CAR-TR-693 F49620-93-1-0576
CS-TR-3174 November 1993

PROBE BASED RECOGNITION OF TARGETS IN INFRARED IMAGES

Sandor Z. Der

Night Vision and Electronic Sensors Directorate
AMSEL-RD-NV-VISP-LET
Fort Belvoir, VA 22060

Rama Chellappa

Department of Electrical Engineering
Computer Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742-3275

Abstract

A probe based approach is used to recognize objects in a cluttered background using an infrared imager. A probe is a simple mathematical function which operates locally on image grey levels and produces an output that is more directly usable by an algorithm. A directional probe image is calculated by taking the difference in grey levels between pixels a set distance apart in a given direction, centered on the probe image pixel. These probe images contain the information necessary for use by an object recognition algorithm in a readily usable, and mathematically describable, form. A parametric statistical image background model which describes the probe images is introduced. The parameters of the probe image model can be readily estimated from the image. Knowledge of these parameters, together with target signatures obtained from Computer Aided Design (CAD) models, allows the likelihood ratio for a given object pose hypothesis versus the background null hypothesis to be written. The generalized likelihood ratio test is used to accept one of the object poses or to choose the null hypothesis. Results of the method applied to a large set of terrain model board images are presented.

Keywords: Object recognition, automatic target recognition, systems and applications, pattern analysis.

The support of Rama Chellappa by the Advanced Research Projects Agency (ARPA Order No. A369) and the Air Force Office of Scientific Research under Grant F49620-93-1-0576 is gratefully acknowledged, as is the help of Sailaja Akunuri and Sandy German in preparing this paper.

1 Introduction

A number of methods [1, 2, 3, 4, 8, 10] have been used to identify objects and discriminate regions in infrared imagery. A relatively simple method is that of Bhanu and Holben [3], who use a relaxation scheme to automatically set a threshold to separate "hot" and "cold" regions of an image. This requires that the objects of interest be quite prominent. Some methods use the physics of objects in the image. Aggarwal and Nandhakumar [1] use infrared emission information together with visible reflectivity information to separate regions based on thermal properties.

A paradigm often used for the extraction of objects from their backgrounds is the detection/segmentation/feature extraction/recognition paradigm (e.g. [10]). It is particularly attractive in situations for which reduced computational complexity is critical, because the stages successively reduce the complexity of the problem to manageable proportions. The chief disadvantage is that the stages introduce errors which cannot always be rectified by later stages. Frequently such algorithms are particularly sensitive to imperfect segmentations. Bhanu gives a survey of such algorithms in [2]. Such algorithms often do not take direct advantage of knowledge of the geometry of objects, relying instead on derivative features such as moments. For applications in which segmentation is particularly difficult, this can be ineffective as many commonly used features require that an intact silhouette be present.

In an attempt to reduce the dependence of the algorithm on the segmentation scheme, a number of authors (e.g. [4, 8]) have used the known geometry of objects of interest to guide the separation of object from background, an approach sometimes called model-based vision. The approach combines the segmentation/feature extraction/recognition stages into one matching process in the hope of improving performance. It is used in [4], in which an energy minimization scheme is used to match templates to image regions. Contour smoothness, edge sharpness, and object shape are all taken into account during optimization, though their relative weights change as the optimization proceeds. One advantage of such approaches is that performance degrades gracefully as the quality of the silhouette degrades.

An alternative to simply choosing an intuitively appealing matching function is to develop an image model and object model that allow the application of a standard test of goodness of fit, preferably one that can be proven optimal (e.g. [5, 9]). This is the general approach taken in this paper. It differs from Margalit, Reed, and Gagliardi [9] in that a more complex image model is required, and from Grimson [5], in that it does not rely on an apriori edge finder which might function poorly for low contrast objects. Grimson obtains empirical background probability distributions for ensembles of edges, and uses them to estimate the number of edge correspondences required to ensure a low false alarm rate. The method requires a set of images containing "false" objects similar to the images to be tested, in order to estimate the probability of false correspondences. The method was developed for the identification of parts in a manufacturing environment. The method developed here is more general, in that the estimated probability distributions describe probe values, which are a more primitive construct than edge correspondences. One advantage of this is that probability distributions can be estimated using only the image with which the algorithm is being tested.

In this paper, image and target models that allow the calculation of the probability of an observed image section are developed. These models are used to define an algorithm which applies the generalized likelihood ratio test to test for the presence of the desired objects. Such an approach has the advantage that the designer knows where to look to improve the algorithm; since the generalized likelihood ratio is well accepted, although not optimal, the image model must be improved if performance is to be improved.

The image model we have used requires that a specific type of probe be defined, where the definition is chosen to correspond with the authors' intuition of what probe function will best differentiate objects of interest from their backgrounds. The probability density function of the probe, or the parameters of this function, is then estimated from the image. The pattern of probes used to search for a given object corresponds to the silhouette of the object, as determined from a CAD model. The probability density function of probe values corresponding to targets cannot be reasonably estimated from the image, because the locations of targets are not assumed known, and because targets occur infrequently, which would cause sample sizes to be small.

Explicit temperature predictions for the scene might be used to predict grey levels, but this would require data that we cannot assume to be known. Accurate prediction of outdoor temperatures requires knowledge of recent thermal history, including whether the object in question is under direct sunlight or precipitation, and the magnitudes of internal heat sources. It is shown in the body of the paper that the variation in irradiation between an object receiving direct sunlight and a shaded object can dwarf variations due to material properties. While knowledge of such information might be realistic in controlled manufacturing environments, it is not realistic for noncooperative targets.

Algorithm designers may take a number of steps to decrease the computational complexity of an algorithm, including prescreening to eliminate large areas of the image from consideration, and decision trees to quickly sort through the set of candidate hypotheses. If these measures are applied appropriately, they should introduce relatively small amounts of error, as compared to the fundamental limitations of the matching scheme itself. The matching scheme should be the heart of the algorithm. Unfortunately, matching schemes tend to be based on heuristic measures that reflect the intuition of the algorithm designer. While intuition can yield effective algorithms, the inherent difficulty in communicating intuitive insights, and the absence of a standardized data set for comparison purposes, makes it difficult to evaluate the relative merits of different algorithms.

Developers of object recognition algorithms have suffered from a lack of images on which to design and test their algorithms. In particular, image sets have often consisted of too few images for adequately understanding of the performance limitations of the algorithm. The algorithm described in this paper has been tested on a set of approximately two thousand images acquired from a terrain model board, encompassing a wide variety of target and background conditions. A description of the image set and algorithm performance results are presented in the paper.

