] By establishing a desired pair-wise recognition accuracy level and seeding this tournament whereby hypothesis pairs asserting highly dissimilar targets are evaluated first, relatively few likelihoods are needed (as determined by the performance estimates) to eliminate the most unlikely targets, while the bulk of the computations are performed later on to select from a reduced number of likely target hypotheses. An "anytime" capability – that is, the ability for the algorithm to return a result after being interrupted – is achieved by progressively increasing the pair-wise accuracy level and executing multiple passes over the tree as time and computational resources allow.

[1]Email: irwinreyes@virginia.edu

Automatic Target Recognition XX; Acquisition, Tracking, Pointing, and Laser Systems Technologies XXIV; Optical Pattern Recognition XXI, edited by F. A. Sadjadi, A. Mahalanobis, S. L. Chodos, W. E. Thompson, D. P. Casasent, T.-H. Chao, Proc. of SPIE Vol. 7696, 769615 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.858293

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 12/5/2017 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

An alternate approach is to use sequential hypothesis testing. Much like the aforementioned tree tournament strategy, sequential hypothesis testing aims to reduce the number of target hypotheses under consideration as additional data is collected and processed. This approach, however, differs from the tree in that rather than evaluating competing hypotheses in a series of pair-wise tests, a sequential test explicitly discards hypotheses whose probability of an incorrect rejection (given that it is indeed true) has fallen below some specified significance threshold.[7] While previous works have explored the use of sequential hypothesis testing in the automatic target recognition domain, those efforts focus reducing the amount of data necessary to reach sound inferences.[8] As such, these one-pass data-limited developments address issues besides those encountered in data-rich applications.

This paper thus demonstrates an anytime sequential hypothesis testing algorithm employed in the context of an M-class target recognition problem using 3D LADAR point cloud data with the aim of reducing computation. This continues previous work in developing anytime algorithms with prescribed maximum allowable error rates. The research further expands upon this by offering a *reject option* which allows this particular approach to assert the null hypothesis should all of the alternate target hypotheses fail to sufficiently account for the observed data, as would be the case in encountering an unknown or unanticipated target. The proposed algorithm accomplishes this by iteratively narrowing down the field of candidate target hypotheses using significance testing. Significance testing compares the distributions of computed log-likelihoods against those expected from each hypothesis under the assumption that the particular hypothesis at hand is true. An anytime capability is achieved explicitly in iterative elimination process.

The remaining sections of this paper are structured as follows. Section 2 details the sequential hypothesis testing algorithm as it is used in this research with a statistical 3D object recognition process. This section also discusses methods of using Gaussian and Johnson family distributions to estimate expected log-likelihood distributions under true hypotheses for use in significance testing. Section 3 illustrates the simulation and data-processing methods employed in evaluating this new algorithm. Section 4 presents quantitative comparisons between the sequential hypothesis testing process and previously-developed uniform-computation anytime algorithms under varying available target hypotheses. Also covered in this section is a demonstration of the new reject option utilized when the algorithm is faced with an unknown target. Conclusions and future work are presented in Section 5.

2. SEQUENTIAL HYPOTHESIS TESTING ALGORITHM

The goal of a hypothesis testing procedure is to choose from among several possible hypotheses the one that best accounts for an observed set of data, if any. A fixed sample size procedure makes this choice after processing all available data. A sequential procedure, on the other hand, processes a variable amount of data, making a decision at each stage as to whether the collection of additional data is warranted.

More formally, suppose that a measurement process yields a sequence of random variables X1, X2, ..., drawn independently from some distribution, which could be one of a $M < \infty$ known distributions or some other unknown distribution. The possibility that the data were drawn from the mth distribution is referred to as hypothesis m, denoted H_m. The possibility that the data was drawn according to some other unknown distribution is referred to as the null hypothesis, denoted H_0.

A fixed sample size test consists of a pair (N_F, ϕ_F), where $N_F < \infty$ is an integer constant indicating the number of sample observations to be processed, and ϕ_F is a decision rule (i.e., a function) mapping the collection of observations $X_1, X_2, \ldots, X_{N_F}$ to one of the M distributions. The decision rule has the following interpretation: If $\phi_F(X_1, X_2, \ldots, X_{N_F}) = m$, then H_m is asserted to be the best explanation for the observations. Of course, this assertion may be incorrect, and the probability that the observations from class j are incorrectly declared to be from class m is $P_{j,m} = \Pr[\phi_F = m \mid H_j \text{ is true}]$. The class conditional probability of error is $\varepsilon_j = \Pr[H_j \text{ not chosen} \mid H_j \text{ not true}] = \sum_{m \neq j} P_{j,m}$.

