1-1-2000

Clutter modeling in infrared images using genetic programming

Liviu I. Voicu

Mosleh Uddin

Harley R. Myler

Anthony Gallagher
University of Central Florida

Julien Schuler

Clutter modeling in infrared images using genetic programming

Liviu I. Voicu
Mosleh Uddin
Frontier Technology, Inc.
12151 Science Drive, Suite 102
Orlando, Florida 32826
E-mail: l.voicu@imath.com

Harley R. Myler
Anthony Gallagher
Department of Electrical and Computer Engineering
University of Central Florida
Orlando, Florida 32826
E-mail: hrm@engr.ucf.edu

Julien Schuler
Electrical Engineering Department
INSA Lyon
69621 Villeurbanne Cedex, France

Abstract. Background clutter characterization in infrared imagery has become an actively researched field, and several clutter models have been reported. These models attempt to evaluate the target detection and recognition probabilities that are characteristic of a certain scene when specific target and human visual perception features are known. The prior knowledge assumed and required by these models is a severe limitation. Furthermore, the attempt to model subjective and intricate mechanisms such as human perception with general mathematical formulas is controversial. In this paper, we introduce the idea of adaptive models that are dynamically derived from a set of examples by a supervised learning mechanism based on genetic programming foundations. A set of characteristic scene and target features with a demonstrated influence on the human visual perception mechanism is first extracted from the original images. Then, the correlations between these features and detection performance results obtained by visual observer tests on the same set of images are captured into models by a learning algorithm. The effectiveness of the adaptive modeling principle is discussed in the final part of the paper.
[S0091-3286(00)01109-0]

Subject terms: clutter modeling; genetic programming; infrared; target detection; visual perception; data modeling.

Paper 990451 received Nov. 14, 1999; revised manuscript received Feb. 21, 2000; accepted for publication Mar. 8, 2000.

1 Introduction

A wide variety of models have been developed in the past decades that were aimed at predicting the probability of detecting or recognizing military targets in a given scene characterized by clutter.[1–9] Most of these models incorporated psychological and physiological findings about the *modus operandi*[6,7] and the acuity[8,9] of the human visual search mechanism. Although some of the classic models have achieved large popularity, a series of recent publications argued about the necessity to modify or update them in order to compensate for some observed limitations.[10,11]

In our opinion, the limitations of these models could be partly due to the fact that they attempt to capture a diversity of distinct factors using simple and general formulas. Even modeling the visual acuity of a single observer is an intricate task, since perceptive abilities of individuals are strongly dependent on subtle and subjective factors.

It is also noteworthy that, as has been pointed out in the literature,[12] significant differences exist between an imager functioning in the visible spectrum and one in the infrared. Indeed, an imager in the visible domain relies on reflected light and displays objects with minor internal intensity variations. On the other hand, an imager functioning in the infrared domain relies on irradiance information. Objects such as running vehicles can have significant temperature variations between their structural parts. These temperature variations generate a highly multimodal intensity distribution across the object surface and make it sometimes very difficult to detect the boundaries of an object in infrared imagery. Some researchers[12] summarize these issues by saying that the primary visual information from a visible-light imager concerns shape, whereas the infrared imagers are more cue-based. Under these circumstances, the criteria developed for predicting performance in the visible domain, such as Johnson's criteria,[13] may be inadequate for infrared. Several authors have focused their efforts on performance prediction specific to infrared imagery.[14,15]

Our conjecture is that a specific model, rather than a general one, may be a solution to some of these limitations. In other words, different models should be generated for different situations, observers, types of imaging equipment, etc. However, in this case, due to the specificity of the model, most of the significant factors involved will be almost impossible to predict analytically. The alternative is to let a heuristic technique learn from examples and generate a model specific to the context under which the examples were generated. This latter approach presents several advantages, given that an appropriate learning scheme is employed. Among these advantages, a very important one is that the necessity of prior knowledge about targets and factors involved is much diminished. Furthermore, this method should be able to generate performance models applicable to both human observers and seekers.

Evidently, the learning scheme outlined must be based on a supervised learning architecture. In the training data file, one must include an objective value that represents the measured performance of the observer/seeker on a set of images. In correspondence with the objective value will be

a list of features considered influential for this performance, such as target contrast and extent, or various clutter metrics from the literature.[3–5,14] In this context, the goal of the modeling system will be to extract the correlations that exist between the observed performance and the features presented and capture them into an analytical model.

In order to accomplish the task defined above, the learning method must meet two criteria. First, the models must be issued in a closed, analytical form, so that they can be interpreted. This criterion disqualifies the neural networks, which have been traditionally employed in predicting the behavior of highly complex processes, as potential candidates for the learning method. Secondly, the proposed method must be able to automatically select the features that are actually relevant from the provided set of features.

An adaptive modeling system that meets these requirements and is based on genetic programming foundations has been developed and is introduced in the next sections.

2 Data Modeling by Genetic Programming

As the principal component of the adaptive modeling system introduced above, we have formulated an approach based on genetic programming[16] (GP). GP has emerged in recent years as a powerful tool in search, optimization, and classification applications. Closely related to genetic algorithms (GAs), GP is a paradigm based on principles advocated by students of natural evolution. GP is concerned with developing simple computerlike programs (or functions). These programs produce a transfer function that models the underlying relations within a data set. The processes that they model can be very complex, subtle, and nonlinear.

In a standard GP, a population of candidate solutions, initially generated at random, undergoes a simulated evolution process. With each generation, a new population is calculated from the previous one by allowing good solutions to replace bad ones and by generating new solutions through operators that mimic natural mechanisms such as reproduction and mutation. Consequently, good solutions progressively spread within the population, while being continually exploited to build possibly better solutions. The measure of goodness for a solution is incorporated into a numerical value called fitness that is assigned to every solution. The solutions to a data modeling problem are represented by models expressed as computerlike programs. These are internally coded as strings of numbers, each of which represents an operator, a variable, or a constant.

The simulation of evolution is realized through genetic operators. The most significant of these operators is called crossover, and it simulates sexual reproduction. In crossover, two solutions from the current population (parents) are selected with probabilities proportional to their fitness; consequently, fitter solutions receive larger chances of being selected. Next, the two solutions selected in this manner exchange pieces of information according to a simple crossover scheme, just like two chromosomes that exchange genes. The result consists of two new solutions (children) that incorporate building blocks of the old ones. Crossover is successively applied to selected pairs until an entire new generation is created. It has been demonstrated that increasingly better solutions are obtained while the evolution progresses, analogous to a population subjected to natural evolution, where the stronger individuals are the ones that pass their genes to the next generation. Eventually, the process is stopped after a predefined number of generations or when a solution with the desired accuracy has been obtained.

Another widely used genetic operator is mutation, which simulates a natural genetic accident and replaces pieces of solutions with randomly generated alternatives. Mutation sometimes acts as a fine-tuning mechanism; other times, mutation provides the necessary boost that helps the search to escape from a local minimum.

The particularities of GP methods make them most suitable for data mining and modeling problems where one is interested in discovering the relations hidden behind numerical data. In data modeling, one seeks to develop analytical models that best fit a set of data. Consequently, the fitness of a model can be any measure of the cumulated error between the model predictions and the actual values from the data set. The evolution process in this case can be referred to as *training*. The advantages of GP methods in modeling are predicted on their ability to accommodate highly nonlinear or subtle behaviors.

The physical mechanism of a GP method consists of a few standard steps:

- *Step 1:* Map the solution space of the problem at hand into a genetic space where each possible solution has a correspondent in a string of numbers called an individual, or entity. In our case, the solutions, or entities, are simple functions that model a data set.
- *Step 2:* Define a population of entities and randomly initialize it.
- *Step 3:* Evaluate the fitness of each individual in the population. The fitness describes the goodness of the solution coded into the individual. In our case, this fitness is a measure of the correlation between the actual and the predicted values. The goal of the evolutionary process is to maximize this correlation.
- *Step 4:* With probabilities proportional to their fitness, pairs of entities are selected for recombination. An operator called crossover that is the algorithmic equivalent of sexual reproduction realizes the recombination. Crossover produces children that compose the entities of the next generation. By selecting parents according to their fitness, it is assured that better solutions propagate their characteristics into the next generations.
- *Step 5:* Apply mutation and other operators. These operators enhance the performance and robustness of the process.
- *Step 6:* If the overall stopping conditions are not met, repeat steps 3 to 6 for the next generation.

In this study, we have employed a highly configurable data-modeling engine based on a hybrid algorithm that combines the robustness of GP with the efficiency of linear regression. The models generated by this engine are internally converted into closed-form, analytical expressions. A comprehensive list of operators is available, including arithmetic and trigonometric functions, min and max, and if-then-else structures. A set of genetic operators such as

crossover and mutation can be selectively applied to instructions or coefficients. Incremental learning can be accommodated by means of model libraries that store models previously developed. Evolving models can access these libraries and call the stored models as compound operators.