The paper is organized as follows. Section 2 discusses the rationale for not making explicit use of the apparent temperature information given by the infrared imager, but rather treating apparent temperature values as grey levels. Section 3 introduces a simple image model, and presents experimental results to justify its use. Section 4 demonstrates how the image model described in Section 3 can be used to perform a generalized likelihood ratio test for a given set of known objects. Section 5 presents experimental results of the recognition algorithm described in Section 4. Section 6 discusses conclusions and future work.

2 Problems with the Thermal Prediction Approach

The image model approach taken here is in contrast to methods used by Nandhakumar and Aggarwal [1] in that it does not directly use knowledge about thermal interactions between background elements and objects of interest to help in their extraction. The reason that thermal prediction information is not used in our approach is that we do not wish to assume that a coregistered visible wavelength image is available. Also, the characterizations of objects required for thermal predictions to be accurate are much more detailed than we would like to assume in our scenario. Presumably such an approach would involve calculating thermal predictions for both man-made objects and background terrain. In our scenario, it would be reasonable to assume some knowledge about the background terrain, such as the general locations of fields and forests, an elevation map on the scale of tens of meters, some idea about the types of vegetation present, and recent meteorological history. It would also be reasonable to assume that exact representations of the objects of interest are available. If this were all the necessary information, thermal prediction could be a great aid to object recognition in outdoor scenes. However, the temperatures of vehicles depend greatly on their recent activity, because vehicles have quite strong internal heat sources including the engine and heater, internal friction, friction on wheels and tracks, etc. For a vehicle that has been recently operated, these heat sources dominate the vehicle's signature. Prediction of vehicle temperature thus requires knowledge of the vehicle's movement, engine status, etc. We do not wish to assume that such knowledge is available. For backgrounds the problem is also more complicated because the exact locations of tree lines and the spatial variations in the densities and types of vegetation are crucial for prediction, but are presumably unavailable. Significant signal variation can be observed even from relatively uniform sparse fields of grass. Temperature prediction is useful if these types of conditions can be controlled, as in indoor scenes, for example.

The thermal prediction approach could be less ambitious, and only predict probability density functions (PDFs) of temperature, or simply reasonable temperature ranges. The usefulness of the latter approach is somewhat questionable, as there are often significant overlaps between natural and background objects, and even when a man-made object stands out, it is not necessarily the object of interest, so that discrimination based on shape is still necessary. While the pdf's of object and background temperature would be useful, prediction would require estimating a priori probabilities such as the probability that a vehicle has moved recently, that its engine is on, that it is shaded by foliage cover, etc. For the background, we would similarly need to know the probability that the soil in the area is of a given quality, that the trees are of a given species, etc. All of the above mentioned factors are first order effects of the object and background temperature.

To demonstrate the difficulty of finding man-made objects in natural scenes based on predicted temperature, a simple thermal model was created to predict the temperatures of materials given meteorological data. The aim was to show that the effects of conditions that cannot be assumed known can be strong enough that the range of temperatures which a given material can have overlaps the temperature ranges of other materials. In particular, the surface of a steel slab on top of soil can have temperatures similar to those of natural materials such as soil, sand, rock, and wood. A more complex model would identify more sets of conditions that produce overlap; for our purposes, a simple model suffices.

The simulation involves a simple slab of material placed on top of dry soil. A finite difference model is applied to the material and the soil. Because the soil has low thermal diffusivity, conduction is assumed to occur only in the vertical direction. The surface effects considered are down-welling irradiation, direct and diffuse solar radiation, free convection, and emission from the material. Solar radiation, air temperature, and down-welling irradiation were obtained from recorded meteorological files. For simplicity, these parameters were kept constant while the finite difference model was allowed to reach steady state. The results for materials in direct sunlight, in the shade, and at night are shown in Table 1. Clearly, whether the object to be recognized is shadowed is more important than its material composition. Of course, objects with internal heat sources might easily exceed the temperature ranges shown. Vegetation was not included in the table because of the crucial role played by evapo-transpiration in the regulation of plant temperature. It can be seen that steel, because of its lower emissivity, has a significantly lower apparent temperature than the other materials. Since man-made objects, often made of steel, typically have strong internal heat sources, it can be seen that these objects have the greatest range of possible apparent temperature values.

Table 1: Predicted temperature and apparent blackbody temperature of materials.

Material	**Sun**	**Shade**	**Night**
Dry Soil	298.5	282.9	280.3
Sand	297.9	282.9	280.5
Steel	302.6	289.5	287.4
Granite	290.2	288.5	277.9
Fir	298.2	282.9	280.4

Material	**Sun**	**Shade**	**Night**
Dry Soil	294.3	278.9	276.4
Sand	290.2	275.5	273.2
Steel	207.2	198.3	196.9
Granite	282.7	281.0	270.7
Fir	292.0	276.9	274.5

The heat equation is written as

$$\left(\frac{1}{\alpha}\right)\frac{\partial T}{\partial t} = \frac{\partial q}{\partial t} + \nabla^2 T$$

where T is temperature, t is time, α is thermal diffusivity, and q is input energy. The thermal diffusivity breaks down as

$$\alpha = k/\rho c_p$$

where ρ is the material density, k is the thermal conductivity, and c^p is the specific heat of the material.

Since we are considering a large flat surface and uniform volume in which the heat flow is exclusively vertical, a one dimensional approximation can be used. The finite difference equations can be obtained by discretizing the above equation as

$$\frac{T_m^{t+1} - T_m^t}{\alpha \Delta t} = \frac{T_{m-1}^t - 2T_m^t + T_{m+1}^t}{\Delta x^2}$$

following the notation in [7], where the external heat flux term has been dropped. For internal nodes at which external heat flux is zero, the finite difference model is simply [7]

$$T_m^{t+1} = F(T_{m+1}t + T_{m-1}t) + (1 - 2F)T_m^t$$

where $T_m t$ is the temperature at the m^{th} node at time t, and the Fourier number F is

$$F \equiv \alpha \Delta t / (\Delta x)^2$$

where Δx is the vertical spacing of the nodes, and Δt is the time step used by the finite difference model. At the surface node, the equation is

$$T_0^{t+1} = 2FT_1^t + (1 - 2F)T_0^t + T_{in}$$

where T_{in} is the temperature change due to energy exchange at the exposed surface. It breaks down as

$$T_{in} = \Delta t(P_e + P_c + P_{ir})\{\Delta x \rho c_p\}^{-1}$$

$$P_e = -\varepsilon\sigma\{T_0^t\}^4$$

$$P_c = h\{T_a - T_0^t\}$$

where ε is the emissivity of the surface material, σ is the Stefan-Boltzmann constant, h is the free convection coefficient of air, T_a is the ambient air temperature, and P_{ir} is the irradiation onto the surface due to direct and diffuse solar, as well as down-welling irradiation. The meteorological parameters, including P_{ir}, were taken from recorded data files and the material constants were obtained from tables in [7].

