A Statistical Methodology for Automatic Target Recognition in Satellite Imagery

John Bart Wilburn
Recognition Research
Tucson, Arizona
Wilburn@dakotacom.net

Abstract: A methodology is presented for Automatic Target Recognition (ATR) of missiles in satellite imagery of a cloud-covered earth. The method is based on the results of a two dimensional local maximum filter applied to a sequence of simulated images capturing the boost phase of a missile, and is an algorithmic estimate of the probability of detection on a trajectory, i.e., a "track". Track association is determined by the significance level for rejecting hypotheses that detection events in sequential frames of imagery are not associated with missile targets or satellite motion.

Introduction:

The problem addressed here is a calculation of statistical confidence in the detection and tracking of missile targets acquired by a satellite sensor in the boost phase of missile deployment. The specific task of the system is to detect a small, relatively bright, moving target against a scene cluttered by clouds. The criteria for success are a high probability of detection and a low probability of a false alarm, and computational efficiency. Further, assumptions for this system must be minimal as follows:

1. The elements of the scene are either background features, e.g., clouds and terrestrial lights, or missile targets.
2. Target radiance is additive.
3. The target moves along a line of predictable shifts between frames of imagery,
4. The target object is small, generally much less than the resolution of the optical system, thus it is represented in the image of the point-spread function (psf) of the optical system, adequately sampled by the focal plane array (FPA), with the peak expected most of the time to be in the instantaneous-field-of-view (IFOV) of one pixel.

The method of detection and tracking proposed is based on the notion that detection is an event, i.e., the occurrence of an isolated and distinct pixel. To satisfy this requirement, the satellite imagery must be filtered in a manner that results in distinct and isolated pixels, and the filter that satisfies this requirement is the local maximum filter[1]. As will be shown, this filter also satisfies the requirements of a high probability of detection and a low probability of false alarm, is computationally efficient, and it is well behaved in the context of cloud-cluttered imagery

The Local Maximum filter:

The local maximum filter is a form of ranked-order (RO) filters developed from theory described in prior work[1 - 3] on RO filters. The application of the local maximum filter to target recognition in a cluttered image is a form of feature extraction derived from a fundamental departure of local RO filters from classical approaches to image filtering. The local maximum filter, and its RO family relative the local median filter, are a window type of filter and function according to the satisfaction of a reflexive relationship between the location and relative values of the data in the sample space of the filter window. Morphological considerations of the filter, then, apply to the structure of the data with respect to the reflexive relationship defining the filter, and the morphology of a particular object, or class of objects, of interest are represented in the filter coincidentally as properties of an object that satisfy the reflexive relationship of the filter. A

presupposition of an extensional property, such as the shape of a particular object, or class of objects, evident in some context, is not represented in the filter.

The approach to target recognition described here is not to focus on suppressing the clutter to reveal the target, but rather to focus on finding the target embedded in the clutter. A canonical description of this approach is looking for properties shared by any significant part of all targets, but not shared by other features in the image. This is the mode of object recognition employed by human observers[4], indeed by all animals. This principle, and our application of it, and may be illustrated by the analogy of a looking for a baseball in a pile of leaves. A baseball has the properties of sphericity, grayness and it has a seam. All baseballs share these properties; wherever you see a baseball under any conditions you see it, it has these properties. If we can see as much as any quarter of any baseball in any situation of the clutter, we see these properties. We expect nothing else in the pile of leaves to have these properties, and if we see anything else with these properties and claim it to be a baseball, then we commit a type-II error - a false alarm. Conversely, if we see some part of a baseball and reject it, then we commit a type-I error.

Filters can be constructed in a variety of geometries (Figure 1) that include adjacent pixels on linear intersecting arms, denoted by the "0", with a single common pixel at the intersection, or center position indicated by " ⊗", that is also the output port of the filter. The orthogonal filter prohibits adjacent maxima on the horizontal and vertical axes while the diagonal filter prohibits adjacent maxima on the diagonal axes. The hexagonal filter prohibits adjacent maxima on all but the horizontal axis while the octagonal filter prohibits adjacent maxima on all axes. The latter is well suited to detection and track of a missile launch in any direction.