3 Feature Selection

An important condition for the success of the model generation process is the quality of the selected features. For our problem, these features must be clearly relevant for the detection process in a wide variety of case conditions. Essentially, they have to capture factors with demonstrated effect on human vision, such as contrast between a target and its local background, internal target structure, area and shape of the target, spectral composition, number and density of confusing forms within the scene, etc. A long series of clutter metrics have been reported in the past two decades,[3–5,14] and some of these features can be readily employed by the modeling system proposed in this paper. The long list of clutter metrics that can be found in the literature demonstrates the difficulty of finding ideal quantitative measures for clutter and target detection performance.

It is our conjecture that another category of features may be obtained through image transforms that exploit and therefore can quantify descriptive elements such as roughness, self-similarity, repetitiveness, or correlation between neighboring pixels. For example, the Gabor transform achieved a large popularity after several studies showed evidence about the relation between its localization mechanism and the mammalian early vision system.[17,18] Several candidate transforms for feature generation are currently under investigation. Among these, the wavelet and Gabor transforms,[19] the Markov random-field models,[20] the gray-level cooccurrence matrices,[5] and the fractal processes[21] have produced promising preliminary results.

Two categories of features can be identified: global, which are calculated over an entire scene and produce a single number for that scene, and local, which are calculated at the target level and describe the degree to which a specific target can be discriminated against its local background. For a specific scene, the two categories of features will carry different weights, depending on the range of the targets in that scene. Small targets, for example, will not affect the global features to an observable extent, due to their limited influence on the statistics at the scene level. In our study, we have implemented features from both categories.

3.1 *Global Features*

The list of global features implemented in our adaptive clutter modeling system is presented below.

The *average target area* (ATA) is the average of the areas of the targets encountered in the respective scene, expressed in pixels squared.

The *RMS clutter*[3] (RMS), due to Schmieder and Weathersby, is a classical clutter evaluation metric. The image is first partitioned into N adjacent windows whose areas are roughly twice as large as those of the expected targets, and then the individual intensity variances σ_i are computed for each window:

$$\mathrm{RMS}=\left(\frac{1}{N}\sum_i \sigma_i^2\right)^{1/2}. \tag{1}$$

The *clutter invariant*[14] (INV) measures the relative difference between the standard deviation σ of the entire scene and the RMS clutter. It has been claimed that this normalized difference is entirely due to intrinsic structure in the image and is background-specific:

$$\mathrm{INV}=\frac{\sigma-\mathrm{RMS}}{\sigma}. \tag{2}$$

The *probability of edge*[4] (POE) is the standard deviation of the probabilities of edge POE_i calculated over N adjacent partition windows:

$$\mathrm{POE}=\left(\frac{1}{N}\mathrm{POE}_i^2\right)^{1/2}. \tag{3}$$

POE_i is the probability of a pixel being an edge pixel. The image is first edge-enhanced and thresholded to produce a binary edge map. The recommended edge detector for this operation is the DOOG[18] filter (difference of offset Gaussians).

The *Kolmogorov-Smirnov test* (KST) measures the difference between the actual edge pixel distribution on a scene and a uniform distribution. It has been argued that a uniform distribution of the edge pixels corresponds to the largest amount of clutter in the scene. First, an edge map of the image is generated and partitioned into adjacent windows of dimensions $M\times M$. For each window, the L-level intensity histogram $\{N_i\}_{i=0}^{L-1}$ is generated first, and then an empirical cumulative distribution function (CDF) $S_N(i)$ is calculated according to

$$S_N(i)=\frac{\Sigma_{j=0}^{i}N_j}{M^2}. \tag{4}$$

The value of the KST feature is given by the sum of the absolute differences between the samples of the CDF and a ramp, which is the CDF of a uniform distribution:

$$\mathrm{KST}=\frac{1}{L}\sum_{i=0}^{L-1}\left|S_N(i)-\frac{i}{L}\right|. \tag{5}$$

The global target prominence metrics are global, yet target-dependent, metrics that measure the degree to which a generic target can be discriminated from local background within a given scene. Different prominence measures can be used; in this implementation, we have selected the contrast- and the edge-content-based measures. The algorithms for calculating the global target prominence features are discussed next.

The *global target prominence in contrast* (GTC) uses a double-window filter to transform the image and is calculated with respect to a certain target present in the scene. For that target, a target region and a background region are created; the target region is the minimum-sized rectangle that encompasses the target, whereas the background region is concentric with the target region and has an area twice as

large and the same aspect ratio as the latter. A so-called realization value is calculated next by subtracting the average intensity of the background region from that of the target region and dividing the result by the standard deviation of the background. Then, a transformed image is obtained as follows: for each pixel in the image, two concentric windows equal in size to the target region and the background region, respectively, are constructed. The difference in average intensity between the target window and the background window is calculated and divided by the standard deviation of the background window. On the normalized histogram of the transformed image, the realization value is marked. Finally, the GTC is defined as the area under the histogram of the transformed image for values less than the realization value. The GTC can be referred to as a target-dependent signal-to-noise ratio. The difference between the target mean and background mean is a measure of the target signal. On the other hand, the standard deviation of the background can be considered a measure of the noise around the target.

The *global target prominence in edge content* (GTE) is similar in principle to the GTC and attempts to evaluate the target prominence based on its edge content. First, a Sobel[22] edge map of the image is derived and transformed by a smoothing operator. The dimensions of the Sobel kernel are equal to those of the target region. A realization value is calculated for the position where the window coincides with the actual target. The area under the histogram for values less than the realization value gives the GTE. This metric quantifies the relative edge information contained in the target region as compared to that of the entire image.

3.2 *Local Features*

The local features employed during this study are calculated over the two target-related regions defined in the previous section, namely, the target region and the background region.

The *target area* (ARE), *aspect area* (ASR), *variance* (VAR), and *entropy* (ENT) are self-explanatory and describe the general appearance of the target in terms of size, shape, and internal structure. The remaining features are presented in Eqs. (6)–(9) and are called the *target-background contrast* (TBC), *target interference ratio* (TIR), *target-background interference ratio* (TBI), and *new target-background interference* (TIN).[23] These features describe the target in relation to its local background and take into account their relative differences in average intensity. It is worth emphasizing that, since the targets are reduced to rectangular regions, errors may occur in estimating the local features due to bleeding between target and background characteristics. The mathematical expressions for the features discussed above are

$$\mathrm{TBC}=|\mu_T-\mu_B|, \tag{6}$$

$$\mathrm{TIR}=\frac{\mu_T-\mu_B}{\sigma_B}, \tag{7}$$

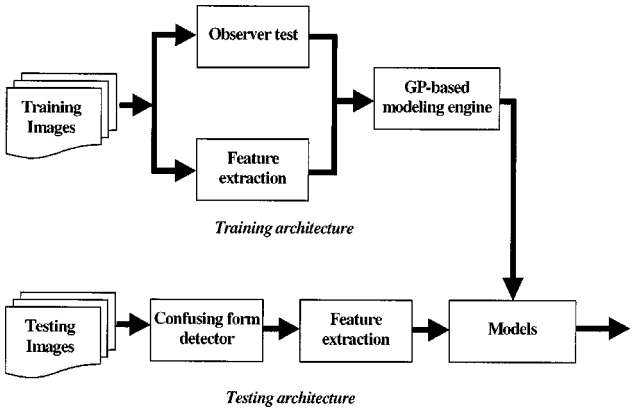


Fig. 1 Block diagram of the clutter modeling system.

$$\mathrm{TBI}=\frac{\mu_T-\mu_B}{(\sigma_B\sigma_T)^{1/2}}, \tag{8}$$

$$\mathrm{TIN}=\frac{\mu_T-\mu_B}{\sigma_B+\sigma_T}, \tag{9}$$

where $\mu_{T,B}$ are the average intensities of the target and background regions, respectively, and $\sigma_{T,B}$ are the standard deviations of these regions.

3.3 *System Architecture*

The dynamic modeling system has the architecture shown in Fig. 1 in two typical instances, namely, training and testing. In training, it was assumed that ground-truth data that describe the target positions and dimensions are available. In testing, a supplementary module called the confusing form (CF) detector has been included to identify potential targets in the scene. Its primary goal is not target detection; instead, it is supposed to detect all the objects in the scene that may be considered targets by a human observer. A variety of algorithms and procedures can be employed in the CF module, such as the hit-and-miss transform[19,24] or similar morphological operators, blob detectors, methods based on edge detection and contour aggregation, or combinations of these.