In a sequential test the value N_S is itself a random variable, being a function of the observations. That is, a sequential test is a pair (N_S, ϕ_S), where $N_S(X_1, X_2, \ldots)$ is called a stopping rule, and ϕ_S, called the final decision rule, maps observations $X_1, X_2, \ldots, X_N$ to one of the M distributions. As with a fixed sample size test, the various error probabilities $P_{i,j}$ and ε_j are values of interest. Additionally, it is useful to determine the distribution of N_S when hypothesis H_j is true. Most sequential tests are characterized in terms of the average sample number (ASN), defined as $E[N_S]$.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 12/5/2017 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

For the two-hypothesis case, the sequential probability ratio test (SPRT) has been shown to minimize the ASN over all tests that have class conditional probabilities of error no greater than some specified ε_1 and ε_2, regardless of which hypothesis is true.[7] At each stage, the SPRT calculates the likelihood ratio of all available data and compares the result against two threshold values. If the likelihood ratio is smaller than the minimum threshold or greater than the maximum threshold, the corresponding hypothesis is selected. If the ratio lies between the two thresholds, the data are taken to be ambiguous, and additional data is collected. The test then repeats with additional data.

Many authors have reported on direct multi-hypothesis extensions of the basic SPRT[9-12]. Unfortunately, it can also be shown that this type of optimality cannot extend to problems with more than two hypotheses. That is, no sequential test can minimize the ASN, subject to upper bounds on the probabilities of error, simultaneously across all true hypotheses.[13] It has recently been shown that the multihypothesis SPRT (MSPRT), which is similar to earlier SPRT extensions, is asymptotically optimal for arbitrarily small error probabilities, ε_j.[14-16] Like the SPRT, all these tests compute the likelihood of the entire data collection under each hypothesis at every stage.

By dropping unlikely hypotheses from further consideration, the number of hypotheses, and thus the amount of processing for each new data sample, decreases with each stage. Suppose that when $\mathrm{H_m}$ is true, and observed data sample X has probability density function (PDF) $\mathrm{f_m(x)}$. Given a collection of conditionally independent observed data $\mathrm{X_1, X_2, \ldots, X_N}$, denote the log-likelihood of $\mathrm{H_m}$ as the sum

$$L_m^n = \sum_{n=1}^{N} log(f_m(X_n)) \tag{1}$$

Note that $\mathrm{L_m^n}$ is a function of the observed data, and so it is itself a random variable. Let $\mathrm{F_m^n(l)}$ denote the cumulative distribution function (CDF) of $\mathrm{L_m^n}$ under the assumption that $\mathrm{H_m}$ is true. Optimal hypothesis testing algorithms are based on the fact that when $\mathrm{H_m}$ is not true, we expect the log-likelihood hypothesis $\mathrm{L_m^n}$ to be small. This implies that when $\mathrm{H_m}$ is not true, we expect $\mathrm{F_m^n(L_m^n)}$ to be close to zero.

Working the other way, assuming that we want to determine whether or not $\mathrm{H_m}$ is true, we note that if $\mathrm{F_m^n(L_m^n)} = \alpha$, then there is a 100α percent probability that by pure chance alone an even less likely sequence of data would be observed if $\mathrm{H_m}$ were in fact true. This can form the basis of a discrimination test for whether or not to drop $\mathrm{H_m}$ at stage N. For example, suppose we pick some arbitrarily small quantity $\alpha = 0.01$ and drop $\mathrm{H_m}$ at stage N from further consideration if $\mathrm{F_m^n(L_m^n)} = \alpha$. We can then be confident that there is less than a 1% chance that we have incorrectly rejected $\mathrm{H_m}$.

With this in mind, we define the sequential target classification procedure as follows.

1. Define α to be the largest tolerable false rejection rate for the problem

2. Initialize the stage number $\mathrm{n} = 0$ and let $\mathrm{M^0} = \{1,2,\ldots,\mathrm{M}\}$ be the set of target classes initially under consideration

3. Increment n by a pre-specified value, collect observation $\mathrm{X_n}$ and compute $\mathrm{L_m^n}$ for each $\mathrm{m} \in \mathrm{M}$

4. Let $\mathrm{M^n} = \{\mathrm{m} \in \mathrm{M^{n-1}} | \mathrm{F_m^n(L_m^n)} > \alpha\}$ be the set of target classes still under consideration after stage n

5. If $|\mathrm{M^n}| > 0$, report

   (a) $\hat{\mathrm{m}} = \mathrm{argmax}_{\mathrm{m} \in \mathrm{M^n}} \mathrm{L_m^n}$ as the most likely hypothesis found at the end of stage n;
   (b) $\mathrm{F_{\hat{m}}^n(L_{\hat{m}}^n)}$ as the significance of that hypothesis; and
   (c) $\mathrm{M^n}$ as the set of feasible alternatives.