The apparent temperature of an object is the temperature of a blackbody that would produce the same emission as the object. Thus, objects with low emissivities have lower apparent temperatures than objects at the same temperature with higher emissivities. The formula is

$$T_{ap} = \varepsilon^{1/4} T_{ac}$$

where T_{ap} is the apparent temperature and T_{ac} is the actual temperature.

3 The Probe Image Model

The image model used by Margalit, Reed, and Gagliardi [9] assumes that the pixel grey levels of an infrared image can be modeled as a Gaussian with slowly varying local mean. For their purposes it was unnecessary to model how the local mean varied, as they were concerned with small targets containing only a few pixels. For our purposes, statistical characterization of the background jumps is crucial.

Our assumption, which will be justified later, is that the difference in grey level between two pixels at a distance d apart can be modeled as a zero mean Gaussian, with variance a monotonically nondecreasing function of d. Such a random variable will be called a probe, and denoted by J. Each probe J will be correlated with neighboring probes, a fact for which our algorithm must account. The variances of the probes, and their correlations with neighboring probes, will vary from image to image, and thus will be calculated locally. The algorithm will then make use of this background distribution to search for desired objects.

Consider an image formed by calculating the value of a probe of a given distance and given direction centered at each pixel of an input image. Thus for an input image $\{x_{ij}\}$ where the first subscript denotes the vertical axis and the second the horizontal, the probe image $\{y_{ij}\}$ for probes of distance d in the horizontal direction is

$$y_{ij} = x_{i+d,j} - x_{i-d,j}$$

Then the random field $Y = \{y_{i,j}; (i,j)\epsilon\omega\}$, where ω is the pixel grid, is modeled as a field of correlated zero mean Gaussians. This correlation is measured for probe images formed in each principal direction. A whitening filter can be applied based on the correlation values, resulting in a field of uncorrelated zero mean Gaussians. Thus, locally, each probe value, after whitening, may be treated as an independent identically distributed (IID) sample of a random variable. This greatly simplifies the hypothesis test used to test for the presence of the target objects. The role of the Gaussian assumption is only to assert that the whitening filter, which causes the pixels to be uncorrelated, also causes the pixels to be independent. The recognition algorithm uses the actual local histogram as a sample probability density function rather than fitting a Gaussian to the histogram, thus reducing the algorithm's dependence on the Gaussian assumption.

The Gaussian assumption was tested by performing Kolmogorov-Smirnov tests of the conformity of sample histogram to a Gaussian distributions. The resulting average, minimum, and maximum K-S test values are recorded in Table 2.

Table 2: Kolmogorov-Smirnov test comparison of empirical probe pdfs vs. Gaussian distribution.

Direction	**Jump Size**	**Average**	**Minimum**	**Maximum**
1	1	0.09613	0.08042	0.10649
1	2	0.08048	0.05586	0.10540
1	3	0.06439	0.03850	0.10678
1	4	0.05153	0.03038	0.10680
1	5	0.04418	0.03017	0.10570
1	6	0.04037	0.03201	0.10559
1	7	0.03961	0.03391	0.10658
1	8	0.03990	0.03523	0.10602
1	9	0.04101	0.03659	0.10599
1	10	0.04223	0.03755	0.10662

Direction	**Jump Size**	**Average**	**Minimum**	**Maximum**
2	1	0.09696	0.08160	0.10536
2	2	0.08432	0.06148	0.10573
2	3	0.07203	0.04626	0.10520
2	4	0.06165	0.03534	0.10562
2	5	0.05431	0.03102	0.10582
2	6	0.04823	0.03296	0.10549
2	7	0.04502	0.03354	0.10588
2	8	0.04327	0.03367	0.10458
2	9	0.04269	0.03502	0.10537
2	10	0.04264	0.03672	0.10519

Direction	**Jump Size**	**Average**	**Minimum**	**Maximum**
3	1	0.10075	0.09855	0.10650
3	2	0.09902	0.09606	0.10688
3	3	0.09543	0.09192	0.10583
3	4	0.09167	0.08601	0.10610
3	5	0.08807	0.08128	0.10480
3	6	0.08459	0.07674	0.10508
3	7	0.08110	0.07237	0.10532
3	8	0.07787	0.06762	0.10571
3	9	0.07458	0.06271	0.10577
3	10	0.07211	0.06010	0.10524

Direction	**Jump Size**	**Average**	**Minimum**	**Maximum**
4	1	0.09554	0.07952	0.10665
4	2	0.07952	0.05508	0.10631
4	3	0.06359	0.03736	0.10525
4	4	0.05104	0.03022	0.10529
4	5	0.04412	0.03041	0.10548
4	6	0.04064	0.03195	0.10548
4	7	0.04003	0.03423	0.10711
4	8	0.04042	0.03553	0.10609
4	9	0.04155	0.03707	0.10528
4	10	0.04271	0.00000	0.10535

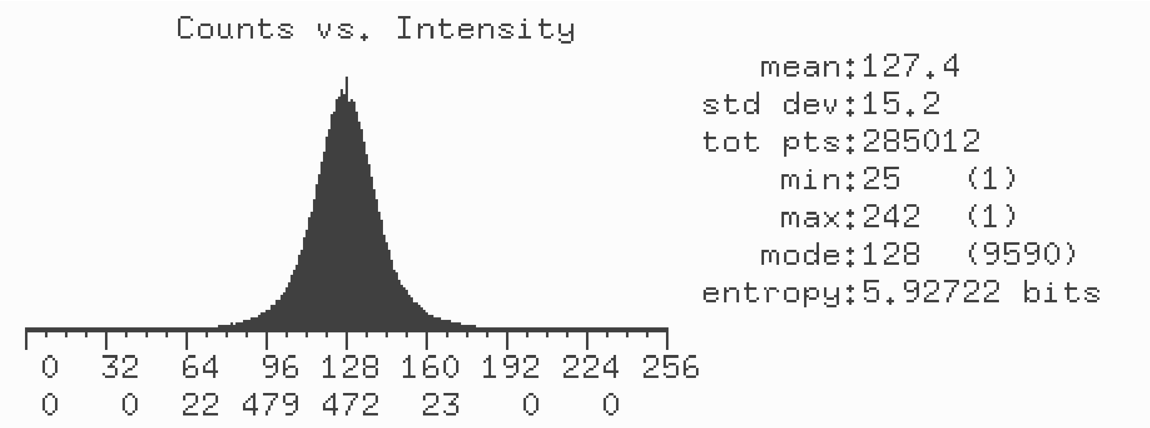


Figure 1: Histogram of probe image.

