AUTOMATIC DETECTION OF MINELIKE TARGETS IN SEVERELY CLUTTERED IMAGES WITH OTHER MANMADE OBJECTS

P. L. Katz, C. J. Eggen, R. M. Haralick, and L. R. Rystrom

February 15, 1995

Paper 2496-81
To Appear in Proceedings of

SPIE Technical Conference 2496
Detection Technologies for Mines and Minelike Targets
17 - 21 April 1995
Orlando, FL

APPLIED PHYSICS LABORATORY
UNIVERSITY OF WASHINGTON
1013 NE 40th Street
Seattle, Washington 98105

AUTOMATIC DETECTION OF MINELIKE TARGETS IN SEVERELY CLUTTERED IMAGES WITH OTHER MANMADE OBJECTS

February 15, 1995

P.L. Katz, C. J. Eggen, R. M. Haralick, and L. R. Rystrom
Applied Physics Laboratory and Department of Electrical Engineering
University of Washington, Seattle, WA 98144

ABSTRACT

A fully automatic algorithm was developed for single-frame detection of minelike objects in realistic shallow water, beach, and nearby land environments. Detection was accomplished in gray scale images, containing representative targets and backgrounds, which had been collected by a down-looking coherent active sensor.

The problem was made challenging by low contrast, partly covered targets, and highly cluttered images including beach vegetation and rocks, complicated natural backgrounds, obscuration (replacement noise) by glint from the surface of the water and distortion within it, and 15 kinds of manmade objects. To deal with these challenges, innovations have been made in automatic background cancellation, in the final thresholding to binary, and in shape and veracity clues for the feature vector used in the final classification step.

Performance is reported for a representative set of 1024 frames. For the majority of background types, including low conventional signal to noise ratio and pervasive instances of clutter and replacement noise patches, the algorithm performed correctly in 92% of the frames. This applies individually to frames identified as "target" where one or more targets existed, and frames identified as "notarget" where no targets existed.

Target-field detection over multiple frames depends upon reliable single-frame target detection. Despite challenging images, performance of our single-frame algorithm appears sufficient for multi-frame target-field detection to proceed with acceptable error rates for the majority of background types encountered in the tests conducted.

1.0 INTRODUCTION

1.1 BACKGROUND

The problem underlying our work is the automatic detection/localization of minefields in shallow water, upward to the beach, and on nearby land, using an active coherent imaging sensor (hereafter referred to as the "camera") viewing the scene from above. Target-field detection, and particularly the reliable identification of regions without target fields, must be performed despite the presence of man made clutter objects, beach and dune vegetation and rocks, variable and complicated natural backgrounds, and variations in optical transmission due to water surface glint and in-water inhomogeneity of the optical path (caustics, etc.)

The images for the work described in this paper were collected during FY 89-90 by a developmental camera which produced images representative of those that could be obtained by an analogous camera under operational conditions. The majority of the data (6 of 8 data sets considered) was collected over representative shallow water, beach, and nearby land sites. The remainder was data of opportunity collected over a group of dry inland sites.

Target-field detection begins with fully automatic detecting of minelike targets in a stream of two dimensional gray scale images (frames) sensed from above by the "camera" system. Single frame information (target detections in frame, along with frame location in reference space) is then integrated in a target-field detection algorithm. Using

the principle that coastal mines must occur in clusters (and thus in many adjacent frames) to be effective, the target-field algorithm uses spatial concentrations of single-frame detections as its criterion.

Here we describe the automatic single-frame detection algorithm, for minelike targets, developed by the Applied Physics Laboratory, University of Washington (APL-UW) for the sponsoring organization, namely the Airborne Mine Detection And Surveillance (AMDAS) program, under the technical direction of Naval Coastal Systems Center (now NSWC/DD, Coastal Systems Station). The algorithm was required to be fully automatic so that the recognition process could be independent of the variations in human judgement, and so that the entire recognition process could be performed, without intervention, in a high speed parallel-processing hardware set (References 1, 7) developed simultaneously under AMDAS funding.

1.2 TECHNICAL PROBLEM DEFINITION

As indicated in Reference 2, the automatic detection problem for minelike targets is a difficult one. Targets of several classes are geometrically relatively well-defined in the sensor frames (though with some random component), but embedded in often severe random backgrounds of various types. Sensor frames typically may also contain one or several classes of man-made objects which are not targets, hereafter called non-target man made objects, or MMOs.

Because the problem here is a single frame target detection problem embedded in a multiple frame target-field estimator, there are two important design implications:

> Small numbers of errors from the single frame algorithm can be tolerated without resulting in a false negative or positive report from the target-field detector based upon clusters;
>
> The most important single-frame measure of performance is correctness of the report "any targets vs no targets," with the correctness of the actual numerical count being of lesser importance.

The problem was attacked in two phases. In an initial phase, the images contained at most one type of target. Severely complex backgrounds were relatively rare, and MMOs virtually absent. Image processing results for this phase, encompassing candidate detection but not classification (per the algorithm overview below), were presented in Reference 2. A second phase introduced more diverse images at two pixel resolutions, including four target types, seven background types (all but one of which are quite complex, spatially) from shallow water, the beach, and nearby land. Fifteen types of non-target MMOs were included, many in substantial numbers. This paper reports on the final delivered version of the single-frame algorithm, end-to-end, including results on second-phase images.

1.3 ALGORITHM OVERVIEW

The overall single frame target detection algorithm is shown in block diagram form in Figure 1-1. Target detection consists of candidate detection followed by classification. Candidate detection in turn consists of the following:

> Gray scale morphological image processing which enhances each input frame (e.g. removes clutter and occlusions, normalizes the background);
>
> Automatic thresholding to a binary image, in which connected regions are then labelled as separate objects and become the detection candidates; and
>
> Obtaining a feature vector for each such candidate region in each frame.

Classification operates in feature vector space to give each candidate an assigned category "Target" (T) or "Non-target" (N); these are referred back to the original frames to produce a count of targets detected per frame.

The first principal contribution of this work is a systems one. End-to-end autonomous detection is demonstrated for difficult targets in badly cluttered images (many with poor local signal-to-background ratio) taken by a high-speed airborne laser-camera system in realistic environments.

The principal areas of innovation in the work reported here lie in the candidate detector (including the feature computation.) Techniques of gray scale morphological processing have been chosen for the initial image enhancement. The image data will be more completely described in Section 2; the following dominant qualities of the images resulted in morphological processing being chosen.

> The images contain a predominance of random sized replacement noise (patches with gray levels totally unrelated to the correct ones), which does not fit the additive linear noise model associated with classical signal processing.
>
> The attributes distinguishing targets from replacement/clutter are mainly geometric (shape/size).
>
> There is often gray scale background tilt far exceeding local target/background contrast.

Appendix A gives a brief review of morphological processing, as it applies to the work here. Reference 3 provides an excellent recent tutorial text; additional material can be found in References 4, 5, and 6. The specific sequence of morphological operations employed, and the associated design rationale, are described in Section 3.

Innovations in automatic thresholding from gray scale images to a binary detection image combine the background cancellation inherent in the morphological closing residue operation (see Appendix A) with two separate models of the histogram of the closing residue image. This combination of innovations (detailed in Section 3.1) rendered the thresholding feasible, automatic, and adaptive to each image, whereas straightforward thresholding of the original gray scale images would have been doomed by the extreme background tilt.

There are three innovations in the feature vector computation, introduced due to observed geometric attributes of candidate regions in the binary detection images. An elongation and major axis computation based on extremal points (Section 3.2.1, Reference 4), rather than an ellipse fit based upon second order moments, was introduced to measure the decidedly irregular and non-ellipsoidal regions commonly encountered. A "spindliness" measure discriminates between more or less compact targets and a spindly type of common processing artifact. A measure of detector saturation allows identification of (likely) spurious candidates arising when the enhancement processing and thresholding is occasionally overwhelmed.

2.0 DESCRIPTION OF GRAY SCALE IMAGE FRAMES

The data used here by APL-UW consisted of 1024 digitized sensor frames of 8-bit grayscale data. These frames consisted of eight sets (odd numbers 01 - 15) of 128 images, each chosen as representative of data from a separate sensor pass made during FY 89-90. Sets 01 - 11 were taken over shallow water, the beach, and nearby land. Sets 13-15 were data of opportunity taken over dry inland sites.

Frames varied in ground size and pixel count somewhat due to variations in altitude and in data collection/digitization. All were approximately 200 rows by 300 columns. Data sets 01 through 05 were of higher resolution, with a pixel edge "footprint" in the field of view representing about 0.6 the length used in sets 07 through 15. Figure 2-1 depicts two typical images at the higher resolution, each containing two minelike targets.

2.1 TARGET CHARACTERISTICS

Targets considered here all appeared darker-than-background (dtb) due to the light background at the test sites. Many displayed low contrast against the local background, to the point of often being difficult to see. Indeed careful inspection of Figure 2-1a is necessary to ascertain that there are not one but two target objects visible.