Essentially, two processes are concurrently conducted in the training stage. The first process is an observer performance test in which a population of subjects is asked to detect the targets in a set of infrared scenes. Based on this test, each image (in the global modeling approach) or target (in the local modeling approach) receives a difficulty measure. A meaningful difficulty measure can be the detection probability for each target, as measured by the test (in global modeling, the difficulty measure for a scene will be the average of the individual target measures). In parallel, a feature extraction process examines the imagery and calculates the set of features that are believed to influence the performance of the subjects. In our experiments, we have used the features discussed above. The outcomes of the observer performance tests constitute the objective data used in training the genetic engine, while the features are treated as explanatory variables or model parameters. The

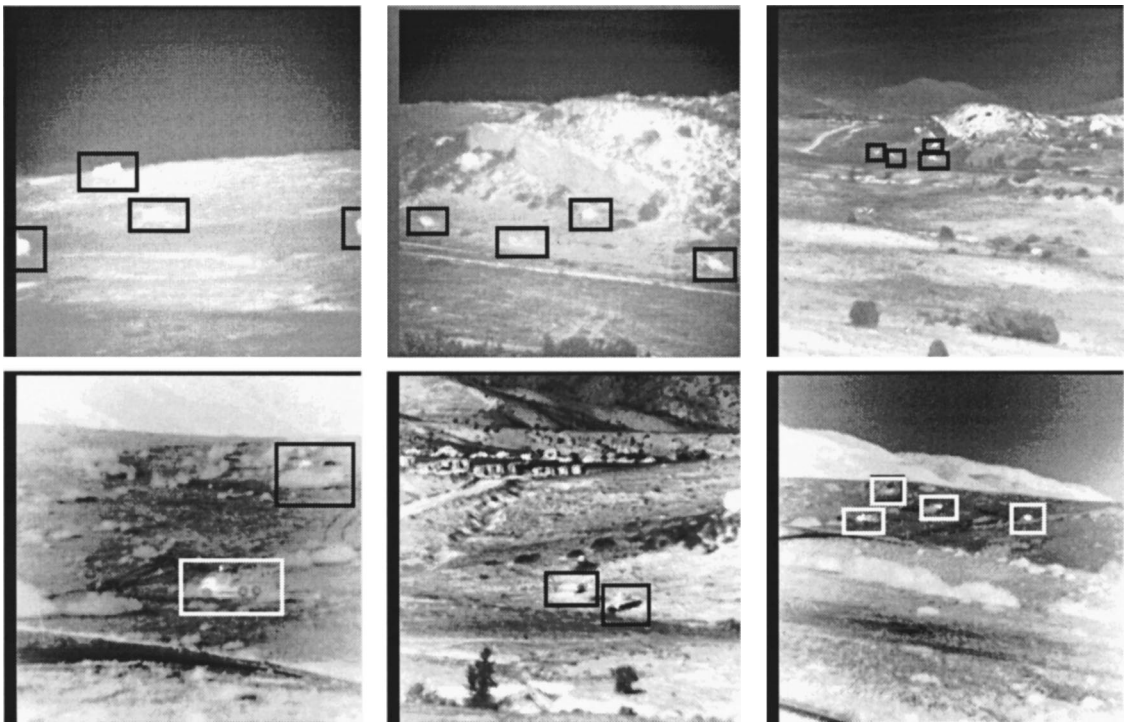


Fig. 2 Examples of IR scenes and targets from the RSTA September 94 data selected for training.

goal of the system is to detect the relations between these features and the observer performance measures.

These relations are then captured in analytical models used in the testing stage. The same feature extractor is applied to the testing images. The positions and dimensions of the potential targets will be indicated by the CF detector if ground-truth data are not available. The models generated in the training stage will use the newly derived features to produce predictions of detection performance and, equivalently, assess the amount of clutter in the testing images.

4 Observer Test

In order to collect objective data for the training set employed in the necessary modeling experiments, we organized an observer test with 33 students enrolled in two perception-related classes offered by the Psychology Department of the Rollins College, Winter Park, Florida. These students had not had military training and were not familiar with infrared imagery, and therefore additional emphasis had to be put on the training aspect of the test. The average duration of the test was about 15 min with a maximum stare time of 15 s per image. The information collected during the test included the locations of shots, the times between shots, the amount of time spent for each image, and subject-related information such as age, gender, training, and experience. An observer test administrator assistant called VASE (Visual Acuity Sensor Evaluator) and developed by Frontier Technology was employed during the Rollins test. This software package incorporates all the necessary tools to design and conduct perception tests and to manage and interpret their results.

4.1 *Imagery*

Sixty-five 256×254 images from the RSTA September 94 database were employed in the Rollins test. Fifteen of these images were included in the training set, and the remaining fifty were used for testing. These images represent infrared scenes in a rocky/mountain landscape and contain between one and seven targets per image and various amounts of clutter consisting of rocks and occasional bushes and trees. The clutter ranges from mild to severe. A total of 132 targets of five different types could be found in the 50 testing images. These types are HMMWV, M60, M35, M543A2, and M113A2. The ground-truth data available for this imagery offer extensive information about the targets (locations, ranges, types, and percentages of occlusion), as well as sensor specifications, weather conditions, etc.

Figure 2 shows a few of the training images, selected to span a variety of case conditions, target positions, sizes, and relative contrast levels.

4.2 *Subject-Related Statistics*

After discarding the results of one of the subjects who averaged about 22 shots per image and evidently did not comply with the instructions, the Rollins test produced the subject-based statistics illustrated in Figs. 3(a) to 3(d). Figure 3(a) shows the distribution of the number of shots per subject. The average number of shots per subject is 126, which means approximately 2.5 shots per image. The distribution of the percentage of targets detected per subject is represented in Fig. 3(b). The maximum of this distribution can be observed around the 56% mark, slightly below the average performance of 60% attained by our subjects. The distribution of the probability of detection, i.e., the percent-

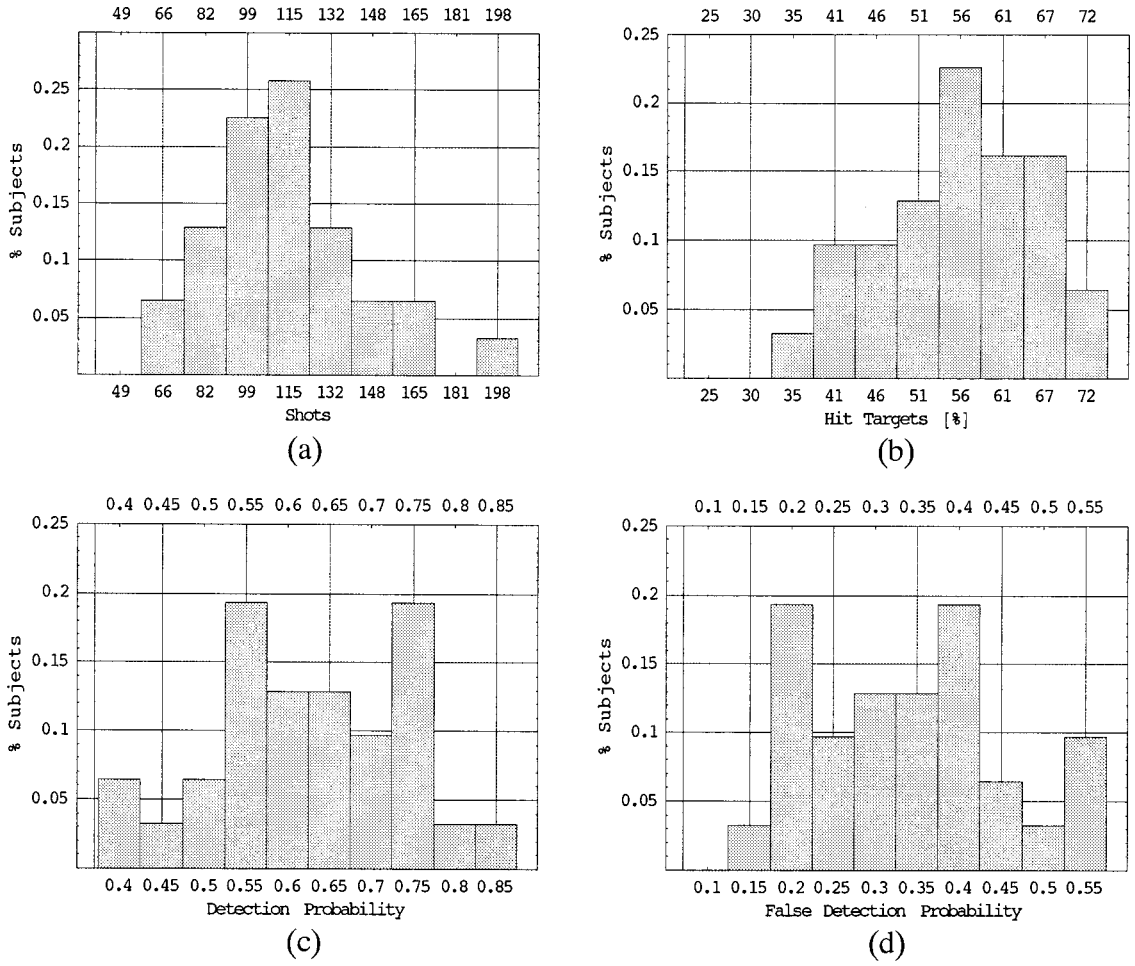


Fig. 3 Summary of the subject-related statistics: (a) distribution of shots (clicks), (b) distribution of the percentage of hit targets, (c) distribution of the probability of correct detection, and (d) distribution of the false-detection probability.

age of shots that were on target, is depicted in Fig. 3(c). Finally, the last plot, depicted in Fig. 3(d), illustrates the distribution of the false-detection probability per subject, defined as the ratio between the number of his false alarms and the total number of shots.

The performance of our subjects is considered satisfactory, given that they had no experience and training with infrared images. Although one may argue that this deficiency will affect the objectives of the test, we believe that the consistency in education, background, and age that characterized the subject population is a strong compensating factor. The models based on these results will predict how a 19-year-old (instead of a trained analyst) can detect targets in cluttered IR images.