6. If $|\mathrm{M^n}| = 0$, report that all of the known target hypotheses have been rejected

7. If $|\mathrm{M^n}| > 1$ go to step 3, otherwise terminate.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 12/5/2017 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Narrowing down the field of viable hypotheses at each stage effectively reduces the number of likelihoods to compute as a direct result of fewer hypotheses to test. A trade-off exists, however, in the form of increased data consumption compared to a fixed-sample uniform-calculation hypothesis test. That is, where a uniform-calculation strategy may test a small number of observations against a large collection of hypotheses, a sequential algorithm more readily tests a larger number of observations against a small group of hypotheses on each iteration.

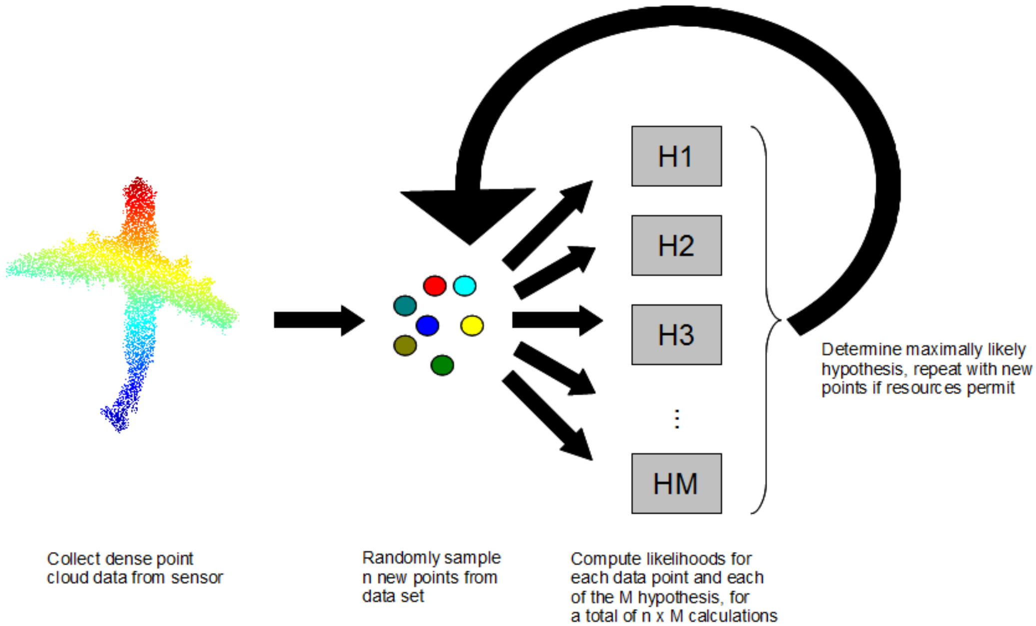

Figure 1. Naïve uniform hypothesis testing strategy.

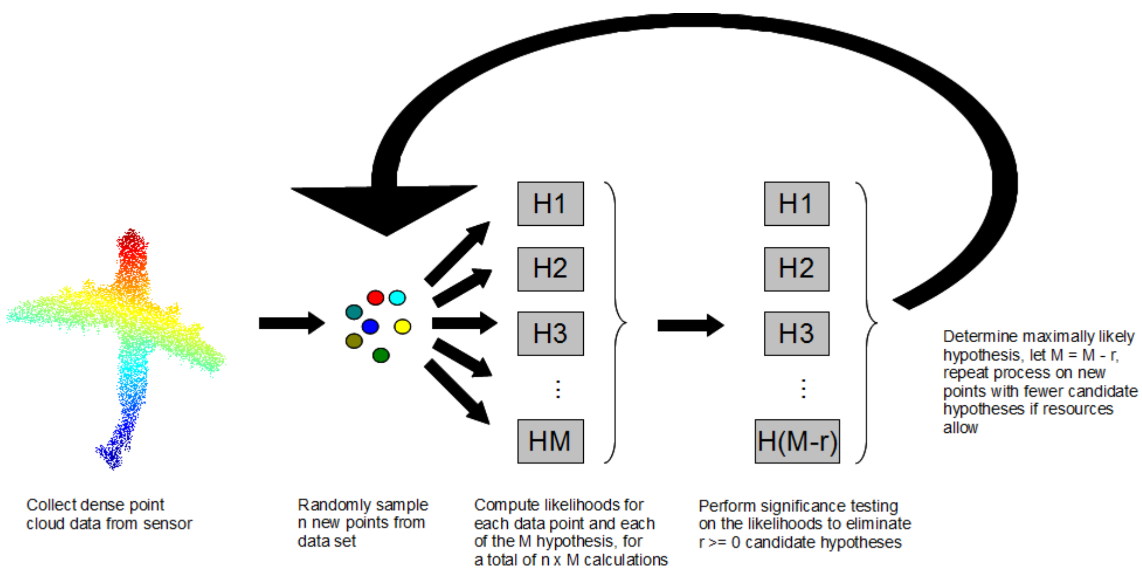

Figure 2. Sequential hypothesis testing algorithm.