Minelike targets considered in this study included "large discs" and "small discs", having respective diameters about 36 and 15 pixels at the higher resolution, as well as "bars" and "long bars" of respective lengths 36 and 108 pixels at the higher resolution. They were located in uniform test arrays in shallow water and on the beach. On land, minelike objects were randomly scattered and placed near to existent non-target MMOs.

The discs, which constituted the majority of the visible targets, could be measured in image displays (at the higher resolution of data sets 01 - 05) as discs/ovals of about 36 and 15 pixels in (major) diameter, respectively, with small variability when cleanly visible. But several effects contributed to a large fraction of targets not having such simple appearances:

> Many targets were in the edge of the image, cut off by the image boundary;
>
> Some targets were physically occluded by being covered with sand from the test site, at the time measurements were made. Their effective shapes in the images ranged from "doughnuts" (e.g. leftmost target in Figure 2-1b) to "horseshoes" to partial discs;
>
> Some targets were occluded by the same "glint" processes, described in the next subsection, that occurred in the background (e.g. leftmost target in Figure 2-1a);
>
> The combination of background variability and low local SNR may make the target boundary obscure, even to an experienced human observer (e.g. leftmost target in Figure 2-1a).

Thus the majority definition of a target was darker than background over "most" of its area, at least "somewhat" round or ovaloid, and possibly with missing area at the boundary or interior. This mitigated strongly against any sort of shape-matching (matched filter or otherwise) that depended upon precise shape.

2.2 NON-TARGET MAN MADE OBJECTS (MMOs)

A development objective was to test and enhance the ability to reject non-target MMOs while recognizing targets. Accordingly 332 MMOs, of 15 types, were included in a total of 260 of the 1024 frames. MMOs, which ranged from substantially smaller to much larger than the targets, included large numbered, striped, or checkered objects (range markers and contrast/resolution calibration markers), buoys, rectangular and elliptical objects 15 to 36 pixels (at the higher resolution) in major dimension, telemetry tower parts, views of a boat and a car, tire tracks, and the top view of humans.

2.3 BACKGROUND CHARACTERISTICS

In general, backgrounds were complex; a large fraction contained (in addition to MMOs) obscuration due to reflection and transmission phenomena in the water and/or natural background clutter. Often these appeared at least superficially targetlike enough to cause false reports from the candidate detection phase, which had to be rejected in the classifier. Occasionally, background objects were so targetlike and strongly visible as to dominate true minelike targets in the final threshold to binary, causing missed targets. Appendix B shows examples of complex backgrounds.

For targets in shallow water, the most significant background characteristic was the presence of surface glint (as in Figure 2-1a.) Surface glint, which appears lighter than background and usually smaller than minelike targets, obscures the basic image under the water (replacement noise) rather than adding to it incrementally. This being the case, and with targets being physically occluded by sand as well, we were strongly motivated to employ Mathematical Morphology rather than more "classical" processing techniques based upon matched filtering, in our algorithms. The morphological opening and closing operations (cf. Appendix A) are well-suited for "papering over" large upward or downward spikes or plateaus in the grayscale image surface, which are best modelled by a (strongly nonlinear) maximum or minimum operation, for example:

Corrupted Image = Max(Original Image, Glint Image)

By comparison, optimal matched filter results are based on additive (linear) noise models. Image backgrounds often had a significant tilt (systematic change from one edge to the other), as exemplified in Figure 2-2. The contrast between object and background was often less than the background tilt; the rightmost object in the image of Figure 2-2, which has extremely low contrast, demonstrates this point. The effective use of simple thresholding (on the "raw" image frames) for detection is precluded in such images. The morphological opening and closing residue operations (cf. Appendix A) inherently cancel tilted backgrounds.

2.3.1 Background Categories

In the 1024 data frames, seven background types were identified by visual inspection prior to any tests of the algorithm reported here on those frames. Examples and underlying mechanisms are given in Appendix B. In summary one background type was simple, five contained substantial potential difficulties for the algorithm, and one (obtained, on opportunity, over dry inland sites) was seemingly impossible.

3.0 TARGET DETECTION ALGORITHM AND RESULTS

As indicated in Section 1.3, target detection is viewed as consisting of candidate detection, followed by classification. Candidate detection is itself composed of gray scale morphological image processing, auto-thresholding (with labelling of objects), and feature vector computation. Candidate detection will be described in Section 3.1, classification in Section 3.2, test protocols and performance results in Section 3.3, and promising future enhancements in Section 3.4.

3.1 CANDIDATE DETECTION

Before proceeding to the specifics of detection we provide top-level pictorial and block diagram representations, Figures 3-1 and 3-2 respectively, that show each image processing substep including glint removal, occlusion removal, gray scale detection, and binary detection. The autothresholds occur within these steps. (Each image processing substep consists of one or more morphological operations.) Thereafter the objects in the binary detection image are labelled and a feature vector is computed for each.

The morphological operations from input image B to binary detection image O are as follows.

Find Glint
$C = 1(B > \mu + 3\sigma)$
$D = B$ *closedby* square w_1
$G = 1(D$ *openres* square $w_2 > T_2)$
$I = (C \cup G)$ *dilatedby* square 3

Eliminate Glint
$K =$ *fill*(D,I)

Restore Target Occlusions
$L = K$ *openedby* square w_3

Normalize Background; Detect Binary Image
$N = L$ *closeres* square w_4
$O = 1(N > T_3)$

In the equations above, the morphological operations *dilatedby*, *closedby*, *openedby*, *closeres*, and *openres* are described in Appendix A; square w is a zero-height square structuring element with edge-length w; and image = 1(condition) denotes a binary image whose value is 1 where the condition is satisfied, and 0 elsewhere.

Sections 3.1.1 through 3.1.3 will now describe the motivation for the morphological processing steps. The subalgorithms for the *fill* operation and for choosing autothresholds T_i based on image statistics will be described in these sections, as will the rules for choosing structuring element sizes w_i based upon target geometries. Thereafter, Section 3.1.4 will discuss the connected component labelling of the binary detection image and Section 3.1.5, the feature vector.

3.1.1 Finding and Removing Glint

The most significant background characteristic for targets in shallow water was seen to be surface glint, a random lighter-than-background replacement noise. Without removing such interference, reliable detection of minelike targets would not have been feasible. Morphological algorithms seeking these darker-than-background targets would have produced numerous false detections on mine-sized "valleys" that happened to occur among the "peaks" represented by the glint.

Glint was eliminated by first finding a binary mask I of glint pixels, then replacing each value under that mask with the image value at the nearest point outside the mask. (The replacement is performed by setting image D to zero under the mask, then filling into the nulled region using nearest nonzero values found by a two-pass algorithm. Section 5.8 of Reference 4 describes an analogous algorithm for the morphological distance transform.)

Pixels qualify for mask I on the basis of two criteria arising from the nature of glint in typical images. In a minority of images, glint appeared as very bright, extended, irregular patches. A majority of glint appearances were relatively light, and smaller than targets. Thus, submask C was based on extreme amplitude alone, and submask G was based on an opening residue for finding objects smaller than square w_2 and brighter than (local) background.

Two steps in finding the glint mask remain unexplained. The initial closing by square w_1 (smaller than w_2) is standard morphology practice, to protect the subsequent opening residue from being "distracted" by tiny dark patches. The final dilation by square 3 remedied occasional loss of corners of glint regions.

Parameters T_2, w_2, and w_1 are set from image statistics and target geometry as follows.

$$T_2 = \max\{\mu + 2\sigma,\ 0.25 \times \max[\text{D } \textit{openres} \text{ square } w_2],\ T_3\}$$

where μ and σ values are measured for the opening residue image and the adaptive T_3 threshold process for *openres* and *closeres* images will be described in Section 3.1.3. To maximize the size of glint found while ensuring that (lighter) gaps between closely spaced (dark) targets are not mistaken for glint, we define $w_2 = (0.6)*(2)^{-0.5} * d$, (where d is the size, in pixels, of a large disc) based on the assumption that actual targets would not be placed closer together than the diameter of the largest disc target. (We correct by $(2)^{-0.5}$ to allow for diagonal placement, and by 0.6 as a safety factor.) To ensure that the initial closing is smaller than the opening residue it protects, and that it does not remove minelike targets, $w_1 = \{w_2/3, d_{min}/2\}$.

Readers familiar with morphological processing will note that we have not used a potential glint remover with fewer steps, namely D = B *closedby* square w_1; K = D *openedby* square w_2. Our rationale, in terms of reducing artifacts arising from the square w_2 structuring element, is given in Reference 2, including images containing artifacts and their impact upon detections.

3.1.2 Occlusion Restoration

Target occlusion, whether by sand or optical interference, constitutes light replacement of otherwise dark target pixels, hence restoration by an opening. Structuring element $w_3 = 0.6$ d, so that occlusions up to half the diameter of the largest target are restored.

3.1.3 Background Normalization and Binary Detection

With preliminary processing completed for the principal forms of small clutter and occlusion, the darker-than-background targets result in the only substantial bright objects in the final gray scale closing residue (grayscale detection) image. The residue process simultaneously removes large-scale background variation (e.g. background tilt.)