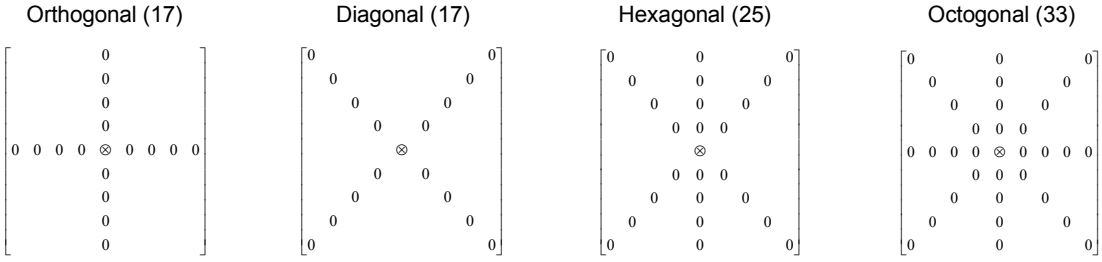

Figure 1. Filter Geometry (Number of Elements)

The local maximum filter functions according to a rank-order relationship of the data sampled by the filter constrained by a *predicate.* A predicate has semantic content as it is either true or false, and the filter is constrained by it to be defined if the predicate is true, or undefined if it is false. The predicate is a description of the data in terms of conditions. If the data do not satisfy the conditions, then the predicate is false and the filter returns a null result. If the data satisfy the conditions, then the predicate is true and the filter returns the result of the filter function at the center position of the window.

The predicate conditions of the local maximum filter involve the value and position of an individual datum with respect to the value and position of all data within the scope of the window, and they apply to each of the 1-d arms of the filter simultaneously for each increment of the filter in the field of data. The predicate conditions are: (a) the datum sampled by the center position of the filter, " ⊗", is simultaneously the maximum value of the data of every axis of the filter, and (b) the satisfaction of a threshold established by the SNR statistics of the image. This threshold defines the lower limit of a reasonably expected target in terms of its SNR determined by its brightness and the standard deviation of the image brightness. If the data sampled by the filter satisfy these conditions, the predicate is true and the filter returns the datum sampled by the center position of the window. If the predicate is not true, then the filter returns a zero. This functioning of the octagonal filter results in isolated and distinct pixels as

intended, but of particular interest is the additional realization that the output image of all local maximum filters is a subset of the input image, thus radiometric information is preserved.

The effectiveness of the octagonal local maximum filter in isolating local maxima is demonstrated (Figure 2) for a simulated cloud scene (102 x 102 pixels) in a solar band from the Synthetic Scene Generation Model (SSGM). The imagery is presented as a bit map on an 8-bit scale, thus all pixel values are on a scale of 0 to 255. The figure compares the pixel imagery and histogram/exceedance statistics before and after application of the filter. The comparisons indicate a significant reduction in the number of pixels at all brightness levels: the total number of pixels decreases from greater than 10^4 to less than 200. The filtered image consists of a pattern of isolated pixels as intended, and the pixel levels of the original, or input, image are preserved in the output, or filtered, image, thus radiometric information is preserved. As can be seen, many regions of a natural image above some threshold do not have local maxima in the sense defined here.