For this algorithm to be practical, one must be able to evaluate the CDFs of the log-likelihoods L_m^n for each m and arbitrary values of n. In general, closed-form expressions of these functions will be difficult to determine. As an alternative, we can assume an approximate form and use sample log-likelihood values to find best-fit distribution

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 12/5/2017 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

consistent with this form. When H_m is true, the quantity L_m^n is the sum of n independent, identically distributed random quantities, each with a distribution identical to that of L_m^1. It is thus necessary to characterize the distribution L_m^1 and then extend that characterization into a corresponding characterization of L_m^n for arbitrary values of n.

This research assumes that the distribution of the sum of log-likelihoods follows a Gaussian distribution, where the mean $\hat{\mu}_1(L_m^1)$ and variance $\hat{\mu}_2(L_m^1)$ of $L_m^1$are the sample mean and variance of a sufficiently large collection of representative sample points. The mean and variance of some arbitrary $L_m^n$are then given as $n * \hat{\mu}_1(L_m^1)$ and $n * \hat{\mu}_2(L_m^1)$, respectively.

3. DATA COLLECTION AND SIMULATION

For the purposes of this research, the Georgia Tech LADAR simulator[17] was utilized to produce the raw point cloud data sets to process under the hypothesis testing schemes examined. 16 3D models of a variety of vehicles from the Princeton Shape Benchmark graphics repository[18] formed the core target library used in this project. The target library was comprised of 4 fixed-wing aircraft, 4 armored land vehicles, 4 civilian automobiles, and 4 commercial trucks. Because the source repository had previously scaled down all objects to fit within a space corresponding to a 1 cubic meter bounding box, the vehicles were appropriately rescaled up prior to use in order to more accurately depict their real-world counterparts. Furthermore, these 3D models were reoriented such that the origin lay in the center of each target and that all were oriented the same way, effectively assuming correct pose estimation. The LADAR simulator generated 50000-measurement point clouds for each of these 16 targets. Measurement points were collected all around each model, except for the undersides, simulating what would be visible to an airborne sensing platform. Information about the target library and the LADAR simulator parameters are summarized in Tables 1 and 2, respectively.

Table 1. List of reference vehicle models used.

FILENAME	DESCRIPTION	SCALING FACTOR
m1148.off	Boeing 757-300	70.7
m1159.off	McDonnell-Douglas MD-90	70.7
m1166.off	Bombardier Learjet 45	17.68
m1168.off	Grumman F-14 Tomcat	18.6
m1411.off	M-1 Abrams	7.9
m1413.off	M-2 Bradley	7.9
m1416.off	Generic WWI-era tank	7.9
m1419.off	Generic modern tank	7.9
m1519.off	Generic 1970s 2-door sedan	4.8
m1524.off	Generic 1980s 4-door compact	4.8
m1532.off	Blocky car model	4.8
m1539.off	Exotic sports car	4.8
m1572.off	Semi-trailer truck	5.7
m1499.off	Pickup truck	5.2
m1570.off	Blocky semi-trailer truck model	5.7
m1574.off	Military truck transport	5.7
m1320.off	Bell 407	12.7

Table 2. LADAR simulator parameters used.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 12/5/2017 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

PARAMETER	VALUE
Object position (x, y, z)	0, 0, 0
Rotation angle	0
Rotation axis (x, y, z)	0, 0, 0
Scale (all, x, y, z)	1, 1, 1, 1
Ground plane	Off
Noise (variance)	$0.0645m^2$ ($10in^2$)
Scene type	Point cloud
Field of view (vertical, horizontal)	15, 15
View position (x, y, z)	-64.9519, 112.5, 75
Position (dist, az, elev)	150, 120, 30 for 25K points 15, 240, 30 for remaining 25K points
Look at (x, y, z)	0, 0, 0
Min/max range	1, 300
Dimensions (Row x Columns)	Controls the number of points produced
Up vector (x, y, z)	0, 1, 0

In addition to the point clouds originating from each of the 16 targets represented in the reference library, 3 other point clouds were generated in order to demonstrate the reject option: two representing Gaussian noise with variance 20m centered about the origin; and the remaining produced from a helicopter model, a target not anticipated by the library.

In order to avoid redundant computation and expedite experimental runs, all the log-likelihoods corresponding to every point-cloud/target-hypothesis pair were computed offline prior to the experiments. Sample means and variances were also pre-computed from a random 1000-point sampling of the likelihoods corresponding to a correct classification under every target hypothesis. These likelihood sets were shuffled point-wise at the beginning of each experimental trial to simulate the unordered nature of collected data.

The experiments performed in this project were designed to evaluate the proposed sequential hypothesis testing algorithm from two perspectives: first, to compare it to a uniform computation strategy and explore the trade-off between computation and data consumption; and second, to demonstrate the reject option offered by sequential testing. These experiments are summarized in the following table. 500 trials were performed for each experiment. A maximum of 5000 likelihood computations was set for each run. A 5% significance level is set for most of the experiments as that was the common value found in the literature.