The final greyscale-to-binary detection is performed by a well-defined autothreshold sub-algorithm based upon image statistics (intensity histograms.) By virtue of the previous processing, the candidate targets to be detected are expected to cover a small fraction of the grayscale detection image and to have the highest amplitudes there (as was emphatically not the case in the original frames.) This dictates that the corresponding intensity histogram will have the majority of pixels close to zero; pixels to be retained will be found in a flat histogram "tail" or a second (or greater) mode or both.

The procedure for setting T_3 was motivated by the need to negotiate "blind" (i.e. automatically, and for all cases) in any histogram having the possibilities just defined. The T_3 threshold contains three steps as follows:

> First the histogram for image N is compressed into fewer bins and smoothed to prevent irregularities and "dropout" bins from distracting the remaining steps (which implicitly assume smooth histograms). Specifically, if 95% of the pixel gray levels lie between 0 and L_{95}, the number of bins is compressed by a factor L_{95} / 10, so that most pixels are in the first 10 bins of the new histogram. Then, a simple 3-point "boxcar" smoother is applied to the compressed histogram (without further compression.)
>
> Next, two separate models of the above histogram are created: T_{3A} is the Otsu threshold of Reference 8. T_{3B} is obtained by fitting an exponential model density curve $P(x) = ae^{-ax}$ to the first three histogram bins, then finding the threshold for a (model) false alarm rate of 0.001, so that $T_{3B} = (1/a) \ln(0.0001)$. Then $T_3 = \max\{T_{3A}, T_{3B}\}$ is the starting value before adjustments.
>
> Finally, further automatic upward adjustments in T_3 (decreases in the number of pixels detected) are applied iteratively; these serve to eliminate pathological cases where the two models both produce threshold values that lie before the first flat spot or upturn in the histogram. Specifically, if h(i) is the number of pixels in the histogram bin at the ith threshold iteration and h_{total} the total number of pixels , then adjust T_3 upward stopping on condition C1 AND C2, each defined as follows:
>
> C1: $[h(i+1) - h(i) < 0.0025\ h_{total}]$ OR
> [threshold is above 99.75% of the total pixels]
>
> C2: $[h(i+1) / h(i) > 0.65]$ OR
> $[h(i+1) < 0.000025\ h_{total}]$

The procedure outlined above, though lacking in elegant simplicity, has proved exceptionally successful at avoiding "misses" of targets clearly visible in the grayscale image on the one hand, and "saturated" candidate detection images on the other.

3.1.4 Labelling the Binary Detection Image

Image O is binary; it is labelled 1 for any of the candidate regions, and 0 elsewhere. Additional processing, to facilitate the computation of a feature vector for each separate detected candidate in the frame, is called the labelling of connected components (CCMs.) This produces a labelled image (denoted Q) in which the background is numbered 0 and the n separate connected objects (candidate regions) are numbered 1 through n. Our CCM labelling algorithm is a memory-efficient two-pass version, described in Sections 2.3.5 and 2.3.6 of Reference 4.

3.1.5 Feature Vector Computation

For each detected candidate region (CCM), a vector of eleven features (clues) is computed. Of the eleven, the first eight are standard.

F1: Area = number of pixels in CCM.

F2: Gray scale Variance = $[1 / (N-1)] \Sigma (I_{rc} - \underline{I})^2$, where $\underline{I}$ = mean gray scale value over CCM.

F3: Eccentricity (based on ellipse fit) = λ_1 / λ_2,
where $\lambda_{1,2} = 0.5 \{ d + f \pm \text{sqrt}[(d - f)^2 + 4(e)^2] \}$

and $$\begin{bmatrix} d & e \\ e & f \end{bmatrix} = 0.25 [\mu_{rr} \mu_{cc} - (\mu_{rc})^2]^{-1} \begin{bmatrix} \mu_{cc} & -\mu_{rc} \\ -\mu_{rc} & \mu_{rr} \end{bmatrix}$$

F4, 5: Mean of Radius = $(1 / N) \Sigma R_i$
Variance of Radius = $(1 / N-1) \Sigma (R_i - \mu_r)^2$
where pointwise radius $R_i = \text{sqrt}[(r_i - \underline{r})^2 + (c_i - \underline{c})^2]$, $(\underline{r}, \underline{c})$ = centroid,
and only boundary pixels of the CCM are used.

F6: Number of Boundary Pixels = Number of pixels in the CCM touching the background.

F7: Measure of Circularity = (Variance of Radius)$^{0.5}$ / Mean of Radius

F8: Gray Scale Gradient (magnitude) = $\text{sqrt}(\alpha^2 + \beta^2)$, for $\nabla(P_{r,c}) = (\alpha, \beta)$,
based on a least square fit of plane $P_{r,c} = \alpha r + \beta c + \varepsilon$ to the gray scale surface over the CCM.

The last three are innovations developed in response to specific problems or requirements encountered during the algorithm development process.

F9: Elongation (based on extremal points) characterizes aspect ratio because a large fraction of observed candidate CCMs bear no vague resemblance to solid ellipses, and thus are likely to be badly characterized by the conventional eccentricity feature F3. The elongation feature was based on the eight extremal points of any CCM as depicted in Figure 3-3 (formal definition: Section 3.2.1, Reference 4.). We used no model of shape, assuming only that the extremal points approximated the extent of the CCM in a broad sense. Thence the two extremal points farthest apart define an approximate major axis (and its length); width comes from the narrowest band, parallel to the major axis, encompassing all extremal points, as indicated in Figure 3-3.

F10: Spindliness = (major axis)^2 / pixel area, approximating the major axis using extremal points (see above). This quantifies a shape difference between true-target and false-detection CCMs (which may have similar elongation.) Many known false detection CCMs were skinny and meandered about regions of the frame.

(Intuitively, they were spindly.) True-target CCMs are rarely spindly. Spindliness of a disc is 1.27, of "compact-looking" objects up to about 1.8, of "spindly-looking" objects 2.2, 2.5 or more. An $n \times 1$ line has spindliness n.

F11: Detector Saturation = (area of all CCMs in frame) / (area of frame) provides a distinguishing measure for CCMs that come from occasional frames in which the detection (probably the thresholding) works poorly, so that the final binary image has a large fraction (say, 20% or more) of its area covered with meaningless objects. By comparison, correctly detected frames, even those containing the maximum number of true targets seen in the data sets, rarely filled more than 10-12% of the binary detector image. To "inform" the classifier of frame credibility (even though the classifier works on individual CCMs) we compute a frame-wise feature of this fraction and assign its value to all CCMs in the frame.

The 11 - dimensional feature vector for each candidate CCM (labelled according to its frame of origin, and location in the frame) is the essential item handed from the detector to the classifier.

3.2 CLASSIFICATION

A binary decision tree was used. The theory of such classifier algorithms is known; see for example Section 4.9 of Reference 4, or Chapters 2 and 3 of Reference 9. In particular we implement Fisher's Linear Decision Rule, at each node, to divide feature space. This can be viewed as separating feature-space into half-spaces by a hyperplane, then linearly dividing each half-space again, etc. A purity (entropy) measure, described in Reference 4, was used to optimally adjust the hyperplane for each node.

Training is done in a space of feature vectors for which the true class is known and used. (The true class is determined by software which compares each candidate detector output CCM with groundtruth; this true class is then appended to the corresponding feature vectors for training.) Given that a human has entered groundtruth information for the training images, thereafter the tree structure, separating hyperplanes, and stopping criteria are all automatically created.

In testing (actual execution of the classifier), the unknown candidates are dropped through the same tree, with the same decision rules (same weighting vectors and thresholds) created in training. Figure 3-4 shows the relationship between the detector, training of the classifier, and testing by the classifier.

3.2.1 Equalization

A persistent problem was found in early development tests. When the numbers of training vectors for one class was much higher than for another class, the resulting training emphasized correct classification of the majority class at the expense of the minority class. Among detection candidates (prior to the classifier) false target detections are usually in the majority by a factor of at least 2.5. This has the potential for resulting in good classifier false alarm rates at the expense of poor probability of correctly classifying true detections.

The cure was to equalize the count - specifically to weight the number of vectors of both classes inversely proportional to the original numbers of training vectors in computation of the best adjustment of the hyperplane for each node and in labelling the terminal nodes as "target" or "notarget" at the end of training.

3.3 TEST PROTOCOLS AND PERFORMANCE RESULTS

The basic test protocol divided the frames into two groups, choosing them alternately in numerical order. Results were generated by training on group a and using the trained classifier to test group b, and the converse, followed by adding the performance totals for the two tests. This is abbreviated "train a test b; train b test a."

Per frame statistics were used to describe performance. In particular we measured the correctness of the report "any targets vs no targets" which is the most important bottom line provided by our single frame target detection algorithm to the multiple frame target-field detection algorithm which is the "consumer" of our results.

In light of the great variability among image backgrounds, performance is defined in terms of regimes of background types. For the regime consisting of the majority of backgrounds, namely backgrounds B1, B2, B4, B5, and B6 described in Appendix B, performance is described below. This constitutes all backgrounds except those labelled "dark blotch" and "dry inland."