4 The Likelihood Ratio

The IID background distribution makes it easy to calculate $P(J \mid B)$, the probability of a given set J of probes given that the set is drawn from the background distribution. The image is assumed to be locally stationary. For each probe, a local histogram is computed and used to calculate the parameters of an assumed Gaussian distribution. Alternatively, the local histogram itself may be used as the pdf of the probe. Making use of the IID nature of the probes allows us to write

$$P(J \mid B) = \prod_i P(J_i)$$

where $P(J_i)$ is the pdf of the probes.

The pdfs of probes associated with a target are more problematic. As previously discussed, prediction of target signatures requires a great deal of information which we do not wish to assume known. Absence of this knowledge makes signature prediction virtually impossible. However, the geometry of the situation provides some hope. The target shape is presumed to be known exactly; in particular, its silhouette shape is known. Since a probe that straddles the edge of the target represents a difference in temperature between a metal object and natural vegetation, it is reasonable to hope that the probe will show a jump discontinuity of some significance. The strength of the jump discontinuity will vary greatly according to environmental and target conditions. For this reason, we will only use probes that straddle the silhouette of our hypothesized target, and we will assume that the magnitudes of these probes follow a uniform distribution over the discrete alphabet of possible probe values. This is essentially a worst case distribution [11].

Consider a set of probes that straddle a given target silhouette. Then the likelihood ratio may be written as

$$L(J) = \frac{k}{\prod p(j_i)}$$

where K is a constant. If a uniform cost function may be assumed for all of the potential hypotheses, then the generalized likelihood ratio test may be implemented. The test will then choose the target pose hypothesis that produces the largest $L(J)$, and will declare that hypothesis if this $L(J)$ exceeds a threshold, but declare "no target" if it does not.

A common technique used to make tests of this type more robust is the use of β Windsorized statistics [13] to prevent abnormally large probe values from distorting the results. In this technique, probe values that exceed a threshold will be treated as if they were equal to the threshold. For the case of infrared imagery, the horizon typically causes a large jump in apparent temperature. Hence, $L(J)$ for silhouette hypotheses that run along this roughly horizontal line will be quite large, even if the probes corresponding to the target side are nonexistent. Using β Windsorized statistics decreases the effect of the horizon pixels, as well as the effect of other edges.

Finding the maximum $L(J)$ is identical to finding the minimum $P(J \mid B)$. Defining τ as the value of the threshold for line processes, we have

$$P(J \mid B) = P(\tau) \parallel \{i : J_i \geq \tau\} \parallel [\prod i : ji < \tau P(J_i)]$$

or

$$-\log P(J \mid B) = - \parallel \{i : J_i \geq \tau\} \parallel \log P(\tau) - \sum_i \log P(J_i)$$

Since τ will be in the tail of the distribution, and it is assumed that for targets a large proportion of the probes will exceed the distribution, the above equation may be approximated as

$$-\log P(J \mid B) \propto - \parallel \{i : J_i \geq \tau\} \parallel$$

In other words, the test simply counts the number of probes along the silhouette that exceed a threshold. The problem becomes one of binary template matching, a subject discussed extensively in the literature (e.g. [14]).

The knowledge of the background distribution allows estimation of the false alarm rate associated with a choice of the threshold τ. Choosing a desired false alarm rate allows comparison of likelihood functions that are generated with different numbers of probes. Define $\gamma(\tau)$ as the probability that a background probe exceeds τ. Then

$$\gamma(\tau) = P[J_i > \tau] = 1 - F_i(\tau)$$

where $F_i()$ is the cumulative distribution function of J_i. If $\vartheta(\tau)$ is defined as the number of probes that must exceed τ in order for the hypothesis to be declared, then the false alarm rate α , which is defined as the probability that a background image portion will be declared to be a target, will be

$$\alpha = 1 - B(\vartheta(\tau), 1 - F_i(\tau), N)$$

Here N is the number of probes used to test the hypothesis and $B(k, p, n)$ is the cumulative distribution function of a binomial random variable, where n is the number of Bernoulli trials, p is the probability associated with each Bernoulli trial, and k is the number of trials with positive result. The independence of the probes allows each probe to be treated as a Bernoulli random variable.

It would be undesirable to simply choose the threshold τ and $\vartheta(\tau)$ a priori. Instead we chose a desired false alarm threshold α, allowed τ to vary over the alphabet of possible probe values, and for each value of τ used the $\vartheta(\tau)$ that gives the desired α . Thus for each image portion presented to the algorithm, a range of τ values was calculated. Since a number of hypotheses need to be compared in the generalized likelihood ratio test, the algorithm actually varies τ, determines the actual number of probes that exceed τ, determines the false alarm rate associated with these numbers, and chooses the τ that provides the lowest false alarm rate. If this lowest false alarm rate is the minimum among all the hypotheses, and is lower than the chosen false alarm rate threshold α, then the object is declared to be a target.

Note that the actual false alarm rate will not be equal to the α value chosen in the algorithm. Actually, α would be the false alarm rate associated with one a priori choice of τ, if there were only one silhouette hypothesis, and if each test were independent. Since the algorithm tests the same silhouette centered at neighboring pixels, it is clear that these tests are not independent. Also, the tests for different silhouettes are not independent, nor are the tests performed at different τs. This makes the calculation of the predicted false alarm rate rather complex, especially when one takes into account the changing size of the silhouettes as a function of range.

One of the parameters to be chosen is the probe distance. There are many factors that influence this choice, some of which are not obvious. These include the following:

1) The quantization of a sharp edge results in an intermediate pixel that has a value between those of the object and its background. This means that the probe should straddle at least one pixel.

2) The blur kernel of the sensor optics further reduces the sharpness of the edge. The blur kernel size depends on how the sensor is focused.

3) Extremely tight probes would make the algorithm quite sensitive to range error, as well as error in the assumption that the targets are on a level surface.

4) Since the variance of the background probes monotonically increases with probe distance, tighter probes provide stronger matches for a given edge strength.

5) As probe distances increase, the probability that a given probe straddles the silhouettes of other target hypotheses increases; thus the number of probes that differentiate between two hypotheses is reduced.

An easy solution that addresses the first four of the above concerns is to vary the probe distance for each probe to determine the distance that gives the strongest probe. This means that at a give jump discontinuity, the probe will be stronger if it is well centered on the jump and somewhat weaker if it is not, but a jump that is not perfectly located is still allowed to influence a hypothesis.

4.1 Implementation considerations

The most obvious steps that can be taken to decrease computation time are the introduction of an effective prescreener, and the use of a decision tree scheme to quickly sort through the hypotheses. In our implementation, an extremely simple contrast box prescreener was used to reduce the number of candidate pixels, but no decision tree was used. The reason was that the algorithm should have a matching function that separates hypotheses in some optimal way, and the prescreener serves only to eliminate some hypotheses at some locations. Algorithm speed was also increased by applying hypotheses at coarse resolution, and moving up to full resolution only if the coarse resolution test result warranted it. No attempt was made to choose optimal settings of thresholds at different resolutions; the purpose was only to reduce computation time.