4.3 *Target-Related Statistics*

Figure 4 illustrates the target-related distribution of the probability of being hit and provides a visual assessment of target detection difficulty in the imagery employed in the Rollins test. It is noteworthy that, while a single target was discovered by all of the participants in the test, there were six targets [shown in Figs. 5(a) to 5(f)] that were not detected at all.

One may speculate at length about the characteristics of the targets illustrated in Fig. 5 that made them virtually invisible to all of our subjects. In Fig. 5(a) the target is small and has very low contrast. Low contrast may also be the reason for missing the targets shown in (b) and (f). In

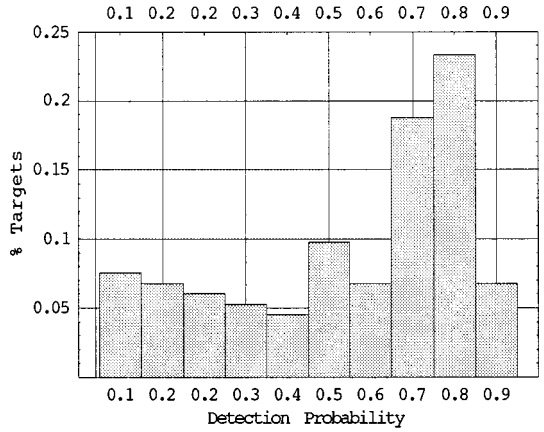


Fig. 4 The distribution of the probability of being detected, calculated over the entire set of 132 targets present in the testing images.

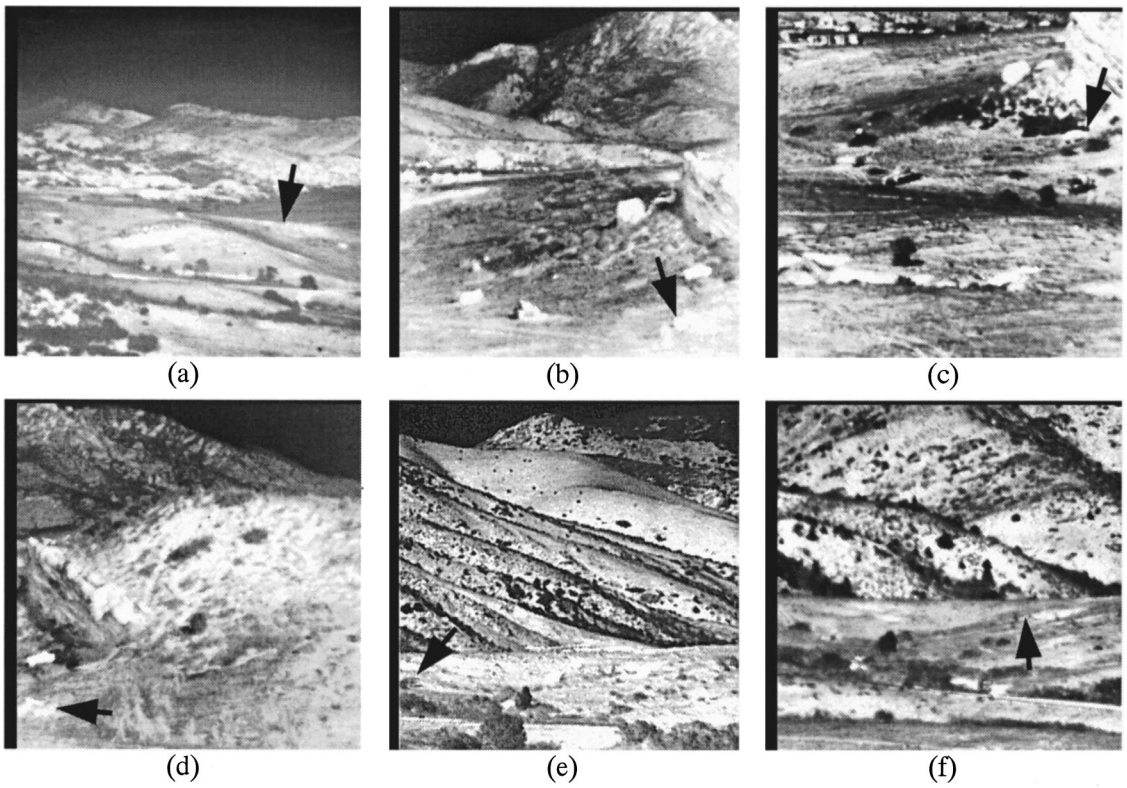


Fig. 5 Difficult detection cases. The six targets above (marked by arrows) did not receive any shots during the Rollins test.

many cases [(b), (c), (d), (e)], the targets that caused problems are situated close to the image borders, and it is well known that the best detection performance is attained on targets situated close to the center of the scene. The target in (c) has good contrast and size but is located in a very cluttered region of the image.

For comparison, we present in Fig. 6 a series of targets that looked very conspicuous to our subjects. Besides the target shown in Fig. 6(a) and recognized by 100% of the participants, Figs. 6(b) to 6(f) depict additional targets that were hit by a large majority of subjects, between 94% (missed by two subjects) and 97% (missed by a single subject).

It is somewhat surprising that even small targets, like the one at the center of image (e), can be easily detectable if they present good contrast and a fairly regular shape. The other targets from Fig. 6 are relatively large, have good contrast, show conspicuous internal structure, and, with two exceptions in (b) and (d), are unique within their scenes.

4.4 *Clutter-Related Statistics*

A few clutter objects consistently selected as targets during the Rollins test are illustrated in Fig. 7. While most of the confusions are justified by the shapes and sizes of the objects, some can be explained by our subjects' lack of experience and ability to interpret visual cues such as range [in (b) and (c)]. The clutter object displayed in (f) represents a tree that is unlikely to be confused with a target by a trained observer.

It is interesting to note, on the other hand, that the images shown in (d) and (e) represent the same scene at different moments in time, although the two images have different contrast and brightness properties and distinct spectral characteristics. The clutter object marked in these images received a significant number of shots in both scenes, which indicates that its selection by the subjects is not accidental.

5 Experimental Results

The results collected during the Rollins test were employed to generate local and global clutter models. The global models incorporate global features, and their goal is to predict the overall difficulty of scenes with respect to target detection. On the other hand, the local clutter models are based on local features and produce the probability that a specific object in a scene will be classified as a target by a human observer.

In the local models, the average probability of detection for each target was used as a performance measure. In the global models, an overall measure for the entire scene was generated by averaging out the individual detection probabilities of the targets from that scene. For computation purposes, all the features have been normalized to their maximum observed values. Although our modeling tool does not require this normalization, we have observed that the evolution process runs faster and produces more accurate models when all of the features range between 0 and 1. This also helps avoid computational artifacts such as, for example, the overflow that happens when an exponential operator is applied to a large number. An alternative and safer normalization process can be performed by dividing the feature instances by their absolute maxima, if these are

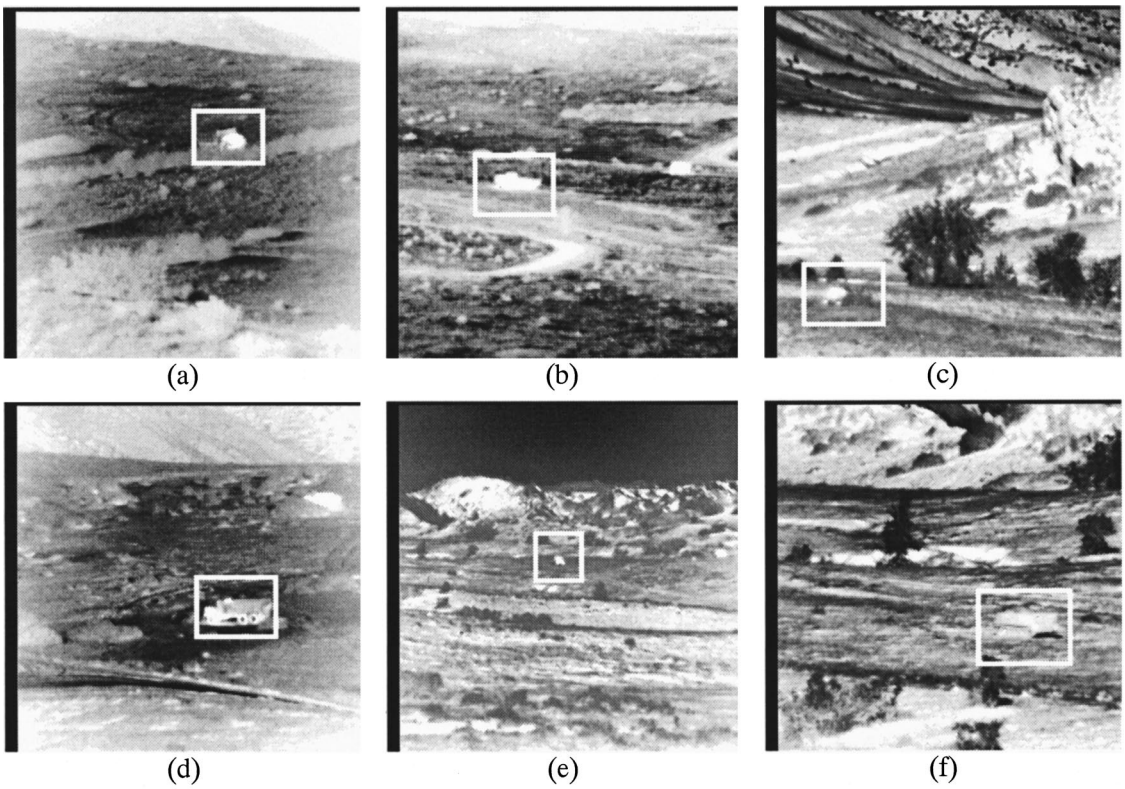


Fig. 6 Simple detection cases. The targets illustrated above were detected by (a) 100%, (b) 97%, (c) 97%, (d) 97%, (e) 94%, and (f) 94% of the subjects.