Frames With Target(s)	103
Correct Frames	95 (92%)
Target Frames / Error Frame	13
Frames Without Targets	477
Correct Frames	437 (92%)
No-Target Frames / Error Frame	12
All Frames	580
Correct Outcome	532 (92%)
Frames / Error Frame	12

This 92% performance (1 error / 12 frames) is the more remarkable because it includes all false positives due to non-target man made objects (MMOs). (Section 3.3.1 considers effects of MMOs on performance.)

Performance for the "dark blotch" background regime (B3) obtains presence vs absence of targets in a frame correctly in only 67% of frames, with equal failure rates in frames with and without targets. The fundamental cause is that within measurements made by the present algorithm, the dark vegetation clumps are just too "targetlike" - indeed often more so than the targets. Innovations in the feature set, with good potential for upgraded performance (but beyond the resources of the work reported here) will be described in Section 3.4.

In the case of the "dry inland" background (B7), the shortfall is entirely in target frames, of which only about 25% are correctly found. (Frames without targets are correct 88% of the time.) This background regime, which occurs in the vast majority of the Set 13-15 frames, and only there, appears too obscure to suggest likely progress; it was recognized as such when it was taken on. Indeed humans, working at interactive displays with enhancement tools, often have difficulty finding the targets in these images.

3.3.1 Performance Against Non-Target Man Made Objects (MMOs)

Viewing MMO-only frames (frames with one or more nontarget MMOs but no targets) by themselves, some 17% have one or more candidate detections that survive the classifier to cause error frames. But since the fraction of frames containing MMOs is only 155 out of 1024, and only some of those are MMO-only, the overall effect on error rate is small.

Furthermore, inspection of performance against specific MMO types reveals that 77% of all false positives due to MMOs result from four types of large objects with high contrast areas, namely resolution panels, contrast panels, number panels, and a boat. Of these four types, three are test-range artifacts. For the remaining types of MMOs, false positives occur only 9% of the time.

3.4 PROMISING FUTURE ENHANCEMENTS

For enhancing performance in the majority background regime beyond 92%, and for establishing algorithm robustness where there is very "targetlike" clutter (as in the "darkblotch" background regime), we see promise in new elements (clues) for the classifier feature vector. These clues combine morphological processing to establish relevant subregions, with gray scale image statistics for those subregions.

The motivation for these clues lies in the following two types of difficult target/background discriminations, which dominate error in the "majority" and "darkblotch" background regimes respectively:

> Artifacts having no obvious basis in background clutter objects, but nonetheless assuming fairly target-like size and shape at the output of the candidate detector;
>
> Background clutter that is target-like in general attributes, and thus is in a sense properly found by the candidate detector and left for the feature vector and classifier to discriminate.

We will call the resulting target-like CCMs processing artifacts and clutter detections, respectively. Because of differences in gray scale image directly over and nearby the respective CCMs, these two types of erroneous detections call for similar but subtly different innovations in classifier clues.

Processing artifacts may be viewed as resulting from targetlike CCMs located randomly in the frame; thus the gray scale image over any such CCM should be very similar to the image over the nearby region. (In this regard, processing artifacts are unlike targets; most visible targets are in some gray scale sense quite different from their surroundings.) Discrimination between target and processing artifact CCMs ought to be improved by clues that compare gray scale statistics over a detected CCM with those over a similar-sized annulus outside that CCM.

Obtaining annuli for the CCMs is a task suited for binary morphology. The task is made nontrivial by the desire to compute all annuli in one or a few passes through the image, suggesting a binary dilation followed by set-subtraction of the original CCMs. Unfortunately such annuli could overlap initially; fortunately good morphological segmentation algorithms exist, based upon the distance transform and watershed concepts (Section 3.5 of Reference 3.) Once each candidate CCM has an annular background region, any of the following clues can compare background b and candidate c:

$$\text{Contrast} = \mu_b - \mu_c$$

$$\text{Normalized Contrast} = (\mu_b - \mu_c) / (0.5\ \sigma_b^2 + 0.5\ \sigma_c^2\ )^{0.5}$$

$$\text{Relative Graininess} = \sigma_b^2 / \sigma_c^2$$

In the case of clutter detections, the gray scale image over the CCM should differ in some important ways from the gray scale image nearby. However those differences may be very similar to the ones for an actual target CCM and its surroundings. But we observe that the gray scale boundary transitions often are more gradual and irregular for clutter detections than for targets. These differences in gray scale boundary transitions should be exploitable by a clue employing statistics of the gray scale image, or its gray scale morphological gradient (Section 7.1 of Reference 3), evaluated over a narrow annulus n (Section 3.2 of Reference 3) from inside to outside the boundary pixels of c, the original candidate CCM. For example:

$$\text{Relative Boundary Irregularity} = \sigma_n^2 / \sigma_c^2$$

$$\text{Gradient Strength (Normalized)} = \mu_n(G) / \sigma_c$$

where G = gray scale gradient image, and μ could also be max, or any order statistic

The clues defined above for broad and narrow annuli may be considered local veracity clues.

In addition to differing from false detections in some form of boundary contrast, it appears that targets may differ from clutter detections or processing artifacts in terms of grayscale uniformity within the CCMs. This suggests a gray scale uniformity feature obtained from the mean or the maximum, over each candidate CCM, of the difference image B *closedby* K – B *openedby* K derived from input image B. Grain size K should be chosen to eliminate the smallest common image irregularities

4.0 CONCLUSIONS ON PERFORMANCE

Successful detection of single frames containing minelike targets has been demonstrated in the majority of backgrounds presented, including difficult backgrounds in shallow water, on the beach, and on nearby land. This performance, 92 per cent correct both for frames with and without targets, should be sufficient to drive the multi-frame detection algorithm which is the "consumer" of its output.

False positives due to (non-target) man made objects (included in the 8% error rate above) have required no special attention, given that frames containing such objects do not occur much more frequently than in our data (155/1024 frames had MMOs.) The few classes of non-target objects offering potential problems if they were numerous, have been documented; most of these are testing-range markers.

Two background types have been delineated, for which a performance shortfall remains. For "darkblotch" backgrounds, containing clutter superficially resembling minelike targets, promising approaches exist, though beyond the scope and resources of the work reported here. The images from dry inland sites were acknowledged as marginal when the project began; indeed a careful human operator often has difficulty there.

ACKNOWLEDGEMENT

The AMDAS program was funded by the Marine Corps Research and Development Acquisition Center (Code AW; presently Code AWT, MARCORSYSCOM -- Mr. Dave Vaughn), with technical direction and sponsorship via the Signal and Image Processing Branch of the Naval Coastal Systems Center (Code 20T) -- Dr. Elan Moritz; presently Code 10T1, Coastal Systems Station, Dahlgren Division, NSWC/DD, Panama City, FL -- Dr. Richard Detsch, acting. The AMDAS program was the first priority Advanced Technology Demonstration (ATD) for the U.S. Marine Corps in FY 89-91.

REFERENCES

1. Somani, A.K., et. al., "Proteus System Architecture and Organization". *Proc. 5th International Parallel Processing Symposium*, Anaheim, California, June 1991, pp 287-294.

2. Rystrom, L. R., Katz, P. L., Haralick, R. M., and Eggen, C. J., "Morphological Algorithm Development Case Study: Detection of Shapes in Low Contrast Gray Scale Images with Replacement and Clutter Noise". In *Nonlinear Image Processing III*, Proc. SPIE, Vol. 1658, San Jose, California, 1992, pp 76-93.

3. Dougherty, E. R., *An Introduction to Morphological Image Processing*. Tutorial Texts in Optical Engineering, Vol. TT9, SPIE Optical Engineering Press, Bellingham, Washington, 1992.

4. Haralick, R. M., and Shapiro, L. G., *Computer and Robot Vision*, Vol. 1. Addison-Wesley, Reading, Massachusetts, 1992.

5. Giardina, C.R., and Dougherty, E.R., *Morphological Methods in Image and Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1988.

6. Serra, J, *Image Analysis and Mathematical Morphology*. Academic Press, Orlando, Florida, 1982.

7. Haralick, R.M., Somani, A.K., et. al.,"Proteus: A Reconfigurable Computational Network for Computer Vision". Proc. SPIE, vol. 1659, Image Processing and interchange: Implementation and Systems, San Jose, CA 1992. pp 54-76.

8. Otsu, N., "A Threshold Selection Method From Gray Level Histograms". *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-9, 1979, pp 62-66.