The algorithm allows great reductions in computation time if certain calculations are performed and their results stored prior to the actual testing of hypotheses. The basic algorithm raster scans the image, and at each pixel, tests the fit of each silhouette hypothesis. To prohibit overlap of targets, the best fit hypothesis within a neighborhood is retained, and nearby hypotheses with lower likelihood ratios are eliminated. The probes used for a given hypothesis are centered at the pixels of the silhouette of that hypothesis.

We choose a set of probe thresholds τ_i such that there are an equal number of probes in the background between each value of τ_i. A great deal of computation can be saved if probe images are constructed, and then scaled so that only the subscripts ("labels") of the thresholds are stored. In other words, a probe image is constructed in which the alphabet is the set of labels, and pixel value j means that the probe at that pixel exceeds thresholds τ_i for $i \in [0, j]$. This means that the significance of the probe in a certain direction at a certain pixel is calculated only once, rather than every time a silhouette hypothesis uses that pixel. This is especially important because the conversion requires floating point calculations while lookups in the probe significance image require only integer operations. On the other hand, this method requires more memory use, since a probe significance image must be stored for each principal direction used.

The introduction of an intermediate step in the calculation of the probe significance image can save even more calculation. As described earlier, the size of the probe at each pixel in each direction is optimized to give the greatest significance. This is done by calculating probe significance for varying probe sizes, choosing the most significant, and then storing that result. The results are probe significance images in which optimization over probe size has already been performed, so that this optimization adds almost no computational cost to the algorithm.

The creation and storage of cumulative binomial distribution tables also saves a great deal of computation time. These tables are indexed by the total number of probes, the number of probes that exceed or equal the threshold label being tested, and the threshold label. Because the α values chosen are quite low, there are computer precision problems associated with these tables; namely, the value returned from a table is often zero. For this reason, an approximation is used for the cumulative binomial distribution that holds for the tail end values [12], namely

$$P\{X \geq K\} \approx b(k,n,p) \times \frac{(k+1)(1-p)}{K+1-(n+1)p}$$

where

$$b(k,n,p) = \begin{pmatrix} n \\ k \end{pmatrix} \times p^k (1-p)^{n-k}$$

The above expression allows logarithms to be taken, thus avoiding the underflow problem.

4.2 Occlusion performance

The performance of the algorithm when a target is partially occluded can be easily determined. Since the algorithm looks only at the silhouette, and all points along the silhouette are treated equally, occlusion simply reduces the number of probes that actually straddle the target. If it is assumed that the probes that are occluded do not, by chance, straddle a sharp discontinuity in the background which causes it to exceed the threshold, then occlusion eliminates a subset of the probes, requiring the remaining probes to be stronger if the object is to be declared a target. Thus the algorithm has the desirable property that it degrades gracefully as occlusion is increased; and partial occlusion can be compensated by stronger edges on the portion of the silhouette that remains visible.

Figure 2 shows the fraction of probes associated with a hypothesis that can give negative results but still allow the algorithm to declare the hypothesis, for a given number of probes and given α. Of course, it makes no difference to the algorithm whether the probes give negative results because a portion of the target is occluded or because a portion of the target silhouette has low contrast. Figure 2 shows that for a given α, as the number of probes decreases, a larger percentage of them must be positive for the algorithm to declare the hypothesis. This is due to the constant false alarm rate (CFAR) nature of the algorithm. The implication is that detection performance suffers as range increases. The imagery used to evaluate the algorithm had a 2.3×1.7 degree field of view with 640×480 pixels. Consider a target pose that is 5 meters wide by 3 meters high. The silhouette would consist of 250 probes at one kilometer, but only 50 probes at five kilometers.