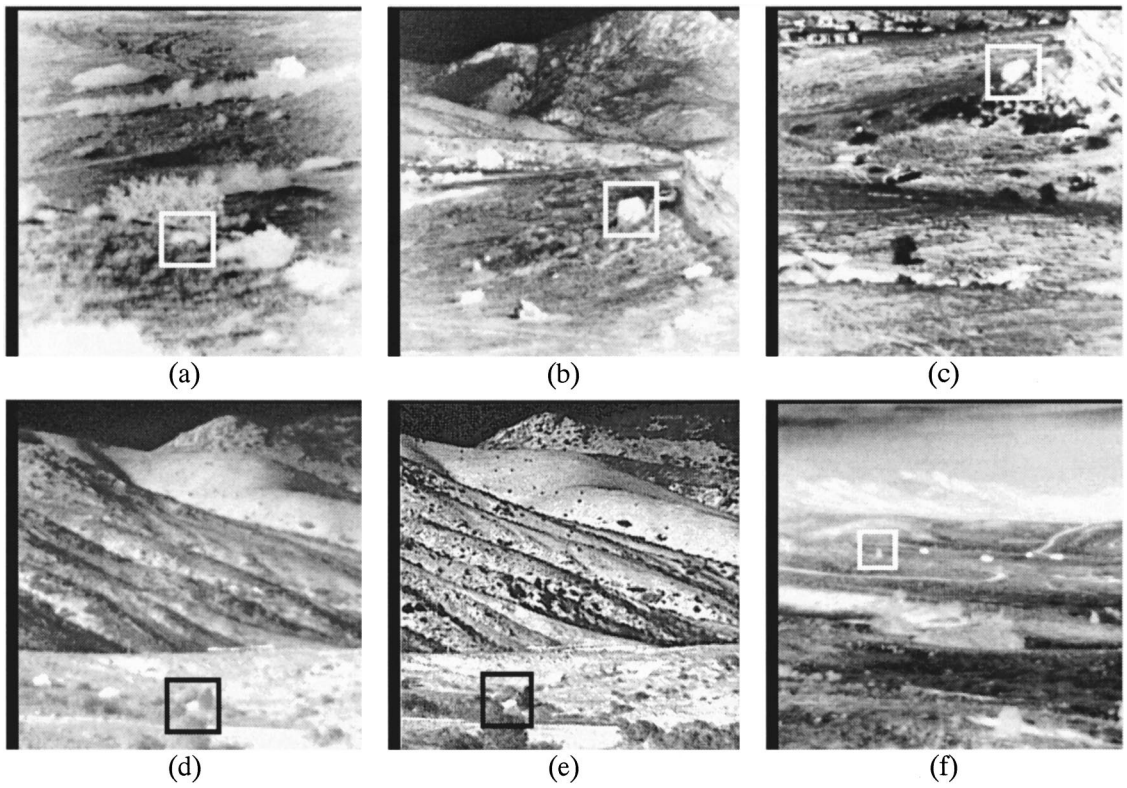


Fig. 7 Clutter objects with a large contribution to the false-alarm statistics. The encircled clutter objects were selected by (a) 72%, (b) 50%, (c) 47%, (d) 56%, (e) 100%, and (f) 37.5% of the subjects.

Table 1 Performance criteria for the global clutter models.

Model	Number of terms	χ^2	R^2	R^2_{adj}
Linear regression	7	0.69274	0.67059	0.61435
G_0	10	0.48557	0.76910	0.71583
G_1	10	0.48607	0.76887	0.71553
G_2	20	0.37310	0.82259	0.70635
G_3	18	0.31470	0.85035	0.75231
G_4	20	0.30667	0.85417	0.75863
G_5	24	0.24032	0.88572	0.78059
G_6	28	0.21723	0.89669	0.76390

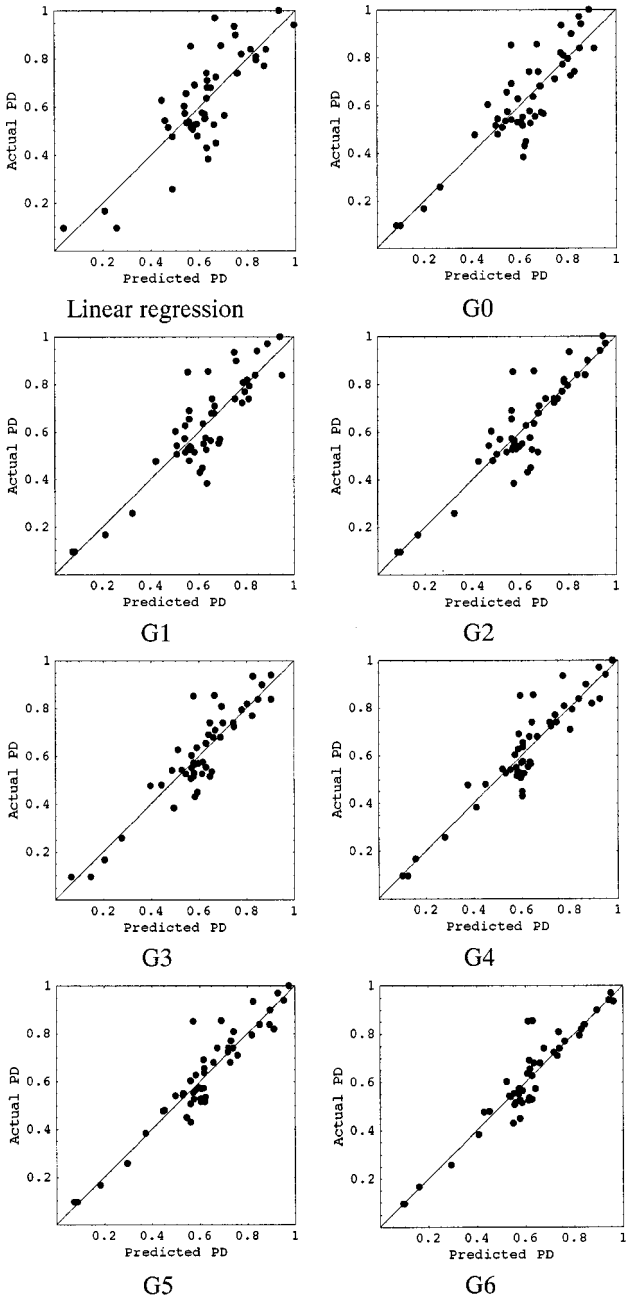


Fig. 8 Validation plots for seven of the models generated for global clutter. For comparison, the linear regression model is also included.

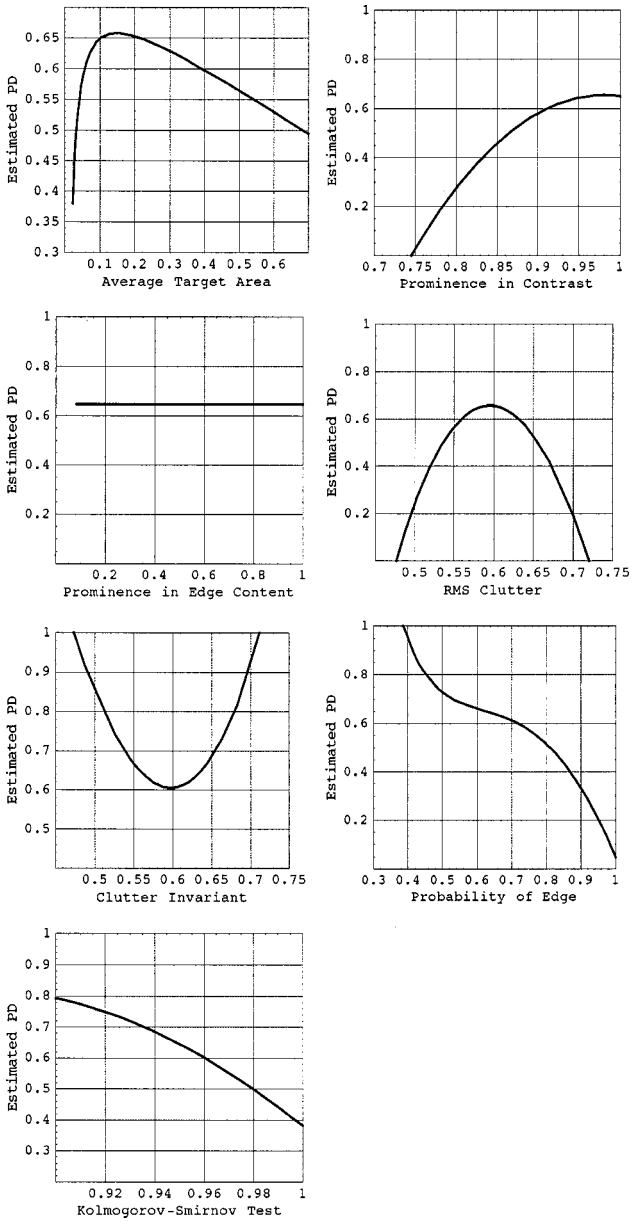


Fig. 9 General influence of global features on the detection performance.

available. The possibility of having to predict instances that exceed the maxima observed in training is thereby avoided.