Table 3. Summary of experiments performed

SEQUENTIAL TESTING VS. UNIFORM COMPUTATION			
DESCRIPTION	**TARGET LIBRARY**	**TRUE TARGET**	**Significance**
Vary the number of target types in the target library (plane, tank, car, truck)	1. All tanks 2. All tanks and cars 3. All 16 targets	Experiments run for each possible true target	5%
Vary the number of targets in each of the four target types (1 – 4 per type)	1. 1 per type 2. 2 per type	Experiments run for each possible true target	5%

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 12/5/2017 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

	3. 4 per type		
EVALUATING THE REJECT OPTION			
DESCRIPTION	**TARGET LIBRARY**	**TRUE TARGET**	
Vary the unknown distribution presented to the sequential hypothesis test	All 16 targets	1. Helicopter 2. Large noiseball (from m1148.off) 3. Small noiseball (from m1519.off) 4. M1 Abrams (to test Type II errors)	1. 5% 2. 3% 3. 1%

The first set of experiments produced resulting data sets containing information about the number of points consumed and likelihoods computed on every stage of each trial, as well as the target hypothesis asserted. This information is used to determine the correct classification rate as a function of measurements necessary and as a function of computations carried out. To evaluate the effectiveness of the reject option, the rate at which the null hypothesis is asserted in those experiments is computed.

4. EXPERIMENTAL RESULTS

This section presents the quantitative evaluation of the sequential hypothesis testing algorithm. The first set of experiments, detailed in subsections 4.1 and 4.2, examined the performance of the sequential hypothesis testing algorithm compared to a naïve anytime uniform-calculation strategy. The uniform-calculation strategy is optimal in the sense that it is guaranteed to converge to near-100% recognition accuracy as the number of likelihoods computed grows arbitrarily large. However, because this requires the computation of likelihoods across *all candidate target hypotheses*, uniform-calculation may not necessarily be optimal in terms of the number of likelihoods evaluated. Sequential hypothesis testing explicitly discards hypotheses as it progresses, so this may offer computational savings under particular circumstances. The experiments in subsections 4.1 and 4.2 explored this trade-off between considering all target hypotheses, evaluating fewer likelihoods, and increasing the number of measurement points necessary to reach sound inferences, all as a function of the number and nature of available target hypotheses.

Subsection 4.3 follows with a collection of experiments exclusive to the sequential hypothesis testing algorithm, whereby the effectiveness of its *reject option* is explored. The sequential procedure was presented a collection of point-cloud observations originating from an unknown distribution. These experiments collected data on the number of computations necessary to reject all 16 candidate target hypotheses and the rate at which the null hypothesis was correctly asserted, all as a function of significance value and the nature of the unknown distribution.

4.1 INCREASING NUMBER OF DISPARATE HYPOTHESES

This series of experiments explored the effect of considering extraneous target hypotheses that are not likely to properly account for the observed data. This was accomplished by first considering only the 4 target hypotheses of the same type as the object under observation. For the purposes of this experiment, the point-cloud observed originated from the M1 Abrams model, so the initial set of hypotheses included all 4 armored vehicles: M1 Abrams, M2 Bradley, WWI tank, and modern tank. Further tests increased the number of types of target hypotheses under consideration: 2 types (armored and automobiles) and all 4 types. The recognition rate was experimentally determined as separate functions of the number of likelihoods computed and data points consumed. These results are graphically summarized in Figures 3 and 4, respectively.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 12/5/2017 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

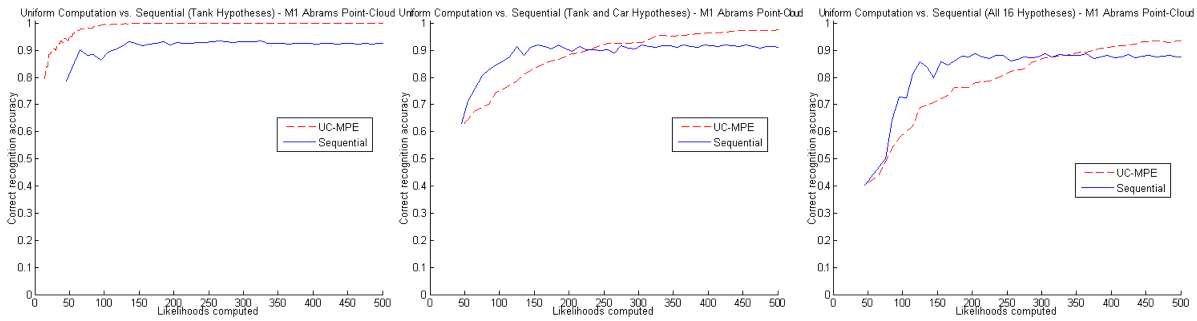

Figure 3. Likelihood computations vs. recognition rate, increasing disparate hypotheses.