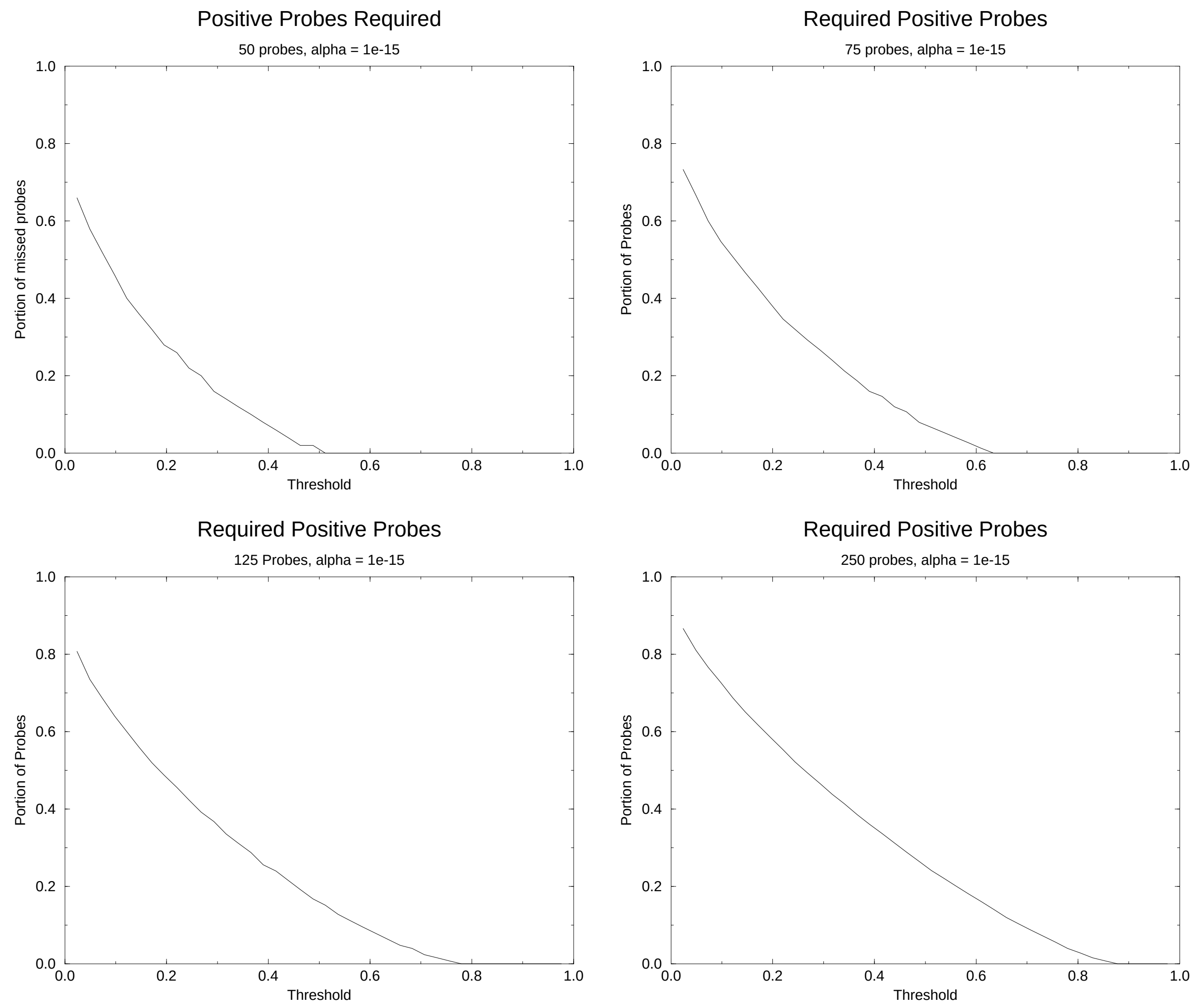


Figure 2: Fraction of positive probes required to declare a target as a function of contrast.

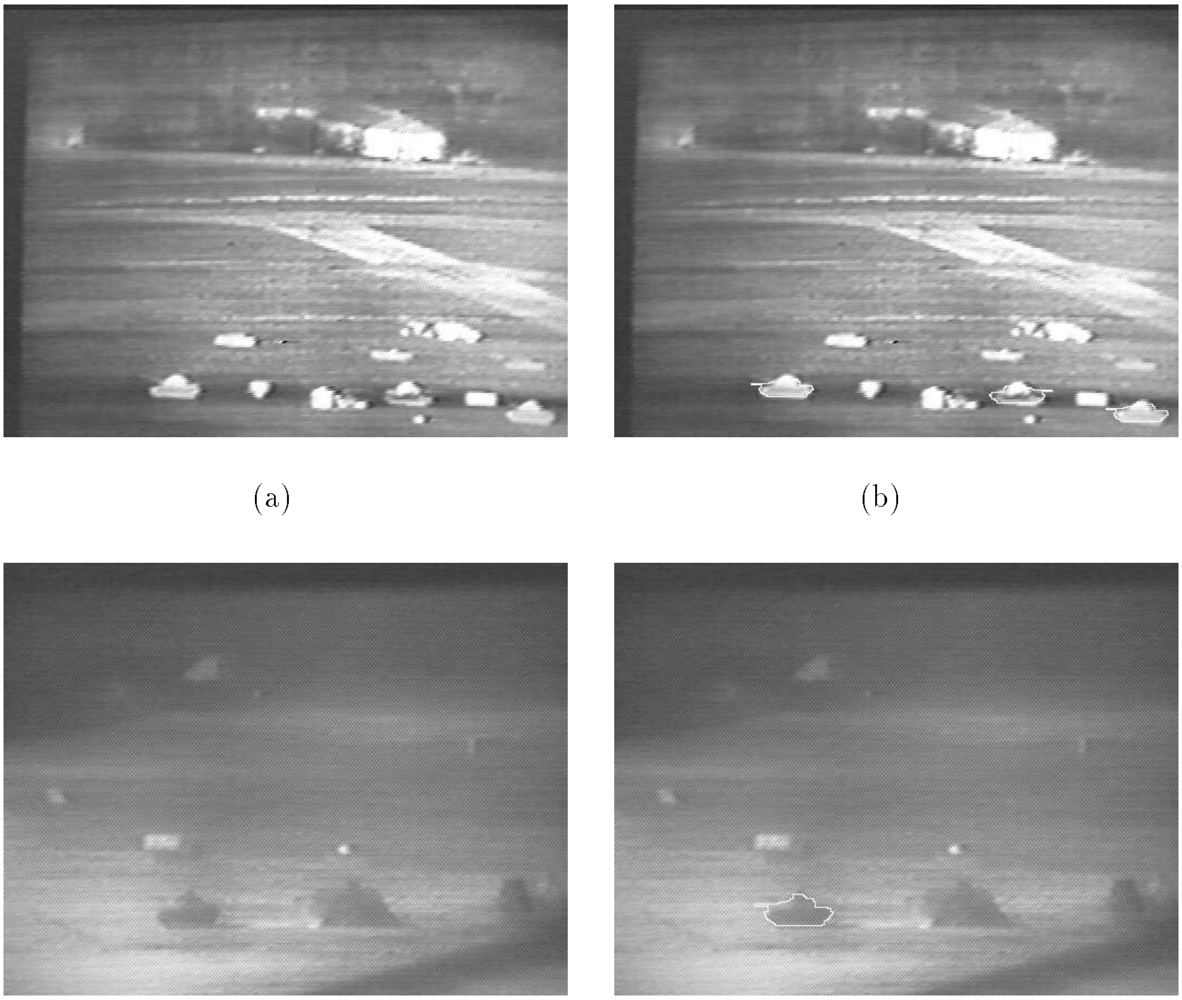
5 Experimental Results

In order to test the performance of the algorithm, it was run on six real FLIR images, as well as about 2000 images obtained from a scale model terrain board. The real FLIR images show high contrast white hot targets in an environment with relatively low clutter. Figure 3 shows two of the FLIR images. The results were scored only for those targets for which CAD models were available. Algorithm detections on vehicles for which CAD models were unavailable were ignored. The result was that for the nine target recognition opportunities, all were recognizable with an α threshold that eliminated any false alarms. Of course, if the threshold were lowered, some false detections would be reported.

The terrain board images are quite difficult in that the background contains a large number of target-like objects, and many of the targets are difficult for human observers to recognize because they have quite low contrasts. Figure 4 shows the sample terrain board images. The detection vs. false alarm curves for the algorithm are shown in Figure 5, together with confusion matrices for some values of α. In order to estimate detection performance for prominent objects, the images were rescored using only those targets that were painted with high reflectivity paints, in other words, those targets whose simulated temperatures were the highest. These results are shown in Figure 6.

Four vehicles were chosen to be objects for the image model, based on the availability of CAD models for the vehicles. The CAD models were ray traced to determine their silhouettes at increments of five degrees in azimuth. The elevation angle was assumed to be zero. These silhouettes were used as the hypothesis silhouettes in the images.