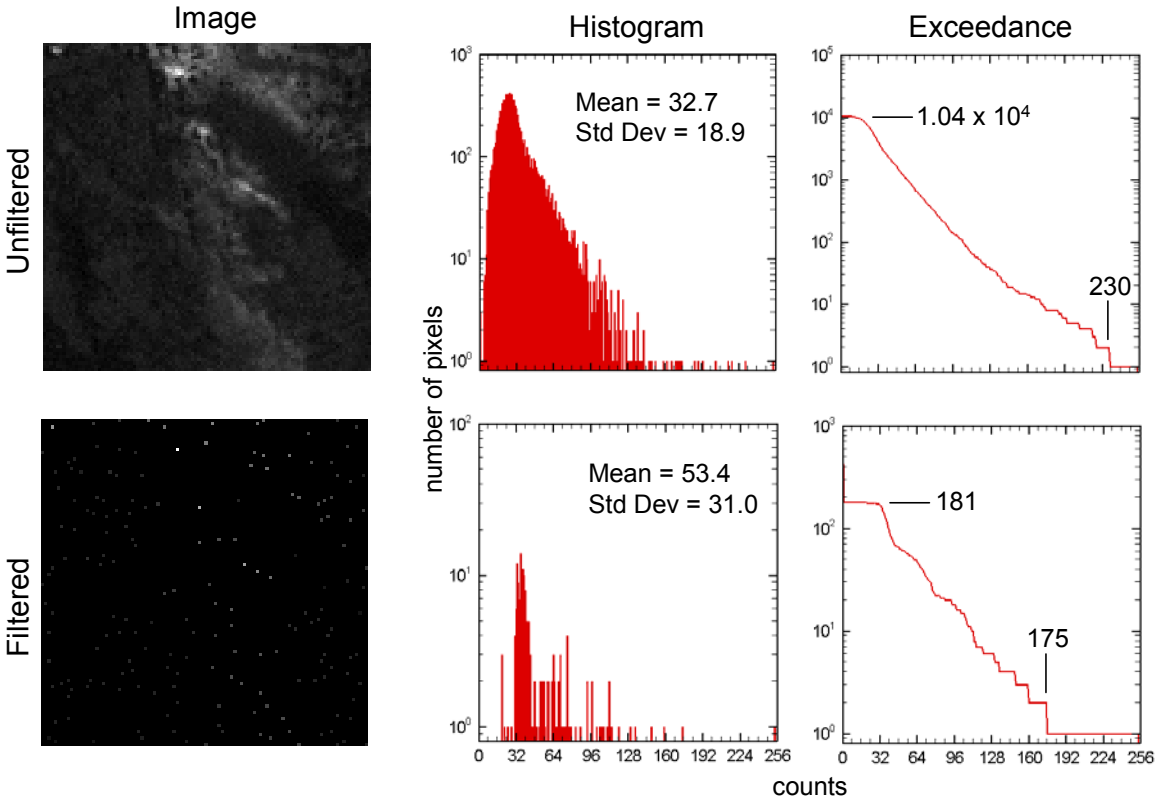

Figure 2. Filter Performance for SSGM Scene

The performance of the filter is specific to a class of image clutter represented by figure 2. The image in figure 2 is one frame of a sequence of 26 frames simulating the continuous nadir view at one frame per second of a cloud cluttered earth by satellite sensor in low earth orbit. This set of imagery is reasonably stationary in the strict sense: the image mean of the set $\langle\langle I\rangle\rangle = 31.1$ with, $\sigma_{\langle I\rangle} = 2.72$, and mean standard deviation: $\langle\sigma_I\rangle = 19.4$ and $\sigma_\sigma = 1.1$. The statistics of the image in figure 2 are: $\langle I\rangle = 33.7$; $\sigma_I = 18.9$, *skew=2.58,* and *Kurtosis=15.42*. The filter performance is described by two parameters: the probability of detection, p_d, and the probability of false alarm, p_{fa}. These probabilities are measured in two kinds of Monte Carlo tests of 300 trials of the filter against the raw image shown in figure 2 for randomly inserted targets of different brightness levels. The p_d is the empirical probability in the Monte

Carlo test that the actual target is one of the output pixels of the filtered image, and p_{fa} is the empirical probability of pixels appearing in the output image that are not the target.

The first kind of Monte Carlo test, shown in figure 3, is of a fixed composite of target plus background randomly inserted in place of a background pixel. The second kind of Monte Carlo test, shown in figure 4, is of a fixed target brightness added to a randomly selected background pixel. These two Monte Carlo tests satisfy different purposes. The first kind is a measure of the filter to detect fixed pixel brightness levels in the test image. This first kind of Monte Carlo test measures the probability of the filter to detect an unknown target parameterized by an average brightness of the pixel containing the target measured over all images having an intensity distribution as shown in figure 2. The second kind of Monte Carlo test is of a simulated target of known, absolute brightness to satisfy a given probability of detection in imagery having an intensity distribution as shown in figure 2. As will be seen, the average signal-to-noise of the target plus background pixels, $<SNR_t>$, measured in the second test, and the $<SNR_t>$ inferred from $<I_t>$ of target pixels, and $<I>$ and σ_I of the total image in the first test are remarkably close.