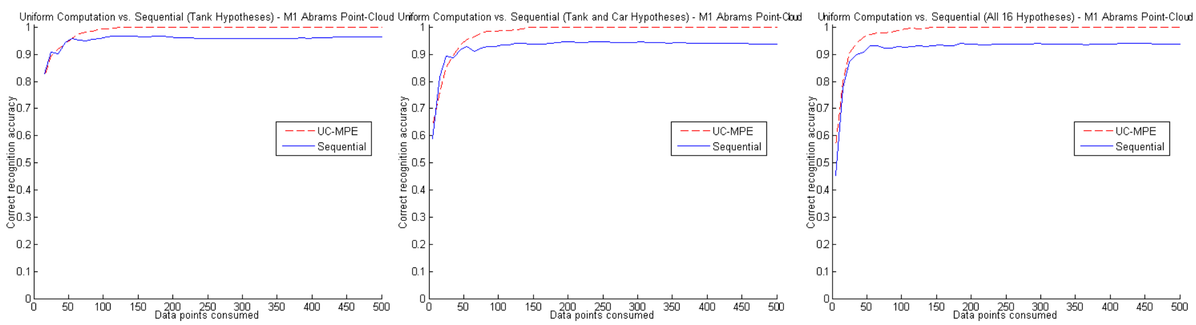

Figure 4. Data consumption vs. recognition rate, increasing disparate hypotheses.

These results highlight the trade-offs posed by the sequential hypothesis testing algorithm. Unlike the uniform-computation strategy, this sequential testing procedure is not guaranteed to converge to near-certain accuracy as the number of points consumed and likelihoods computed arbitrarily increase. Rather, sequential hypothesis testing may converge to a lower recognition rate, as explained by the fact that there exists a non-zero probability that the correct hypothesis is incorrectly discarded; uniform testing considers all hypotheses equally. These experiments, however, reveal that sequential testing reaches its convergent recognition rate in fewer likelihood computations under the event that there are many unlikely hypotheses to consider. Figure 3 illustrates this observation: where there are only a few highly similar target hypotheses (armored vehicles only), sequential testing offers only minimal benefit because the similar distributions posed by the available hypotheses are not conducive to the sequential test's elimination process; however, where there are many more disparate hypotheses and their corresponding dissimilar distributions, sequential testing discards those hypotheses and thus invests fewer computations on unlikely models. In the 2-type test, sequential testing reaches an 85% accuracy rate in 35% fewer computations than the uniform calculation scheme. Similarly, in the 4-type test, sequential testing achieves 85% in 55% fewer computations. In terms of data consumption, sequential testing has no advantage, as expected.

4.2 INCREASING NUMBER OF SIMILAR HYPOTHESES

The following results arose from experiments identical in structure to those presented in subsection 4.1. However, instead beginning with a small initial set of similar distributions to test and considering additional dissimilar hypotheses, these tests examined the effect of starting with a small initial set of disparate hypotheses and increasing the number of hypotheses in each type. That is, hypotheses from the armored, automobile, aircraft, and truck types were present in all these tests, but the number of hypotheses in each type was varied. As before, the M1 Abrams tank was designated as the true target.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 12/5/2017 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

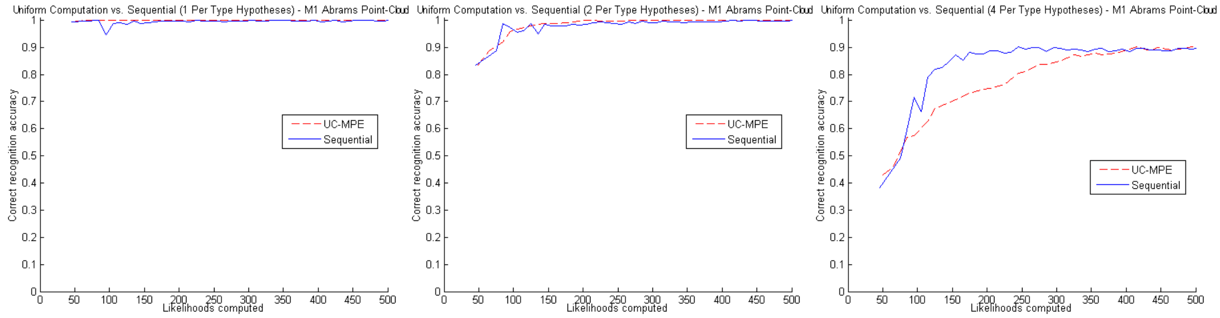

Figure 5. Likelihood computations vs. recognition rate, increasing similar hypotheses.