9. Breiman, L., et.al., *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA, 1984.

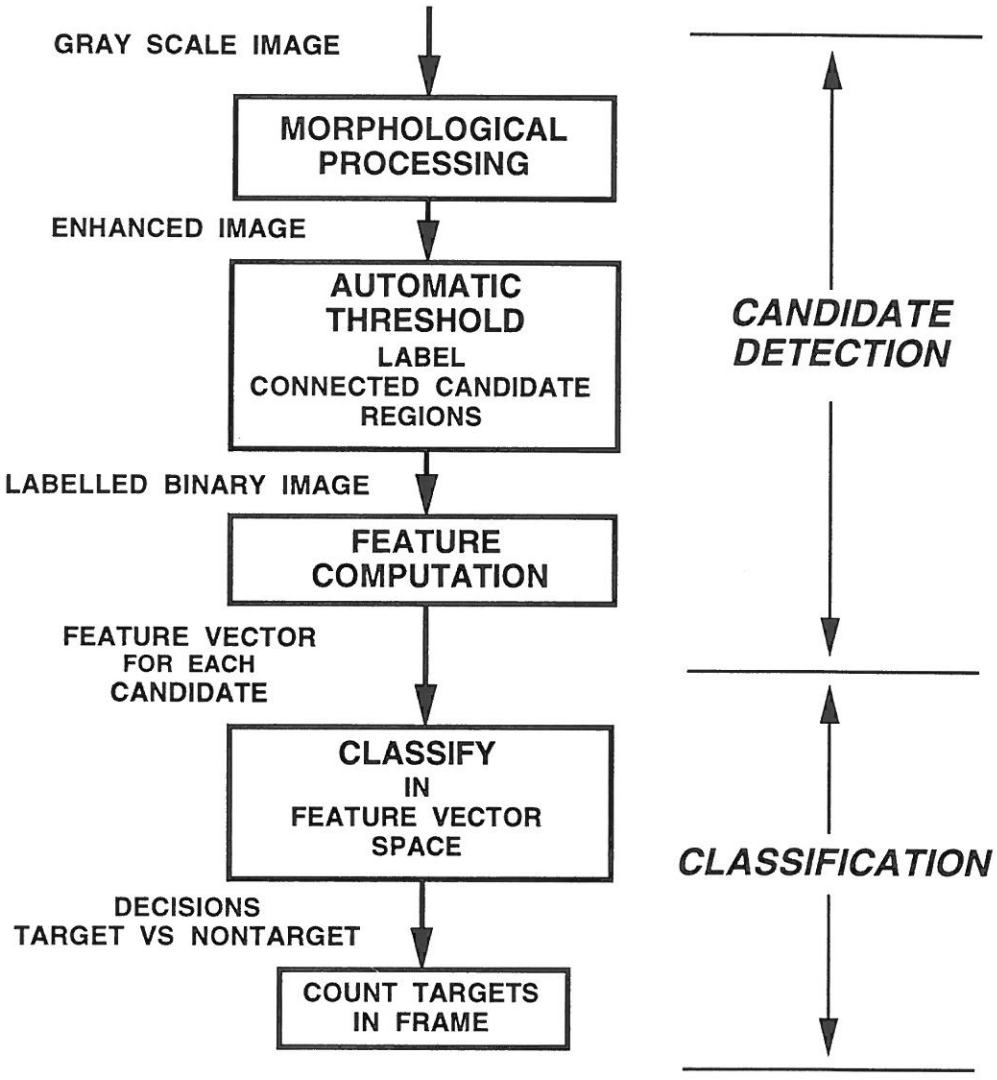
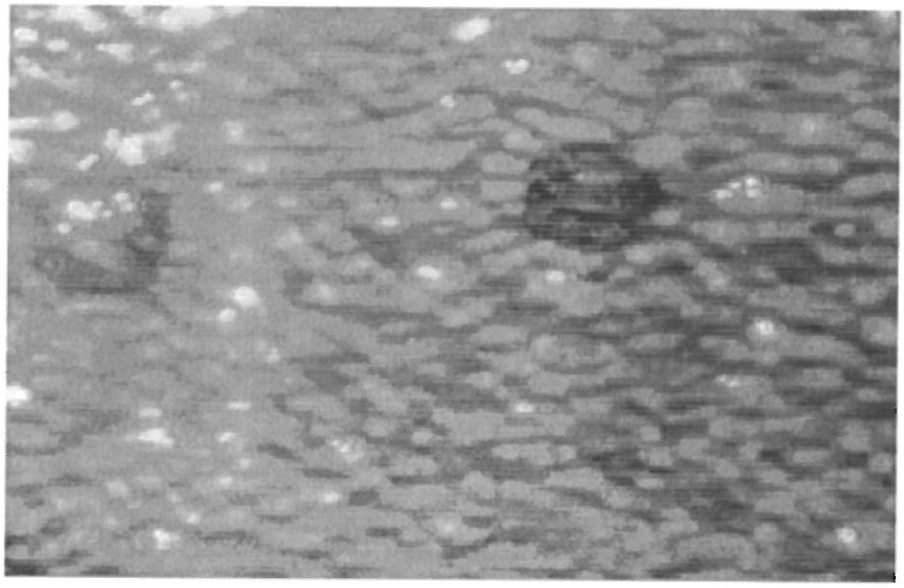


FIGURE 1-1. SINGLE - FRAME TARGET DETECTION ALGORITHM.

(a)



(b)