The set of terrain board images were produced in 1990 by personnel of the U.S. Army Night Vision and Electronic Sensors Directorate [6]. The simulated infrared images were obtained by capturing images of a scale model terrain board using a camera sensitive in the visible region of the spectrum, and an 8 bit digitizer. The grey levels of the image were then reversed, so that grey level i became grey level $255 - i$. This provides a reasonable simulation of the appearance of infrared imagery, as discussed below. The targets were painted so that when inverted, the target signatures correspond to those typically found in infrared imagery. The advantage of using terrain board imagery is that conditions can be controlled, allowing



(a) (b) (c) (d)

Figure 3: Recognition results on real FLIR images (a) and (c). The recognized targets are highlighted in (b) and (d). Only targets for which CAD models were available were considered for experiments.

repetition of experiments and choice of operating conditions. Background objects can be placed at will, allowing the user to increase clutter or occlusion. Terrain board imagery can also be gathered much more cheaply than real imagery. The drawback, of course, is that inverted video is not a perfect approximation of infrared imagery. In order to justify use of this imagery to test the algorithm, a test was made to see if the probe image model, which matched nicely with a set of actual FLIR imagery, also matched the inverted visible imagery captured from the terrain board. Table 2 shows the results of the Kolmogorov-Smirnov test

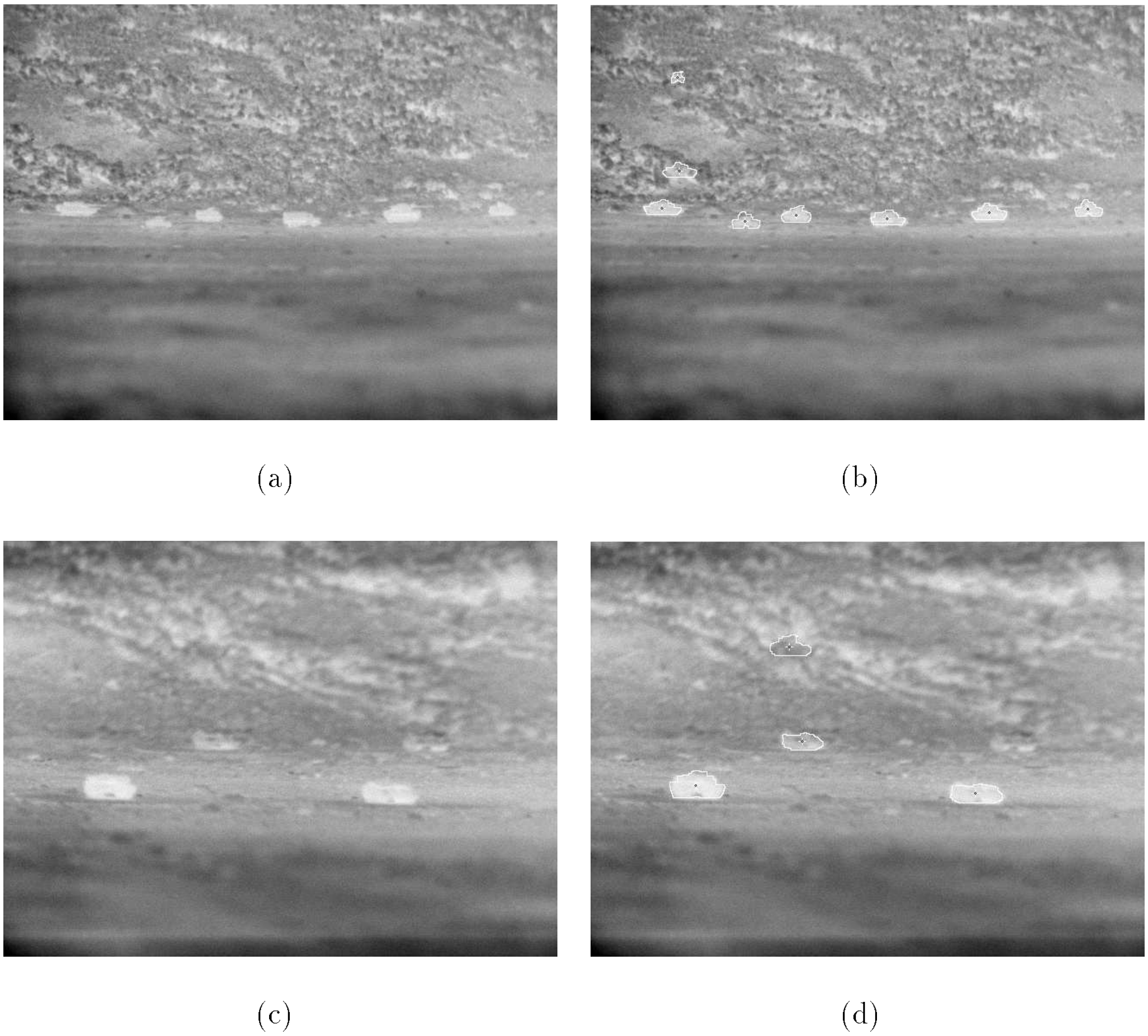


Figure 4: Recognition results on NVL Terrain Board images. (a) and (c) are typical images fed to the algorithm. Recognized targets are outlined in (b) and (c). Note the presence of two false alarms in (b) and one in (d).

on a set of terrain board images. Since the probe image model applies to the terrain board images, it is reasonable to test the algorithm on this set of images.

The defensibility of the notion of using inverted visible imagery as a simulation of infrared imagery depends on the application. Clearly, the approach of Aggarwal and Nandhakumar [1], which uses infrared and visible imagery together to estimate the thermal properties of objects in the image, would not be well served by this imagery, not just because the in-

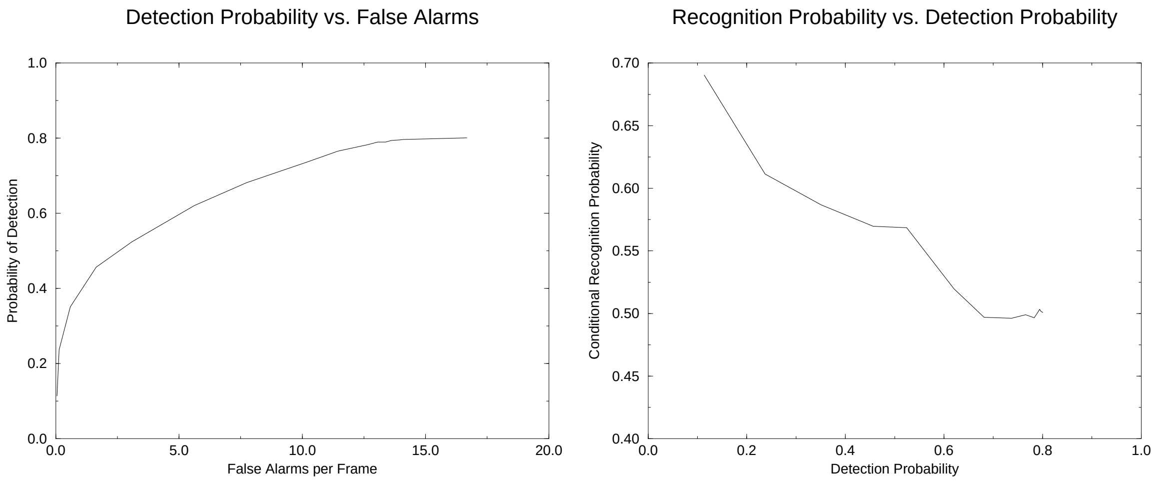


Figure 5: Detection and recognition probabilities on the entire terrain board data set.