Several experiments have been conducted under distinct genetic settings. The maximum allowed number of terms in the models has also been verified. For every model, three performance criteria well known in the statistics community have been recorded, namely, χ^2 (the sum of residuals squared), R^2 (the coefficient of determination), and R^2_{adj} (the adjusted coefficient of determination). Since it penalizes models with a large number of terms, the last criterion is preferred by some researchers and has been selected as the fitness value in our experiments.

5.1 *Global Clutter Models*

A series of global models was generated under the specifications described above. The models presented in this sec-

Table 2 Expressions for the selected global clutter models.

Model	Expression
Linear regression	$9.76399+0.285727\ \mathrm{ATA}+1.07117\ \mathrm{GTC}+0.223051\ \mathrm{GTP}$ $-6.47041\ \mathrm{INV}-1.7147\ \mathrm{KST}-0.679379\ \mathrm{POE}-7.73056\ \mathrm{RMS}$
G_1	$-37.6271+78.2385\ \mathrm{GTC}-23.971\ \mathrm{GTC}^2+0.73807\ \mathrm{INV}^2$ $+0.15297\ (\mathrm{ATA}+\mathrm{INV})-8.68121\ \mathrm{INV\ KST}-5.3\ (\mathrm{GTC}-\mathrm{POE})$ $-6.83856\ \mathrm{GTC\ POE}-8.93719\ \mathrm{GTC\ RMS}-16.1695\ \log \mathrm{GTC}$
G_3	$35.4946e^{\mathrm{INV}}-25.8371e^{\mathrm{KST}}-22.7465e^{\mathrm{RMS}}+82.7759(\mathrm{ATA}-\mathrm{GTC})$ $+133.5\sqrt{\mathrm{GTC}}+50.0688\ \mathrm{GTC}-\frac{27.5094\ \mathrm{GTC}}{\mathrm{KST}}-19.1543\ \mathrm{ATA\ KST}$ $-\frac{2.27767\ \mathrm{KST}}{\mathrm{INV}}+14.3713\ \mathrm{KST}^3-73.6726(\mathrm{INV}-\mathrm{POE})+\frac{1.95375\ \mathrm{KST}}{\mathrm{POE}}$ $-\frac{0.01305\ \mathrm{POE}}{\mathrm{ATA}}+10.0294\ \mathrm{INV\ POE}-7.94026\ \mathrm{POE}^2-69.965(\mathrm{ATA}+\mathrm{POE})$ $+\frac{9.05898\ \mathrm{RMS}}{\mathrm{GTC}}+\frac{1.92621\ \mathrm{RMS}}{\mathrm{INV}}+0.28166\ \mathrm{RMS}^3+16.6671\ \log\ \mathrm{RMS}$
G_5	$2.0272\sqrt{\mathrm{ATA}}+0.59598\ \mathrm{ATA}^2+39.7252e^{\mathrm{INV}}+4.30419e^{\mathrm{POE}}$ $+\frac{49.0303}{\mathrm{GTC}}+160.632\sqrt{\mathrm{GTC}}-2.71025\ \mathrm{ATA\ GTC}-74.1211(\mathrm{ATA}+\mathrm{GTC})$ $-140.405\ \mathrm{INV}^2+84.2337\ \mathrm{INV}^3-388.665\sqrt{\mathrm{KST}}-23.4616\ \mathrm{ATA\ KST}$ $-\frac{44.3663\ \mathrm{KST}}{\mathrm{GTC}}+62.3976\ \mathrm{KST}^3+96.1657(\mathrm{ATA}+\mathrm{KST})-\frac{0.71396}{\mathrm{POE}}$ $+\frac{5.93756\ \mathrm{KST}}{\mathrm{POE}}+126.17\sqrt{\mathrm{POE}}-\frac{0.015\ \mathrm{POE}}{\mathrm{GTP}}-73.9839(\mathrm{KST}+\mathrm{POE})$ $+167.351\sqrt{\mathrm{RMS}}+90.9979\ \mathrm{KST\ RMS}-327.94\ \mathrm{RMS}^2+176.056\ \mathrm{RMS}^3$

tion are labeled G_i, $i=0,\ldots,6$, and their performance criteria are illustrated in Table 1. For comparison, the performance of the linear regression model derived from the global data is also included. The plots in Fig. 8 show the correlation between the model predictions and the actual values. Ideally, the points in these plots should be collinear along the line $y=x$, which is also represented.

The generated models can now be used to investigate the relation between the background clutter and each of the features considered during this experiment. Seven global features have been employed in our study, and a graphical representation of the dependence between the outputs of our models and these variables is not feasible. A simple possibility is to draw projections of the seven-variable function on each of the feature planes. This can be accomplished by setting six of these features to predefined values and varying the remaining one within a feasible interval. As predefined values, we have chosen (although the appropriateness of this choice is arguable) the average feature values over the entire set of samples.

Figure 9 presents these projections for model G_6, which is our best performer in terms of χ^2. The seven features have shown consistent behaviors in other models as well, although these behaviors did not always follow the ones dictated by our intuition. The target area (ATA) displays a significant maximum, which suggests that there may exist an optimal size for the targets in order to be detected. The performance slowly decreases above this threshold. This behavior is possibly caused by a number of fairly large targets that were missed, some of which are shown in Figs. 5(b), 5(c), and 5(e). The dependence on the contrast prominence (GTC) is increasing, as expected. On the other hand, model G_6 does not depend on the target prominence in edge content (GTP). The RMS clutter (RMS) shows a prominent maximum in midrange. Intuitively, this is consistent with our expectations. Indeed, it is reasonable to assume that in order to detect the targets, one needs a fair amount of intensity variation in the scene. For scenes that are very uniform or exhibit a lot of intensity variation, the detection task is likely to be a difficult one. It is important to note that the RMS has been found to be a very stable and robust clutter metric throughout this experiment. The profile of the clutter invariant metric (INV) is complementary to that of the RMS clutter, as anticipated by Eq. (2). Also consistent with our expectations is the monotonically decreasing behavior of the probability of edge (POE). Finally, the Kolmogorov-Smirnov test (KST) behaved in disagreement with our theoretical predictions, which affirm that the probability of detection should increase with this metric. These inconsistencies may result in part from the data in-

Table 3 Performance criteria for the local clutter models.

Model	Number of terms	χ^2	R^2	R^2_{adj}
Linear regression	8	7.36779	0.39437	0.35498
L_0	9	5.14557	0.57704	0.54952
L_1	10	4.76134	0.60862	0.57975
L_2	20	3.78980	0.68847	0.635628
L_3	19	3.75963	0.69090	0.64173
L_4	28	3.13270	0.72765	0.65695
L_5	30	3.70570	0.72294	0.64417
L_6	39	3.05191	0.74913	0.64663

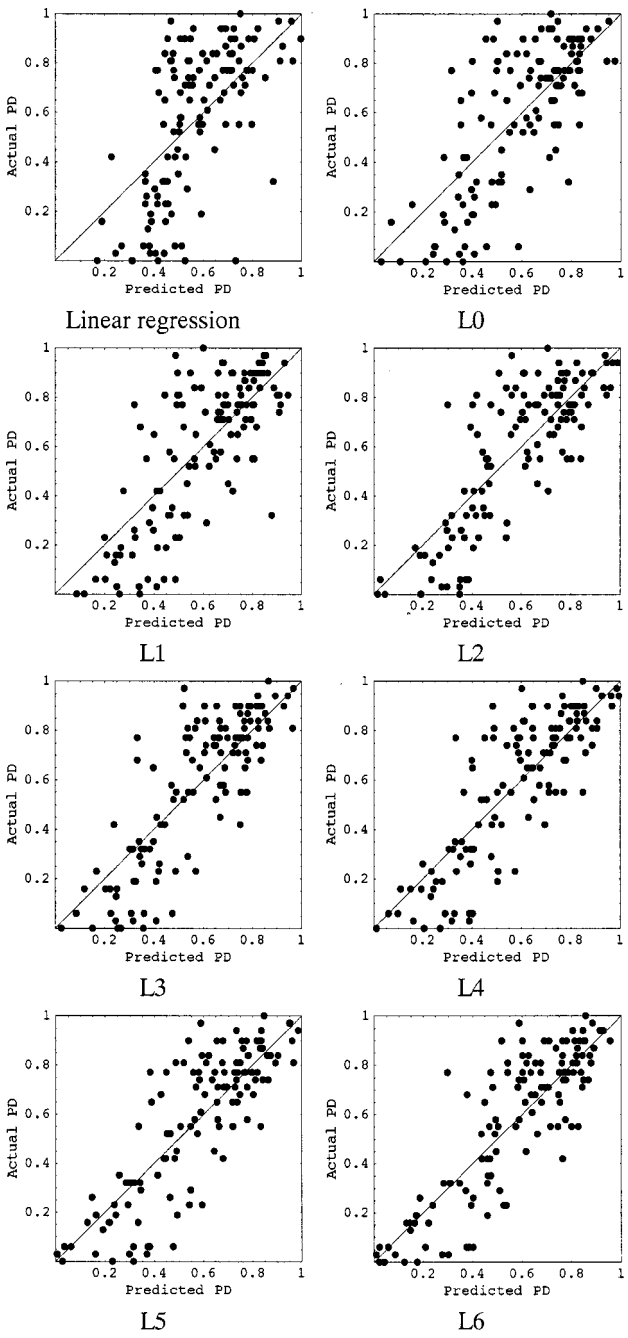


Fig. 10 Validation plots for seven of the models generated for local clutter. For comparison, the linear regression model is also included.