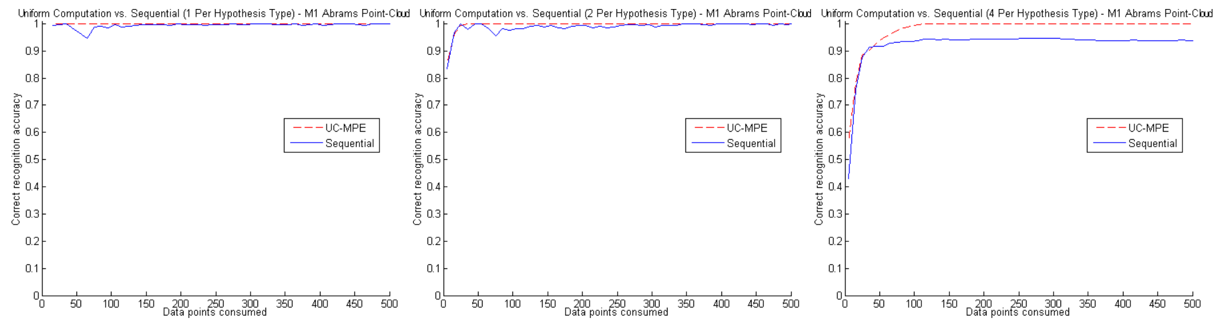

Figure 6. Data consumption vs. recognition rate, increasing similar hypotheses.

These results provide further insight into the observations noted in subsection 4.1. While the sequential testing procedure may converge to a lower recognition rate than uniform computation, this is not always necessarily the case, as evidenced by 1-hypothesis-per-type experiment. When there are a small number of highly dissimilar distributions associated with those hypotheses, the sequential test can quickly (in terms of computations needed) and correctly eliminate improbable hypotheses through significance testing. This is the expected outcome, as the opposite was observed in the previous subsection where there was a small number of hypotheses within the same vehicle type. As the number of hypotheses within each type increases, the sequential test's converging accuracy rate decreases. This is explained by the fact that once the field of viable hypotheses has been narrowed down to the only the most likely distributions, the elimination stage of the sequential test becomes a hindrance in that it poses the possibility that the correct hypothesis is erroneously rejected. As before, this formulation of sequential hypothesis testing offers little by means of performance relative to the number of world observations needed.

4.3 REJECT OPTION

Unlike the uniform computation algorithm, the sequential hypothesis testing algorithm does not need to assert any of the available target hypotheses. This occurs in the event that all the hypotheses are rejected through significance testing, as what might be desired if the algorithm is presented with data inconsistent with all the known models. The experiments presented in this subsection examine the effectiveness of this reject option as a function of the observations processed and the significance values. Three different unknown distributions – 2 Gaussian noise fields generated from small and large targets, respectively, and a helicopter point-cloud – are used to evaluate the rate at which the null hypothesis is correctly asserted. The M1 Abrams data set from before is used to estimate the rate at which the null hypothesis is wrongly asserted. The full set of 16 reference target hypotheses from before is considered.

Table 4. Null hypothesis assertion rates and average computational cost.

	SIGNIFICANCE = 5%	SIGNIFICANCE = 3%	SIGNIFICANCE = 1%
SMALL NOISEBALL	100%, 105 computations	100%, 105 computations	100%, 105 computations
LARGE NOISEBALL	100%, 103 computations	100%, 105 computations	100%, 105 computations

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 12/5/2017 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

HELICOPTER	100%, 105 computations	100%, 105 computations	100%, 105 computations
M1 ABRAMS (Type II Error)	2.6%, 607 computations	2.6%, 336 computations	0.6%, 272 computations

It is shown that the reject option is particularly effective with the given experimental condition of 16 target hypotheses grouped into 4 target types of 4 targets each. When processing an entirely unknown target, as was the case with the Gaussian noise and helicopter point-cloud inputs, the sequential hypothesis test consistently rejected all the proposed models in roughly the same amount of computation each time. The significance bound plays little part in the correct null hypothesis assertion rate and the computational work necessary to reach that. This value, however, has a major effect on the type II error rate. Processing data with ever tighter significance bounds reduces the rate and cost at which the null hypothesis is incorrectly asserted. However, given the observation from the previous experiments that the significance test's class conditional error rate may be as high as 10% - 15%, this suggests that the previously-mentioned gain may be offset by a higher chance that an incorrect hypothesis is asserted.

5. CONCLUSIONS

This paper has outlined and demonstrated a sequential hypothesis testing algorithm in the context of a multi-class automatic target recognition problem. The experiments performed on this formulation of sequential testing show that there exists a trade-off when processing observation data in this manner. While a sequential testing algorithm is not guaranteed to converge to an optimal accuracy level as more computations are performed, it may however reach a lower accuracy bound in fewer computations than a comparable non-eliminating uniform-calculation approach. This accuracy bound is heavily influenced by the number and nature of target hypotheses under consideration. As such, it is safe to conclude that sequential hypothesis testing is appropriate for use in situations where computational savings are needed at the expense of some degree of recognition accuracy when there are a large number of disparate hypotheses to consider. Moreover, this sequential testing algorithm proves itself capable of making no claims about a set of observations when none of the viable hypotheses sufficiently explain the data. The trade-off here, though, comes in the form of a chance of incorrectly rejecting the true hypothesis during significance testing.