The first kind of test is parameterized by the signal-noise (SNR_i) of the target *pixel* for the i^{th} level of brightness: T_i.

$$SNR_i = T_i / \sigma_I \text{, where } \sigma_I \text{ is the standard deviation of the image brightness.} \quad (1)$$

The filtered image is subject to a threshold test by setting to zero all pixel output of the filter satisfying the integer calculation of $p_n < SNR_i * \sigma_I$. The purpose of the test was to describe the behavior of the filter parametrically in terms of p_d and p_{fa} as a function of a variable SNR_i in the space of 0 to 7.5. The p_d and p_{fa} results are shown in figure 3.

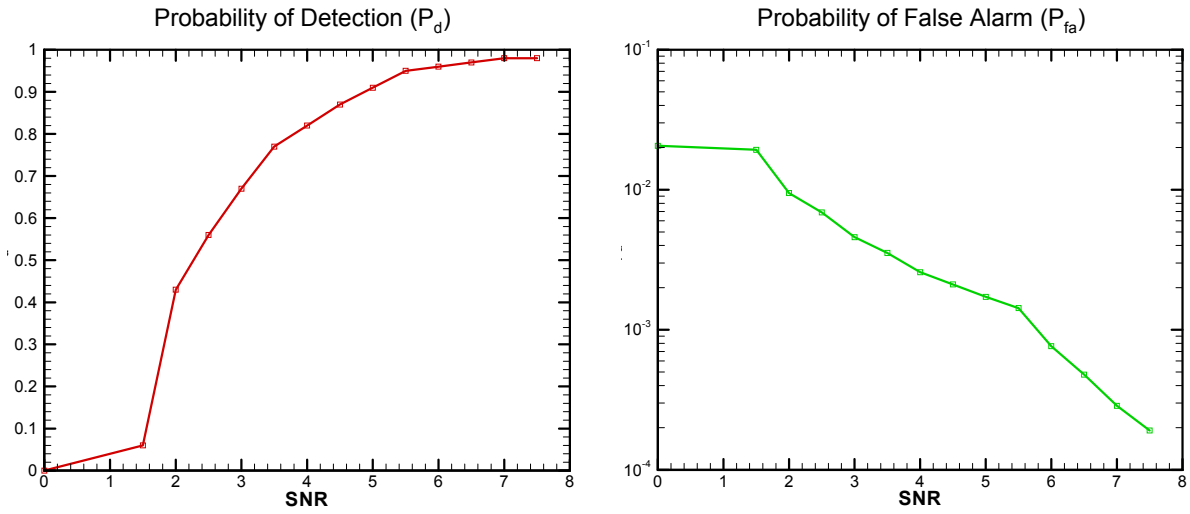

Figure 3. Composite Target Pixel Detection and False Alarm from Monte Carlo Statistics

The results of the Monte Carlo Testing show that the filter is well behaved and that the region of a reasonable target is one having a SNR ≥2.5 satisfying a probability of detection of p_d >0.5 and a probability of false alarm of $p_{fa} = 6.4 \text{ X } 10^{-3}$.

The second kind of Monte Carlo test follows from the first kind with a criterion for detection of p_d >0.5. In this test, we fix the p_d and search for the minimum absolute target brightness satisfying this criterion as a function of the threshold of the filter in terms of the SNR. The results are curves of absolute target brightness as a function of SNR, or p_{fa} by figure 3, parameterized by p_d. The average (target + background) pixel signal-to-noise, <SNR_t> satisfying p_d is computed over all measures of SNR_t in the Monte Carlo tests of the image.