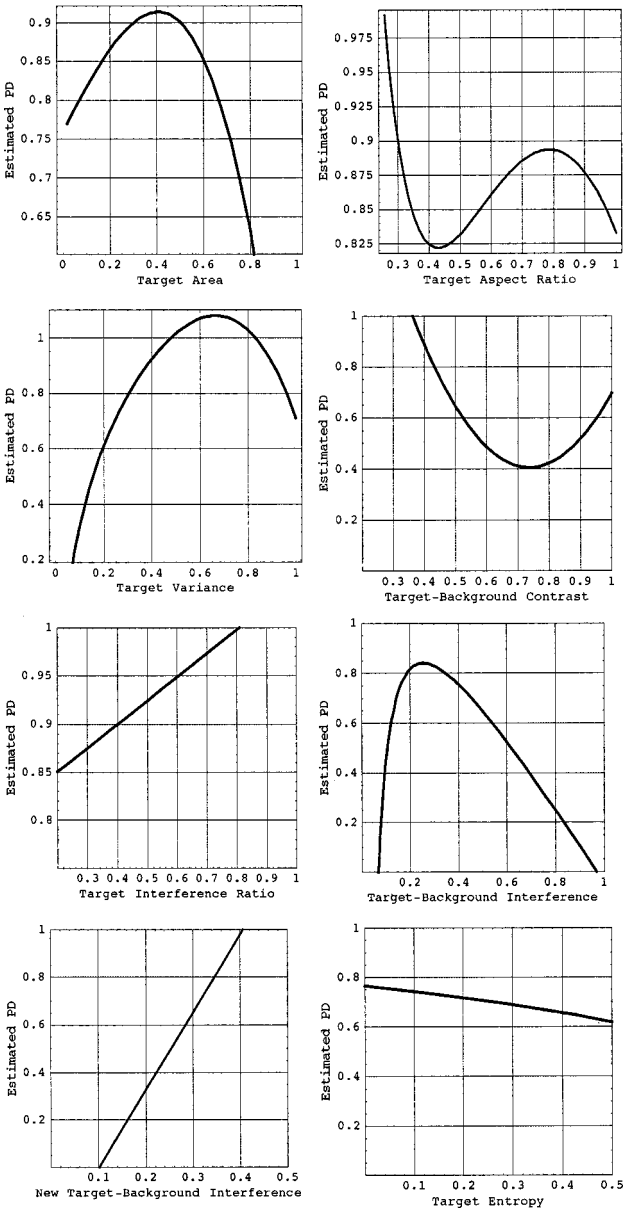


Fig. 11 General influence of local features on the detection performance.

sufficiency in the test. In addition, there is undoubtedly a great deal of cross-correlation between our features, and accordingly it is conceivable that some of the combinations of features used in drawing the projections are not physically possible.

For illustrative purposes, the expressions of a subset of the global clutter models generated throughout this study are presented in Table 2.

5.2 *Local Clutter Models*

In the local clutter modeling experiment we have followed the same guidelines as in global modeling. This subsection presents seven selected models built during this experiment and labeled L_i, $i=0,\ldots,6$. The same model performance criteria as in the global modeling case have been monitored (Table 3). Figure 10 contains the correlation plots (predicted versus actual) for a selection of local clutter models, and Fig. 11 shows the individual influence of the features on the output of model L_4.

As a general conclusion, the robustness of the local models seems to be affected by the approximation of the targets with rectangular boxes. The target area (ARE) exhibited similar behavior to that in global modeling, with a maximum in midrange. The aspect ratio (ASR) has a bimodal profile, and we suspect that its relevance for target detection is minor. Besides, for small targets the aspect ratio is quite an unstable feature. The target variance (VAR) metric shows a prominent maximum about the 0.6 mark,

Table 4 Expressions for the selected local clutter models.

Model	Expression
Linear regression	$0.30491+0.146614\ \mathrm{ARE}-0.0335958\ \mathrm{ASR}-0.0516514\ \mathrm{ENT}$ $-0.543521\ \mathrm{TBC}-2.73696\ \mathrm{TBI}+2.40095\ \mathrm{TIN}+1.55645\ \mathrm{TIR}+0.17753\ \mathrm{VAR}$
L_0	$-0.10102\ \mathrm{ENT}+1.34486\sqrt{\mathrm{TBC}}-\frac{0.20812\ \mathrm{TBI}}{\mathrm{TIR}}$ $+\frac{0.31917\ \mathrm{TIR}}{\mathrm{TBI}}-1.47918(\mathrm{TBC}+\mathrm{TIR})-0.33075(\mathrm{TBI}+\mathrm{TIR})$ $+1.2755(\mathrm{TIN}+\mathrm{TIR})+0.9821\ \mathrm{VAR}-1.12454\ \mathrm{VAR}^3$
L_1	$-4.49967\ \mathrm{ARE}^3+\frac{1.76127}{\mathrm{ASR}}+3.67527\ \mathrm{ASR}^3+5.82934e^{\mathrm{ARE}}-6.68816e^{\mathrm{ASR}}$ $-0.30404e^{\mathrm{ENT}}+4.98517e^{\mathrm{TBC}}-1.16551(\mathrm{ARE}+\mathrm{ENT})-5.53485\ \mathrm{TBC}^2$ $-4.7782(\mathrm{ARE}+\mathrm{TBC})+1.23373(\mathrm{ENT}+\mathrm{TBC})+\frac{0.0005\ \mathrm{ENT}}{\mathrm{TIN}}$ $+1.0382\ \mathrm{TIN}-\frac{0.02294\ \mathrm{TBC}}{\mathrm{TIR}}-\frac{0.1874\ \mathrm{TBI}}{\mathrm{TIR}}+1.50349\ \mathrm{ARE}\ \mathrm{TIR}$ $-0.84614\ \mathrm{ASR}\ \mathrm{VAR}-1.90775\ \mathrm{VAR}^2+2.5069(\mathrm{ASR}+\mathrm{VAR})+6.72968\ \log\ \mathrm{ASR}$
L_4	$-6.24693-0.74531\mathrm{ARE}-1.13651\ \mathrm{ARE}^3+\frac{0.2825}{\mathrm{ASR}}+\frac{0.53831\ \mathrm{ARE}}{\mathrm{ASR}}$ $+3.6108\ \mathrm{ASR}-1.92572\ \mathrm{ASR}^2-0.25391e^{\mathrm{ENT}}+7.85078e^{\mathrm{TIN}}$ $-4.29496e^{\mathrm{VAR}}+0.17687\ \mathrm{ARE}\ \mathrm{ENT}+0.28074\sqrt{\mathrm{TBC}}+4.28414\ \mathrm{TBC}^2$ $+\frac{0.00102\ \mathrm{ASR}}{\mathrm{TBI}}-\frac{0.2513\ \mathrm{TBC}}{\mathrm{TBI}}+2.19599\ \mathrm{TBC}\ \mathrm{TBI}-2.51213(\mathrm{TBC}+\mathrm{TBI})$ $-9.50362\ \mathrm{TBC}\ \mathrm{TIN}-5.61252\ \mathrm{TIN}^2+0.45453\ \mathrm{TIN}^3-0.01244(-\mathrm{ENT}+\mathrm{TIN})$ $-0.15024(\mathrm{TBC}+\mathrm{TIN})-1.62557\ \mathrm{ARE}\ \mathrm{TIR}+4.5423\sqrt{\mathrm{VAR}}$ $+1.92152\ \mathrm{ARE}\ \mathrm{VAR}+1.46482\ \mathrm{TIR}\ \mathrm{VAR}+4.95654\ \mathrm{VAR}^2-1.16427\ \mathrm{VAR}^3$

similar to the profile of the variance-based feature used in global modeling, namely, the RMS clutter. This suggests, once again, that a human needs a moderate degree of variation, inside and outside the target, in order to perform a target detection task in infrared images. The target-background contrast (TBC) did not behave as expected and shows a minimum in midrange. The positive influences of the target interference ratio (TIR) and the new target-background interference ratio (TIN) on the detection process are in agreement with our predictions based on Eqs. (7) and (9). Indeed, the greater the difference between the average intensities of the target and background regions, respectively, the better the chances for the target to be detected. This argument fails, though, in the case of the target-background interference (TBI), whose behavior is considered an inconsistency. Finally, the entropy of the target (ENT) shows a mild negative contribution to the detection process. Alternative local features must be investigated and validated against testing data.