While this project touched on a basic formulation of sequential hypothesis testing and its performance under a variety of available target hypotheses and known and unknown input observations to process, there are still other aspects that are worth exploring. The significance testing procedure in this formulation assumes that the distribution of log-likelihood may be modeled as Gaussian distribution with moments scaling to the amount of data processed. Although this approach proved straightforward to implement, a Gaussian distribution may not be the best fit for the data collected. Other finer-grain methods such as the Johnson distribution family or polynomial chaos methods may provide for more nuanced and accurate significance testing, thereby reducing the type II error rate observed. Additionally, the explicit relationship between available hypotheses, the specified significance value, and the significance test's suboptimal recognition convergence is well worth deeply understanding in order to better allow systems-designers to manage acceptable error rates. These advanced distribution estimation methods and questions surrounding the aforementioned suboptimality should be explored in future work.

ACKNOWLEDGEMENT

This work was supported by the Air Force Research Laboratory grant FA8650-07-1-1-1113.

REFERENCES

[1] Marino, R. M., Davis, W. R. Jr., "Jigsaw: A Foliage-Penetrating 3D Imaging Laser Radar System," Lincoln Laboratory Journal 5(1), 23-36 (2005).

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 12/5/2017 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

[2] Wang, C., Thorpe, C., Suppe, A., "Ladar-Based Detection and Tracking of Moving Objects from a Ground Vehicle at High Speeds," IEEE Intelligent Vehicles Symposium (IV2003)

[3] Tan, S., Stoker, J., Greenlee, S., "Detection of Foliage-Obscured Vehicle Using a Multiwavelength Polarmetric Lidar," IEEE Intelligent Vehicles Symposium (2003)

[4] Zhou, X., "Statistical Model-Based Object Recognition from Three-Dimensional Point-Cloud Data," PhD dissertation, Department of Systems and Information Engineering, University of Virginia (2008)

[5] DeVore, M., Zhou, X., "Minimum Probability of Error Recognition of Three-Dimensional Laser-Scanned Targets," Proceedings of SPIE, Automatic Target Recognition XVI, 6234 (2005)

[6] Redard, D., "A Performance-Estimation-Based Tree Algorithm For Reducing Computation in Automatic Target Recognition," MS thesis, Department of Systems and Information Engineering, University of Virginia (2008)

[7] Wald, A., "Sequential Tests of Statistical Hypotheses," The Annals of Mathematical Statistics, 20(4), 117-186 (1945)

[8] Jouny, I., Garber, F.D., "M-Ary Sequential Hypothesis Tests for Automatic Target Recognition," IEEE Transactions on Aerospace and Electronic Systems, 28 (2), 473 (1992)

[9] Sobel, M., Wald, A., "A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution," The Annals of Mathematical Statistics, 20(4), 502-522 (1949)

[10] Armitage, P., "Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis," Journal of the Royal Statistical Society, Series B, 12(1),137-144 (1950)

[11] Simons G., "A sequential three hypothesis test for determining the mean of a normal population with known variance," The Annals of Mathematical Statistics, 38(5),1365-1375 (1967)

[12] Lorden G., "2-SPRT's and the modified Kiefer-Weiss problem of minimizing an expected sample size," The Annals of Statistics, 4(2), 281-291 (1976)

[13] Ghosh, B.K., Sen, P.K., editors. Handbook of Sequential Analysis. Marcel Dekker, 6, sec. 9.2 (1991)

[14] Baum, C.W., Veeravalli, V.V., "A sequential procedure for multihypothesis testing," IEEE Transactions on Information Theory, 40(6), 1994-2007 (1994)

[15] Dragalin, V.P., Tartakovsky, A.G., Veeravalli, V.V., "Multihypothesis sequential probability ratio test - part I: Asymptotic optimality,". IEEE Transactions on Information Theory, 45(7), 2448-2461 (1999)

[16] Dragalin, V.P., Tartakovsky, A.G., Veeravalli, V.V., "Multihypothesis sequential probability ratio test - part II: Accurate asymptotic expansions for the expected size," IEEE Transactions on Information Theory, 46(4), 1366-1383 (2000)

[17] Dixon, J.H., "LADAR Simulator User Manual," jdixon@ece.gatech.edu (2007)

[18] Princeton Shape Retrieval and Analysis Group – Princeton Shape Benchmark, http://shape.cs.princeton.edu/benchmark/

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 12/5/2017 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use