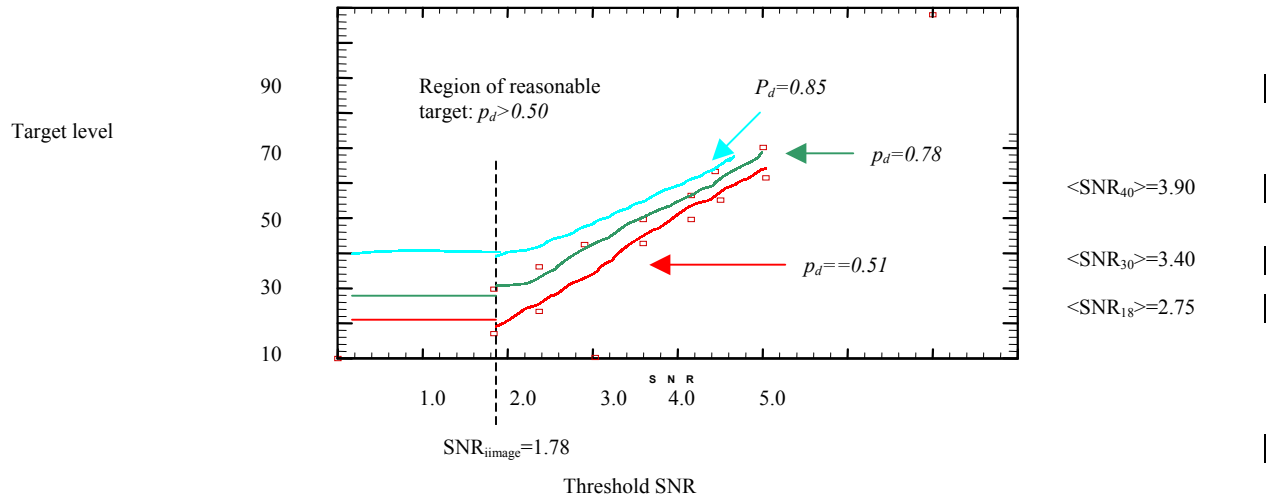

Figure 4. Target Level Satisfying a p_d from Monte Carlo Statistics

The distribution of the probability of detection of a given target level in this test follows the distribution of the image intensity (figure 2) according to conventional signal detection theory in terms of the skewed statistics of the image data. Thus it follows that the target level for p_d=0.50 is constant for thresholds less than or equal to the SNR of the image, as that threshold level corresponds to the image mean value, and increasing for thresholds greater than the image mean. By the same reasoning, we expect that target levels for p_d>0.5 is constant to thresholds somewhat greater than the SNR of the image, and increasing thereafter, and we see that this is indeed the case. The results show that the minimum target level for p_d=0.51 is *target level=18* having a $<SNR_t>=2.75$ at a threshold of *SNR=1.78*, or by figure 3, a $p_{fa}=1.58 \times 10^{-2}$, and increases to *target level= 26* ($<SNR_t>=3.18$) at a threshold of *SNR= 2.5* ($p_{fa}=6.4 \times 10^{-3}$). The results for higher probabilities of detection follow a similar pattern: The minimum target level at p_d=0.78 is *target level=30* ($<SNR_t>=3.40$) at a threshold of *SNR= 2.0* ($p_{fa}=1.2 \times 10^{-2}$), and for p_d=0.85, the minimum target level is a *target level=40* ($<SNR_t>=3.9$) at a threshold of *SNR= 2.5* ($p_{fa}=6.4 \times 10^{-3}$). These results compare favorably with the inferences drawn form figure 3, and the expected results from signal detection theory.

Target Recognition:

The schema for target recognition combines the notions of detecting local maxima and persistent linear motion. Persistent linear motion is a high probability of detection on a line of predictable shifts between frames, and it is referred to as a *track*. Determining a high probability of detection on a track is *track association*. We may combine these notions with our assumptions of the imagery and conclude that any track consistent with background motion, which we know a priori, is not like a target, and any track inconsistent with background motion is like a target. In this way, track association inconsistent with satellite motion, i.e., inconsistent with background motion, constitutes target recognition.

We may illustrate the schema for recognition of missile targets in boost phase by application of the Octagonal filter to a dynamic sequence of 25 SSGM images simulating sensor acquisition at the rate of 1 frame per second of a missile target. The missile simulated in this set of imagery is a theater missile and has a variable target pixel brightness with $\langle SNR_t \rangle = 3.8$ and a $\sigma_t = 1.7$. The results of the Monte Carlo test of the filter shown in figure 3 indicate that we may expect this target to have a $p_d \cong 0.8$, thus it is a target we may reasonably expect to detect and it is a typical target by definition. Recognition of the target is determined by the measure of confidence we may have in associating detection of this target with a specific track as a reasonable and typical target.