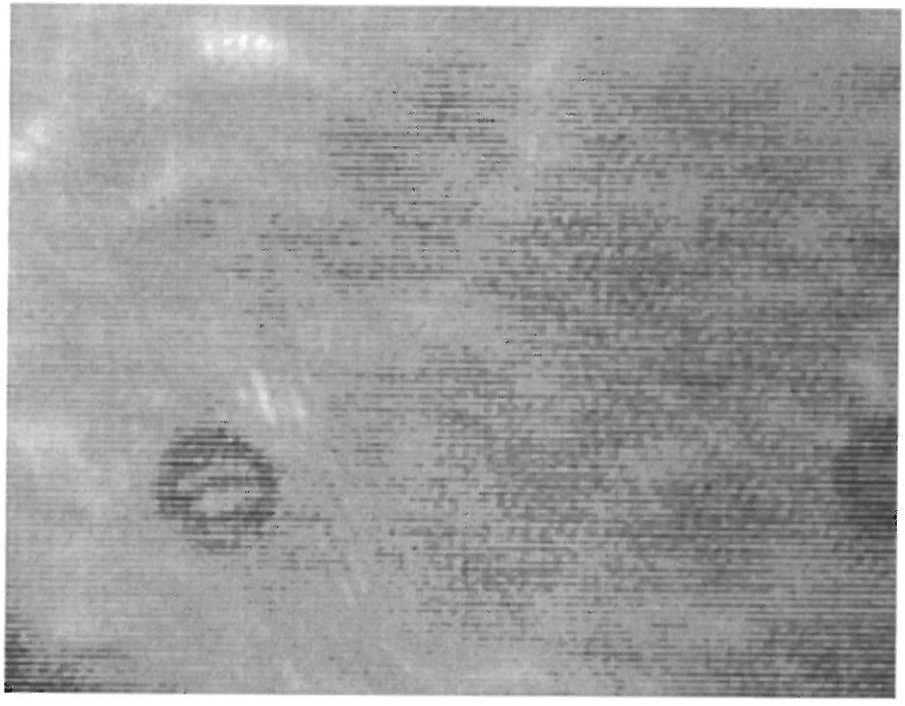


FIGURE 2-1. TYPICAL IMAGES, EACH WITH TWO TARGETS
(a) DEMONSTRATING GLINT; (b) DEMONSTRATING TARGET OCCLUSION

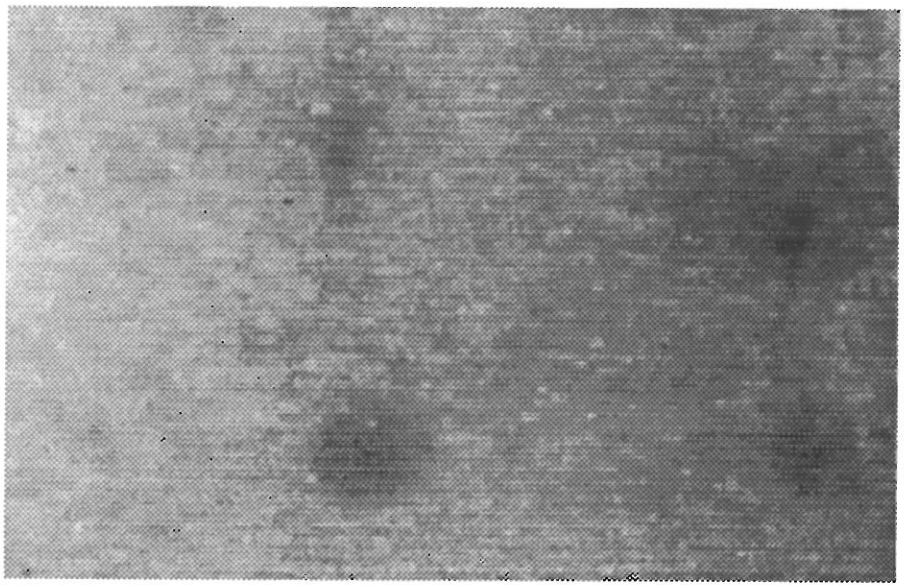


FIGURE 2-2. (U) IMAGE WITH BACKGROUND TILT EXCEEDING CONTRAST BETWEEN TARGET AND BACKGROUND

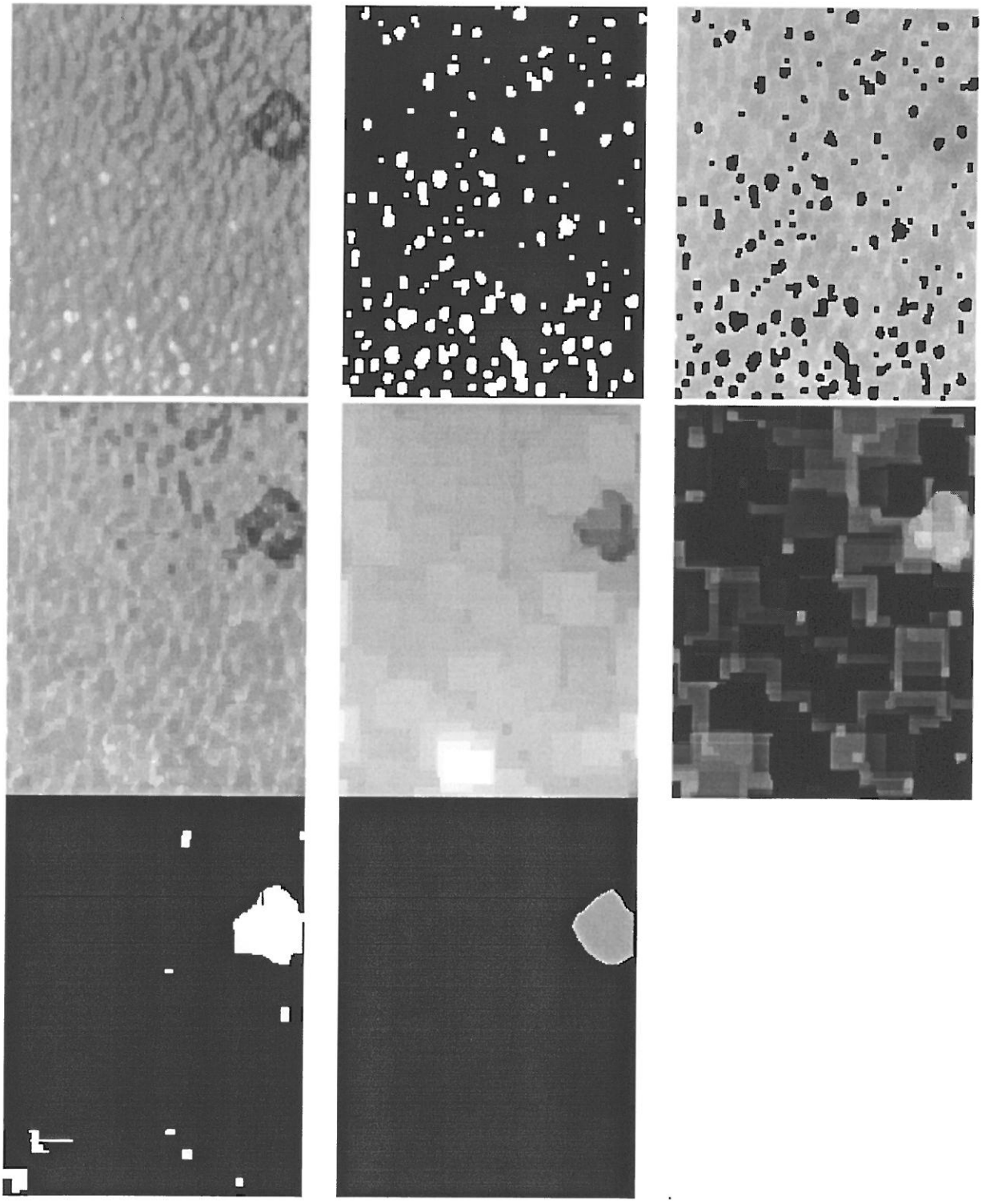


FIGURE 3-1. CANDIDATE DETECTION BY MORPHOLOGICAL IMAGE PROCESSING
Row 1: Original (B), Glint Mask (I), Glint Nulled; Row 2: Glint Eliminated (K), Occlusion Restored (L), Background Normalization / Gray Scale Detection (N), Row 3: Binary Detection (O), Ground Truth.