The expressions for three models selected from the set discussed in this section are given in Table 4.

6 Conclusions and Future Research

This study describes a new approach to clutter modeling, referred to as dynamic modeling assisted by genetic programming. The purpose of this study was to open a new direction of research in this area rather than solve the clutter quantification problem. We believe that dynamic modeling can become a powerful instrument in various domains that require the systematic analysis of large amounts of data.

The models generated throughout this study have illustrative value. Due to data insufficiency, we could not validate them on testing data. Furthermore, the overall significance of the Rollins test is questionable, due to the small number of participants and images employed, and the inappropriate background and training of these participants. Nonetheless, it has been shown that data modeling based on genetic programming achieves better prediction accuracy than regression modeling. Besides generating clutter models, this approach can also be used to identify the features that have the greatest effect on detection performance. It is noteworthy that similar features behave consistently with one another in both global and local modeling. A very good example is the similarity between the RMS clutter (global) and the target variance (local). An additional advantage of the dynamic modeling is that it can essentially predict both human observer and seeker performance.

Further research is necessary in infrared clutter modeling in order to determine alternative features that are in closer relation with the detection performance indicators, capture unexplored aspects of targets that are inherently used by humans to discriminate them from the background, and take advantage of contextual cues.

An important aspect of data modeling is the statistical significance of the models generated. Our experience has shown that a conflict arises at some point in time between the accuracy and the prediction capabilities of our models. In theory, a model generated by genetic programming can achieve infinite accuracy if a proper coding method has been chosen and a long enough evolution time has been allowed. However, this is oftentimes accomplished at the expense of the generalization and prediction properties of the resulting model. Consequently, one must not aim at obtaining extremely high accuracy, but rather at obtaining meaningful models. A robust method to estimate the statis-

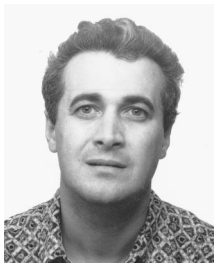
tical relevance and the prediction capabilities of these models as early as in the training stage is yet to be established.

Acknowledgments

This work has been funded under the US Army Aviation and Missile Command (AMCOM) Contract No. DAAH01-98-C-R097. We greatly appreciate the support and advice received from Lisa Cannon and Ron Passmore. The authors wish to thank Drs. Marni Bekkedal and Bonnie Walker from the Rollins College in Winter Park, Florida, for their contribution to the organization of the observer performance test. We acknowledge the support received from George Lukes and Raymond Rimey in the acquisition of the RSTA September 94 image data used in this study.

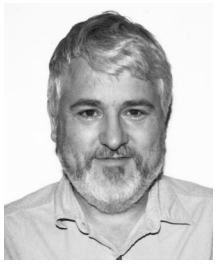
References

1. T. Morawski, C. Drury, and M. Karwan, ''Predicting search performance for multiple targets,'' *Hum. Factors* **22**(6), 707–718 (1980).
2. A. Akerman III and R. E. Kinzly, ''Predicting aircraft detectability,'' *Hum. Factors* **21**(3), 277–291 (1979).
3. D. E. Schmieder and M. R. Weathersby, ''Detection performance in clutter with variable resolution,'' *IEEE Trans. Aerosp. Electron. Syst.* **AES-19**(4), 622–630 (1983).
4. S. Rotman, G. Tidhar, and M. Kowalczyk, ''Clutter metrics for target detection systems,'' *IEEE Trans. Aerosp. Electron. Syst.* **30**(1), 81–90 (1994).
5. G. Waldman, J. Wootton, G. Hobson, and K. Luetkemeyer, ''A normalized clutter measure for images,'' *Comput. Vis. Graph. Image Process.* **42**, 137–156 (1988).
6. G. Waldman, J. Wootton, and G. Hobson, ''Visual detection with search: an empirical model,'' *IEEE Trans. Syst. Man Cybern.* **21**(3), 596–606 (1991).
7. T. Arani, M. H. Karwin, and C. G. Drury, ''A variable memory model for visual search,'' *Hum. Factors* **26**(6), 631–639 (1984).
8. T. Caelli and G. Moraglia, ''On the detection of signals embedded in natural scenes,'' *Percept. Psychophys.* **39**(2), 87–95 (1986).
9. H. R. Blackwell, ''Contrast thresholds of the human eye,'' *J. Opt. Soc. Am.* **36**, 624–643 (1946).
10. J. D'Agostino, W. Lawson, and D. Wilson, ''Concepts for search and detection model improvements,'' *Proc. SPIE* **3063**, 14–22 (1997).
11. R. Kistner, W. Pibil, T. Meitzler, D. Bryk, E. Sohn, and D. Bednarz, ''AMSAA/TARDEC moving target perception experiment and analysis,'' in *Proc. Eighth Annual Ground Target Modeling and Validation Conf.*, pp. 201–205 (1997).
12. J. S. Sanders, M. S. Currin, and C. E. Halford, ''Visual perception of infrared imagery,'' *Opt. Eng.* **30**(11), 1674–1681 (1991).
13. J. Johnson, ''Analysis of image forming systems,'' in *Image Intensifier Symp.*, pp. 249–273, U. S. Army Eng. Res. and Development Lab., Fort Belvoir, VA (1958).
14. W. R. Reynolds, ''Toward quantifying infrared clutter,'' *Proc. SPIE* **1311**, 232–240 (1990).
15. N. Ben-Yosef, K. Wilner, S. Simhony, and G. Feigin, ''Measurement and analysis of 2-D infrared natural background,'' *Appl. Opt.*, 2109–2113 (1985).
16. J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, Boston (1992).
17. S. Marçelja, ''Mathematical description of the responses of simple cortical cells,'' *J. Opt. Soc. Am.* **70**(11), 1297–1300 (1980).
18. R. A. Young, ''Simulation of human retinal function with the Gaussian derivative model,'' in *Proc. IEEE Conf. Comput. Vision Patt. Recogn.*, pp. 564–569 (1986).
19. D. P. Casasent, J.-S. Smokelin, A. Ye, and R. Schaefer, ''Correlation filter fusion for detection: morphological, wavelet, and Gabor methods,'' *Proc. SPIE* **2026**, 6–16 (1993).
20. R. Chellappa, ''Two-dimensional discrete Markov random field models for image processing,'' in *Progress in Pattern Recognition 2*, L. N. Kanal and A. Rosenfeld, Eds., pp. 79–112, Elsevier, North-Holland (1985).
21. A. P. Pentland, ''Fractal-based description of natural scenes,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **6**(6), 661–674 (1984).
22. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, MA (1992).
23. R. A. Peters II and R. N. Strickland, ''Ån image complexity metric for automatic target recognizers,'' presented at the ATR Systems Technology Conf., Oct. 1990, Naval Surface Warfare Center.
24. D. Casasent, R. Schaefer, and R. Sturgill, ''Optical hit-miss morphological transform,'' *Appl. Opt.* **31**(29), 6255–6263 (1992).



Liviu I. Voicu is a scientist with Frontier Technology, Orlando, Florida. He received his PhD in electrical engineering from the University of Central Florida in 1997 and his engineering diploma degree from the University Politehnica, Bucharest, Romania, in 1988. His main research focus is in data mining with emphasis on data modeling, bio-mimetic algorithms, and fractal imaging.

Mosleh Uddin received his BSEE from Bangladesh University of Engineering and Technology and MSEE from University of Central Florida in 1994 and 1998, respectively. Currently he is working for Frontier Technology, Inc., as a member of the technical staff and studying for his PhD at University of Central Florida. His research interests include digital signal processing and image processing.



Harley R. Myler is a professor of electrical engineering in the School of Electrical Engineering and Computer Science at the University of Central Florida in Orlando. He received his BSEE from the Virginia Military Institute in 1975 and the MSEE and PhD from New Mexico State University in 1982 and 1985, respectively. His main research focus is in machine intelligence with emphasis on image understanding and intelligent control. He has written four books, thirty refereed journal articles and fifty conference papers. He is a senior member of the IEEE, a member of SPIE and a Tau Beta Pi Eminent Engineer.

Anthony Gallagher graduated in 1998 from the University of Central Florida, Orlando, Florida, with a BS in computer engineering. He is currently pursuing a MS in electrical engineering at the University of Central Florida. His expertise is in image processing, machine vision and pattern recognition, genetic algorithms, and data mining.

Julien Schuler is a student in the Electrical Engineering Department at INSA Lyon, France. In 1999 he was awarded an internship with I-Math Associates, Orlando, Florida, where he worked in image processing and infrared clutter modeling. Julien Schuler is currently attending an engineering program in Sevilla, Spain, where he is working in automatic path recognition and 3D laser sensors. His research interest is focused on image processing, path recognition, and genetic algorithms.