The detections of the target and a background feature suggest a hypotheses of a target track $r_t = x - 4y$, i.e., a shift of $\Delta x = 1$ and $\Delta y = -4$ between frames, and a satellite motion track of $r_s = -2x-3y$. Detections by the local maximum filter satisfying the hypotheses of r_t and r_s at a threshold of SNR=2.5 are shown in figure 5 as a composite of frames plotted on an arbitrary frame of the input set of imagery for reference. The detection events are indicated by a bright cross reflecting a tolerance in detection of +/- 1 pixel to allow for non-integer multiples of pixel movement between frames. Failures of the filter to detect an object satisfying this hypothesis are shown by a null result.

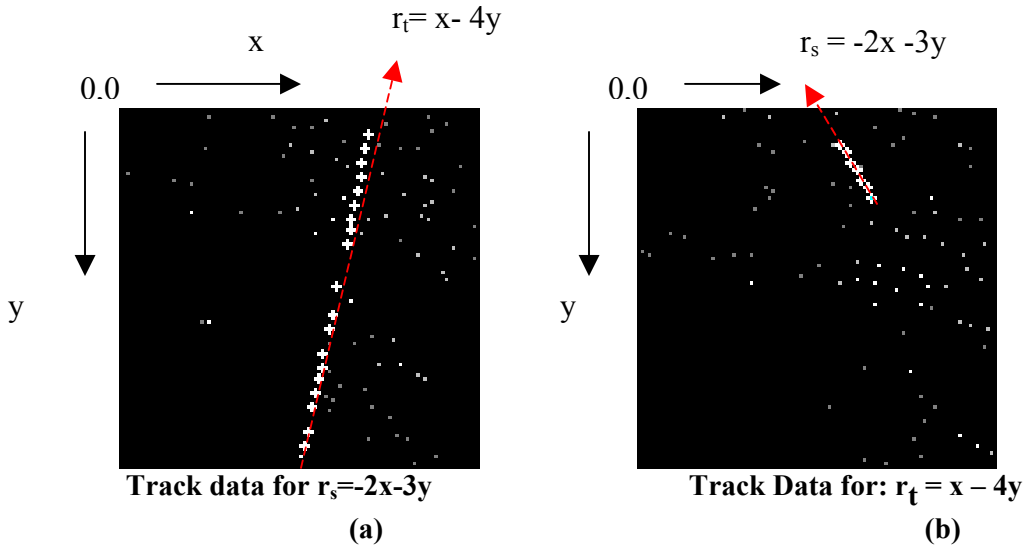

Figure 5: Track data for 25 sequential frames

The notion of detection being an event enables this approach to result in an assertion of target recognition bounded by the probabilities of type-I and type-II errors as follows. The data in figure 5 reveal that the filter detected an object on track r_t in m=20 of N=25 frames, and we employ the binomial probability law (1) to calculate the type-I and type-II errors, α and β respectively, for hypotheses of p_j *on* r_t .

$$P\left(x \geq m \middle| p_j\right) = \sum_{i=m}^{N} \binom{N}{i} p_j^i q_j^{N-i} \qquad (1)$$

$$\beta_j = 1 - P(x \geq m \,|\, p_j)$$

$$\alpha_j = P\left(x \geq m \middle| p_j\right)$$

To find the maximum likelihood estimate of p_d, we plot the α and β errors as a function of p_j, *=0 -1.0* as shown in figure 6. From figure 6, we can see that a 99.6% confidence region of p_d is found by observing that the type-I error =0.002 at *p=0.5,* and the type-II error =0.002 at *p=0.945,* thus we have a probability of 0.996 that 0.5 <p_d <0.945. The maximum likelihood estimate of p_d is the value of p_j where *the type-I error= the type-II error =0.5,* and that occurs for p_i*=0.78*.