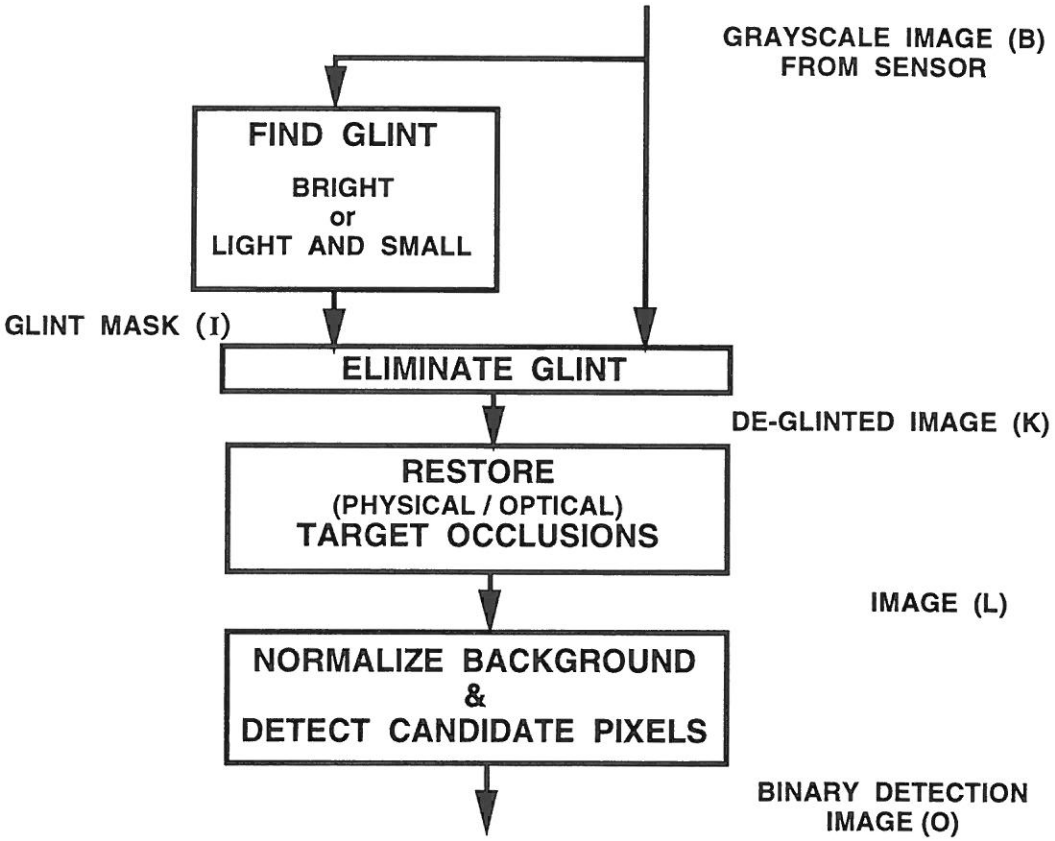


FIGURE 3-2. FLOW DIAGRAM OF CANDIDATE DETECTION

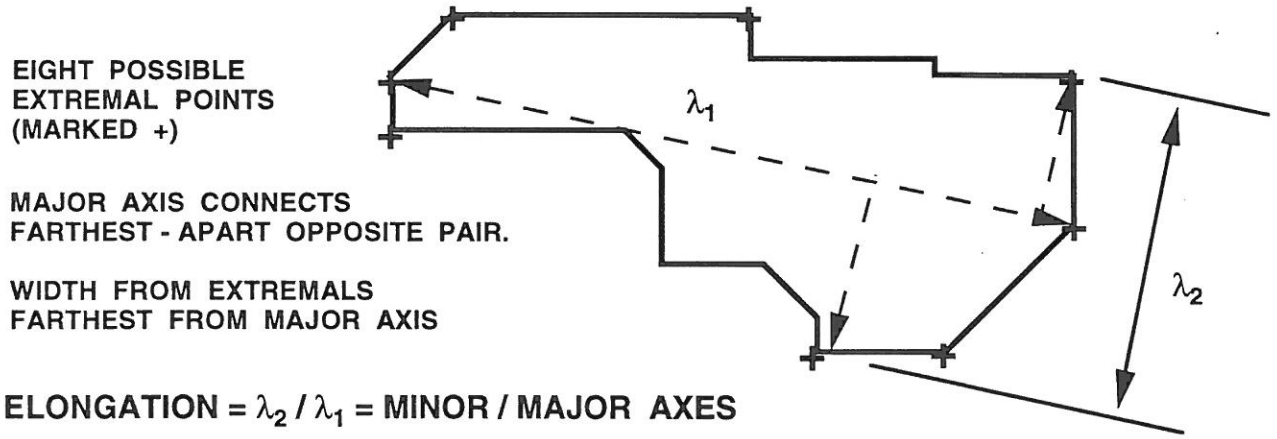


FIGURE 3-3. FEATURE 9 -- ELONGATION BASED ON EXTREMAL POINTS

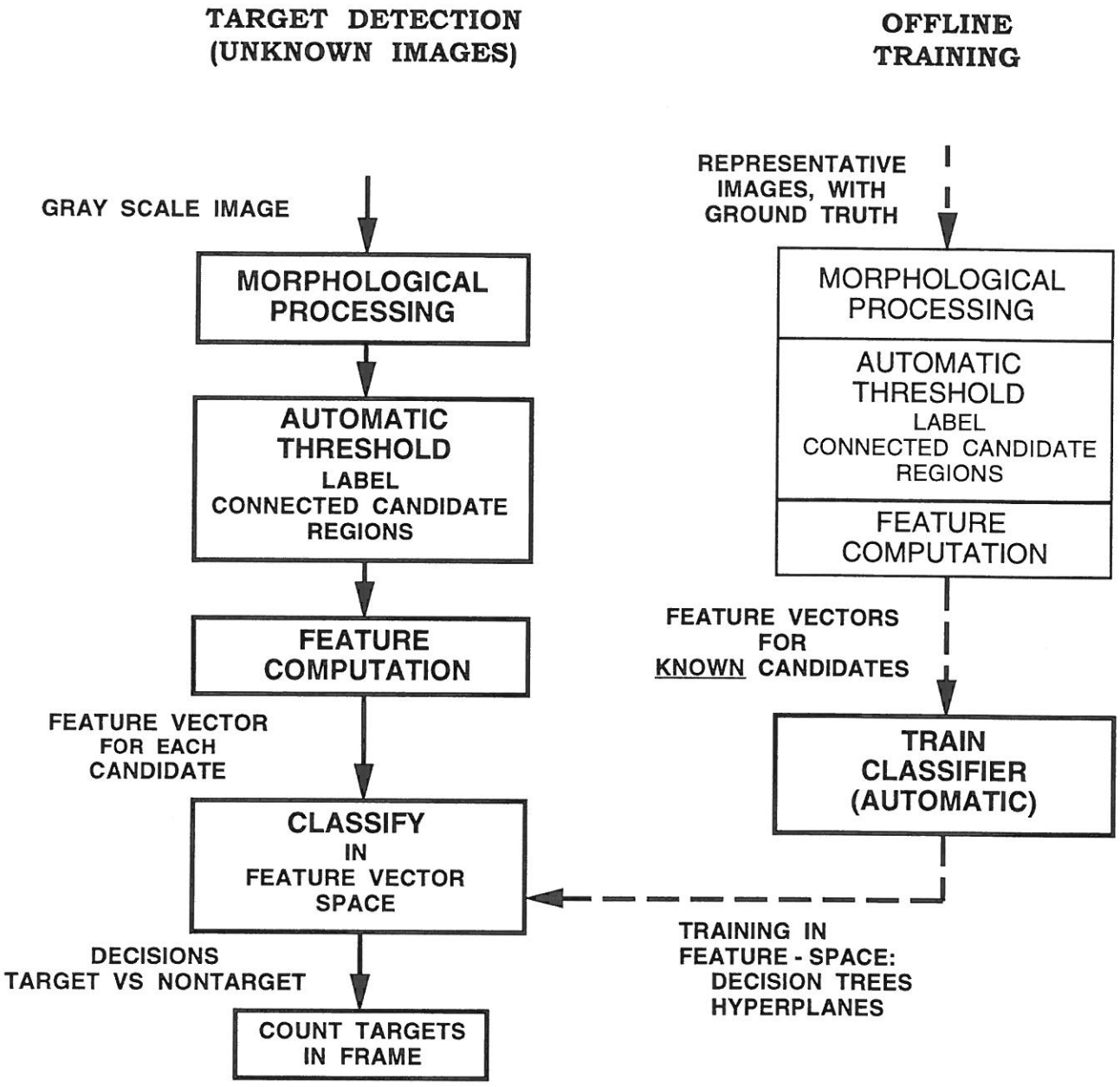


FIGURE 3-4. AUTOMATIC OFF-LINE TRAINING OF CLASSIFIER, USING IMAGES WITH KNOWN GROUND TRUTH

APPENDIX A BRIEF REVIEW OF MORPHOLOGY

Candidate detection is based on morphological image processing, and specifically the gray scale closing and opening operations depicted in the first column of Figure A-1. Furthermore, differences between an image and its opening (as depicted in the second column of Figure A-1), and between an image and its closing, can be used to detect objects smaller than a given set (structuring element) and ignore background variation. Our design process is based on such visualizations and the geometric effect, in successive gray scale surfaces, of a sequence of morphological operations.

The closing, opening, closing residue, and opening residue operations are defined, in turn, from dilation and erosion primitives, as follows:

$$G \textit{ closedby } D = (G \textit{ dilatedby } D) \textit{ erodedby } D; \quad G \textit{ openedby } D = (G \textit{ erodedby } D) \textit{ dilatedby } D.$$

$$G \textit{ closeres } D = (G \textit{ closedby } D) - G; \quad G \textit{ openres } D = G - (G \textit{ openedby } D) .$$

where G is a gray scale image and D is a structuring element. (The latter fills a role analogous to that of the FIR filter kernel in linear processing.) Note that since closings are always extensive ($G \leq G$ *closedby* D) and openings are always antiextensive (G *openedby* $D \leq G$), per Reference 3, residues are always non-negative.

The underlying gray scale morphological primitives are defined as follows. For gray tone image G, over set Γ, and structuring element D over set Δ, gray scale dilation and erosion are:

$$(G \textit{ dilatedby } D)(x) = \max\{ G(x - z) + D(z) \}, \text{ where the max is over } z \in \Delta, (x - z) \in \Gamma;$$

$$(G \textit{ erodedby } D)(x) = \min\{ G(x + z) - D(z) \}, \text{ where the min is over } z \in \Delta.$$

The output image domains change according to $\Gamma \rightarrow \Gamma$ *dilatedby* Δ for dilation and $\Gamma \rightarrow \Gamma$ *erodedby* Δ for erosion. The gray scale primitives are generalizations of binary dilation and erosion, defined as follows, for arbitrary sets F and L.

$$F \textit{ dilatedby } L = \{x |\ x = f + l \text{ for some } f \in F \text{ and some } l \in L\}, \text{ and } F \textit{ erodedby } L = \{x |\ x + l \in F \text{ for all } l \in L\}.$$

(Note that that in the algorithm contained in this paper, only one binary morphological operation is used, dilation of the glint mask by a 3X3 pixel structuring element, to enlarge that mask by a pixel in all directions.) References 3-6 consider the properties of morphological operations extensively.

Note that image size (domain) changes during the gray scale dilation and erosion operations. This is a mathematical resolution of the edge condition problem common to shifting mask algorithms (including linear filters.) Mathematically, results expand where they can be fully defined outside the original domain, and are left undefined where the lack of image data outside the edge would require local modification of the definitions. In applications, padding is added to the edges of the original image so that the sequences of morphological operations never encounter the shrinking and growing, within the original image frame. By so doing, and trimming afterward, we can always display and consider images the size of the input image. (In computational practice, literal expansion of the image may, in fact, be replaced by conditions and special functions that operate at and near the edges, and have equivalent effect to padding.) The spatial extent of padding depends upon the extent of the structuring element. Choice of appropriate padding values is a design decision equivalent to assumptions upon objects only partly visible in the original image. We used "level" padding; each pixel in the padded region had value equal to the closest pixel in the image.

Most significantly, the choice and use of a padding rule must be made if unexpected and potentially undesirable effects are to be avoided not only at the edges, but within a structuring element width of them. For example, Figure 9 of Reference 2 demonstrates that to properly detect even the simplest shapes near the boundary of a clean, noiseless image, proper padding must be performed.