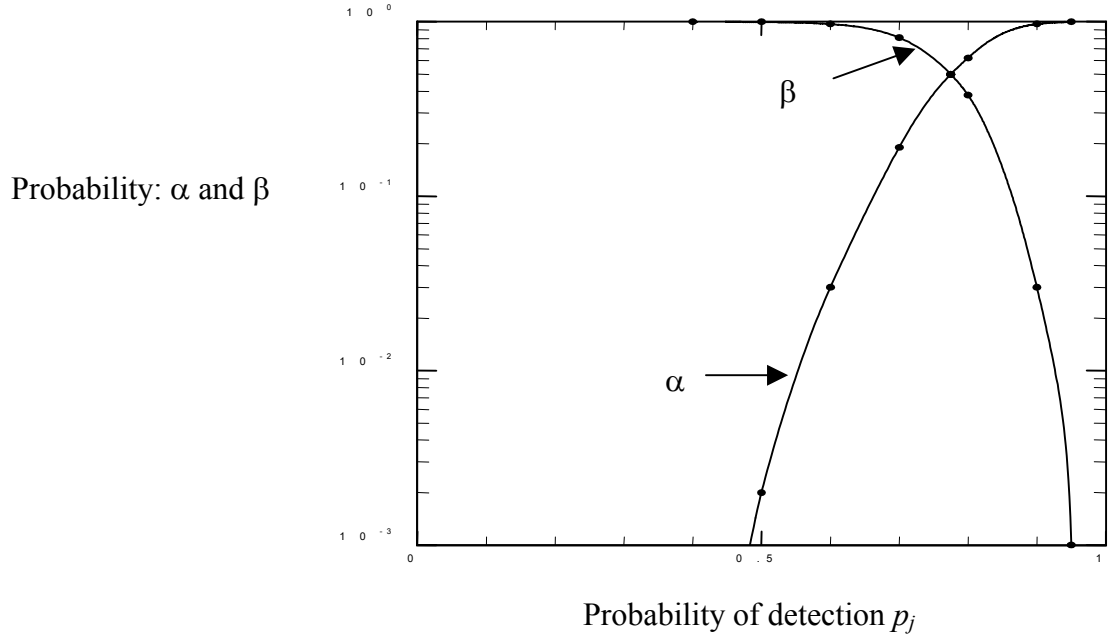

Figure 6: Type-I and Type-II probabilities of p_d

Estimating p_d is an intermediate step to asserting an object to be a target. We may combine the estimate of p_d with the notion that it is specific to a track and define target recognition as: p_d *>0.5 on a track.* We may then accept or reject a null hypothesis that an object is *not* a target on that basis with the alternative hypothesis being that the object *is* a target; the decision determined by the respective probabilities of either hypothesis being true. The probability of the null hypothesis being true is given by the type-I error = 0.002 evaluated at p_d =0.5, thus we may confidently reject the null hypothesis, knowing that there was only a 0.002 chance that it was true, i.e., a significance level of 0.002, and accept the alternative hypothesis that the object *is* a target on a track, or *it is recognized*, with a probability given by the type-II error = *0.998* of being true.

Tracks are found by hypotheses of a track on all pixels based on neighboring pixels within range of possible motion in subsequent frames. The hypotheses are confirmed in the manner shown here to be either satellite tracks and ignored, target tracks and recognized, or pixels that cannot be confidently associated with any track and ignored as false alarms. The development of this method is dependent on the fidelity of the simulation of the targets, and particularly dependent on the fidelity of the simulation of the imagery of a cloud cluttered earth produced by a satellite sensor.

Attribution:

This material is based upon work supported by, or in part, the US Army Research Laboratory and the US Army Research Office under contract/grant: G 44394-MA. Presented[5]: 8th Army Conference on Applied Statistics.

References

1. Wilburn, J. B., "Development of the local maximum variety of ranked-order filters", *Journal of the Optical Society of America A* (JOSA A), **19**, pp 1994-2004, 2002.

2. Wilburn, J. B., "Theory of ranked-order filters with applications to feature extraction and interpretive transforms", in *Advances in Imaging and Electron Physics*, ed. P. Hawkes, Harcourt-Brace Academic Press, **112**, pp233-332, 2000.

3. Wilburn, J. B., "Developments in generalized ranked-order filters", *Journal of the Optical Society of America A* (JOSA A), **15**, pp. 1084-1099, 1998.

4. Wilburn, J. B., "A possible worlds model of object recognition", *Synthese*, Kluwer Academic Publications, **116**(3), pp. 403-438, 1998

5. Wilburn, J. B., "Automatic Target Recognition is Satellite Imagery", Proc. *8th Army Conference on Applied Statistics*, 31 October 2002.