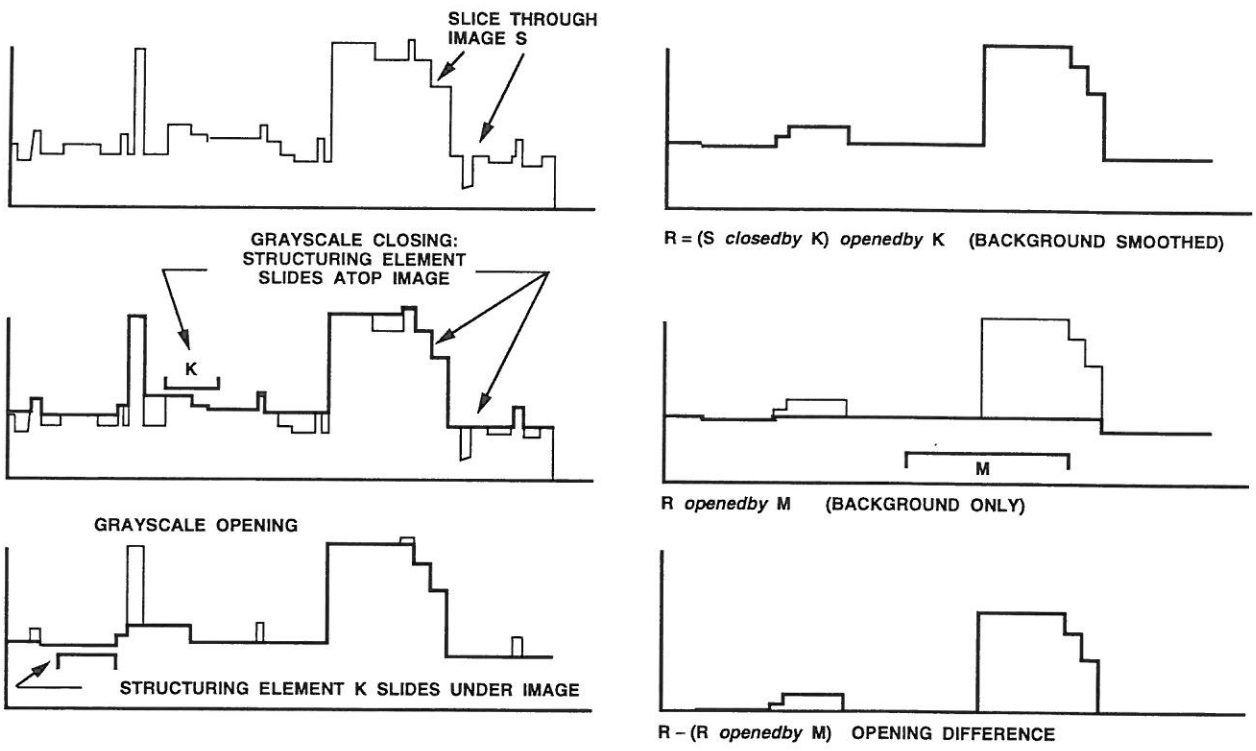


FIGURE A-1. GEOMETRIC VISUALIZATION OF GRAY SCALE MORPHOLOGY. COLUMN 1: CLOSING, THEN OPENING OPERATIONS; COLUMN 2: OPENING RESIDUE (DIFFERENCE). Note the zero background level in the residue, corresponding to cancellation of background variation simultaneously with final detection.

APPENDIX B BACKGROUND TYPES

Seven background types were identified by APL-UW investigators based upon viewing the 1024 test images prior to algorithm testing. Type B1 occurred mainly at site 1 (sets 01-11), with a few at site 2 (sets 13,15); types B2-B6 occurred only at site 1; type B7 occurred only at site 2. Backgrounds are depicted in Figure B-1.

B1: Flat Background. These occur in all data sets 01 through 15; such backgrounds are typical of the beach domain. There is some fine-grain variability, but (except for MMOs) there is no natural clutter even beginning to look target-like. Some images display background tilt, and some display corners with dark shading (particularly the lower left corner.)

B2: Glint Background. Glint arises from reflections of active illumination at the water surface. It is often much brighter than background as illustrated in Figure B-1. Alternatively it may merely be lighter than background and cover areas mainly smaller than targets. Where the glint "objects" are extensive enough to cover a significant fraction of the frame, glint cleaning will not occur, and the remaining glint has to be eliminated by the occlusion removal and background normalization preceding final thresholding.

B3: Dark_Blotch Background. This background arises from natural clutter objects, mainly dark vegetation clumps on and near the beach. Many blotches superficially resemble large and small disc minelike objects, or larger versions of those same targets. In general the blotch edges are more irregular than minelike target edges, and the blotches are a bit "grainier" than are targets. (Here graininess is used analogously to the term in film photography, though the physical mechanism is very different.)

B4: Mottle Background. A relatively infrequent phenomenon, mottle results from scattering of the illumination within the water column. It is distinguished by compact, light, generally target-sized "islands" close together with narrower dark bands "channels" separating them. There are also a few target-sized dark areas, larger than the channels (although these blotches are very different from those of the previous section, which were due to actual objects.)

B5: Glint_Dark_Blotch Background. These backgrounds, found mainly in data sets 03 and 05, occur when the physical processes creating extensive glint leave target-sized dark voids between the glint.

B6: Sparkle Background. Sparkle is an extreme form of glint. The potential difficulties are two. First the very dark voids between glint "peaks" could appear as strong, well defined targets. Secondly, the glint appears in such large clumps, and with so much very bright area, that the glint removal algorithms are at least partly foiled. Relative valleys within areas of surviving glint peaks are then potentially prone to false recognition as targets.

B7: Dry Inland Background. This background, found only in data sets 13 and 15 taken at site 2, was on the one hand highly textured with a variety of dark (and to a lesser extent light) shapes, and on the other hand essentially impenetrable. Virtually every target was difficult for a human, with high-level vision skills, to discern (e.g. In Figure B-1, there is only one target in either of the dry inland examples -- in the second of the two, about 25% of the way from left to right, and 40% up from the bottom.) Visually, the predominant texture was undulating in character, which is to say many mild peaks or indentations existed. These, coupled with the extremely low contrast of actual targets, mitigate against a good detection rate without excessively many false positives.

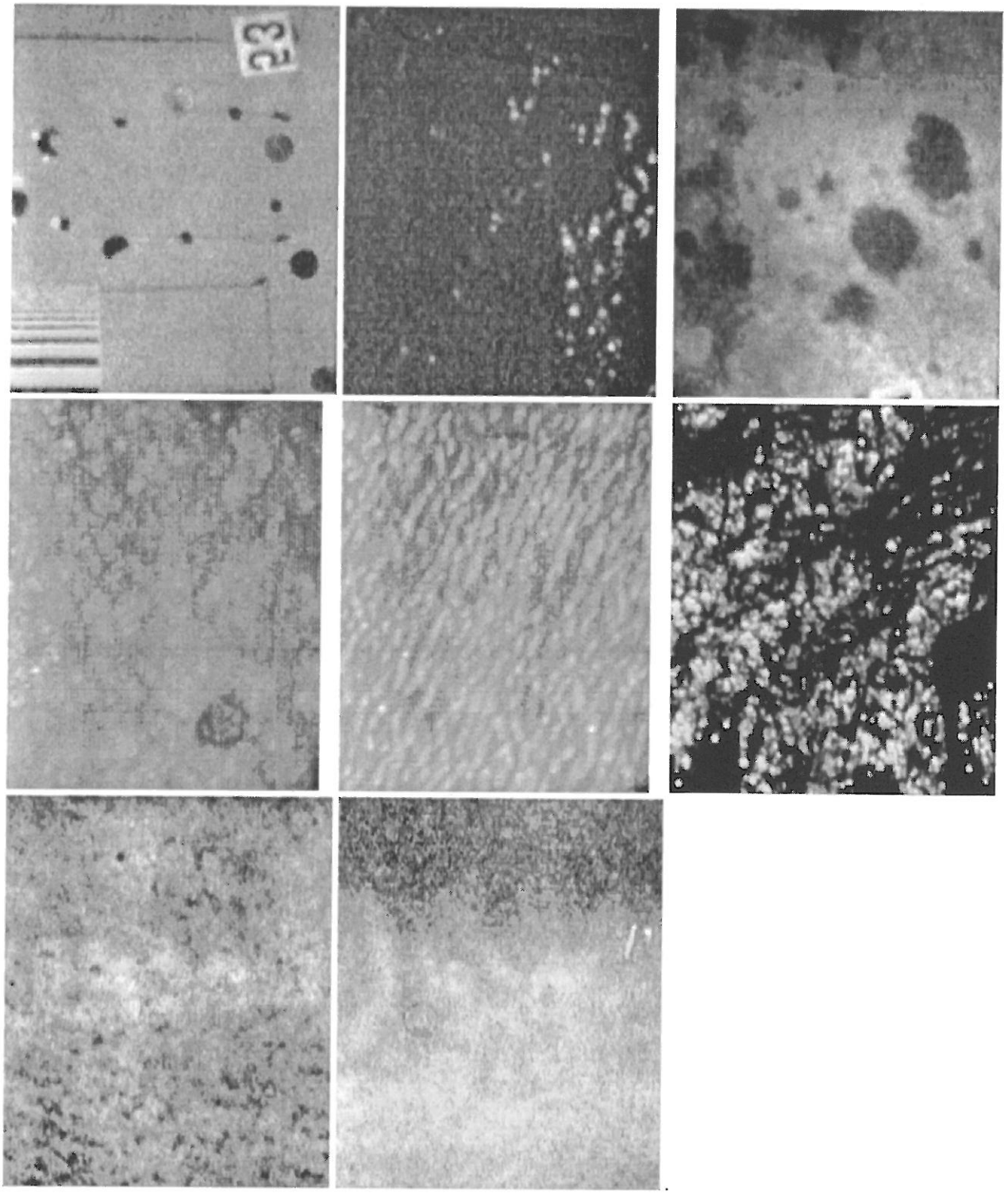


FIGURE B-1. BACKGROUND TYPES, AS DESCRIBED IN APPENDIX B TEXT
Row 1: Types B1, B2, B3; Row 2: Types B4, B5, B6; Row 3: Type B7 (two examples)