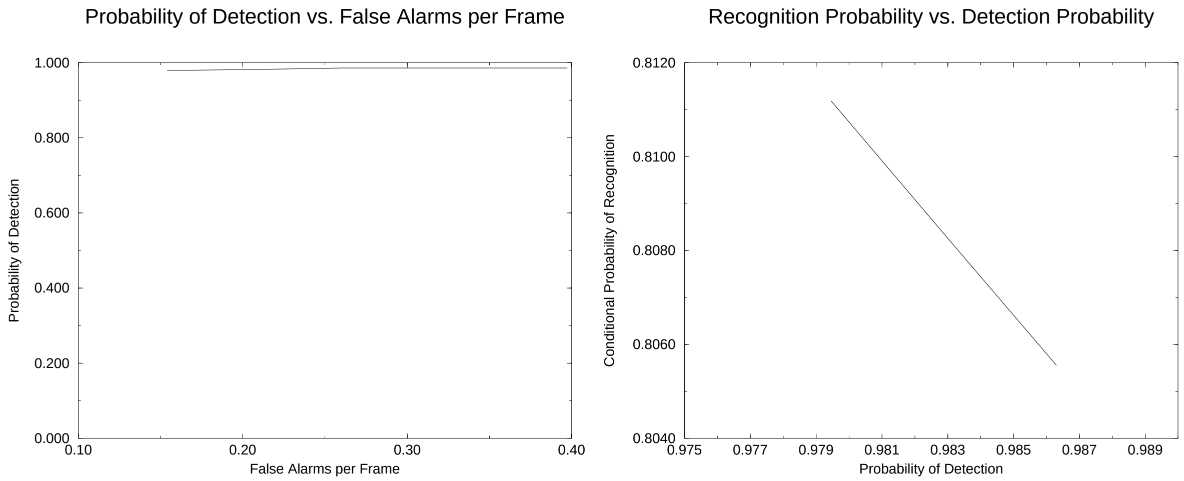


Figure 6: Detection and recognition probabilities on high contrast targets in terrain board imagery.

frared and visible would be essentially the same, but because they rely on relatively subtle phenomenology present in infrared imagery. For algorithms that do not make use of thermodynamic phenomena, the emphasis is on the structure of the objects in the image; the exact grey level values of specific objects are not significant, as long as the contrast between objects and their surrounding backgrounds is great enough. Under certain conditions, as shown below, image inversion is a good approximation to the difference between visible and infrared images. In any case, the difference is that the structure of objects remains the same, while grey level values change significantly. If one is looking for structure and not attributing any significance to grey levels other than using significant changes in grey level to mark locations where there is probably either a change in material or a jump in physical location, then image inversion is a sufficient model for testing algorithms. The majority of object recognition algorithms fall into this category.

Consider a set of objects that are diffuse radiators and can be modeled as grey bodies, which means that reflectivity and emissivity are related as $e = 1 - r$, and the absorptivity $a = e$. Over a somewhat narrow temperature range, the radiance may be treated as a linear function of temperature, as $R = ceT$. For steady state conditions, the irradiation on a surface q can be written as

$$q = R + C_v + C_d = ceT + h(T - T_a) + g(T - T_g)$$

where C_v is energy convected from the surface, C_d is energy conducted away from the surface, h is the convection coefficient, and g is a constant related to the thermal diffusivity of the material. For dry soils the last term plays a minor role, and so may be dropped. Solving for temperature and using $R = ceT$ gives

$$R = ce(T_a + \frac{q - ceT_a}{h + ce})$$

Since, for grey bodies, the reflected energy is proportional to $1 - e$, the radiance should be a linear function of e for inversion to be valid. Figure 7 shows a plot of the radiance as a function of emissivity. Clearly for the conditions outlined, inversion is a reasonable approximation. The conditions tend to hold best for mineral substances, soils and rocks, and not well for vegetation. Thin foliage such as leaves and grass blades, because of transpiration, tends toward the ambient air temperature. Soils and rocks tend to have emissivities between .85 and .95, hence that part of the curve is the most important. Of course, the description above does not apply to objects with internal heat sources. For our experimental setup, those objects were deliberately painted so that inversion would produce reasonable results in the infrared, rather than making them accurate in the visible.

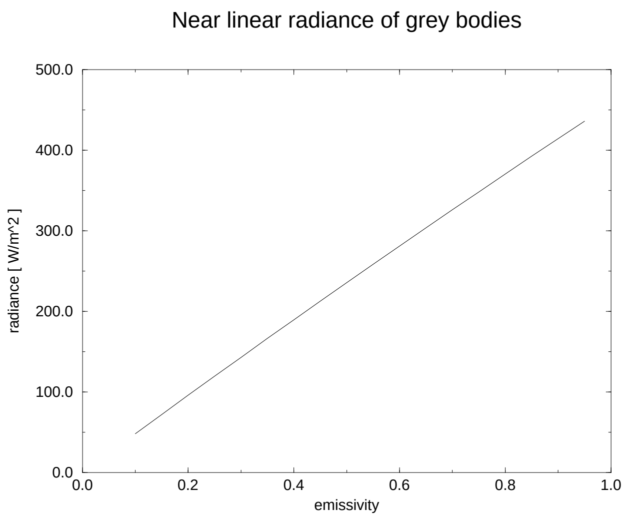


Figure 7: Radiance vs. emissivity at 300K.

6 Conclusions

This paper provides a parametric image model for infrared images, and an associated object recognition scheme. The goal of the work is to use knowledge about the geometry of the set of permissible objects to extract those objects. The algorithm degrades gracefully under decreasing contrast and increasing occlusion. It should be noted that the method allows probes to be defined in any way that assigns a single value to the probe, and causes the probes to have a distribution that is similar to Gaussian. Preliminary experiments suggest that almost any simple combination of sums and differences of grey level values meets this criterion. Also, probes of different types can be combined easily using the likelihood function. Thus, for example, one might use a corner detector at appropriate regions of the silhouette, and edge operators on the standard edges. The probabilities of each probe can be calculated from the local background region as described in the body of the paper.

References

[1] J.K. Aggarwal, and N. Nandhakumar, "Multisensor fusion for automatic scene interpretation", in *Analysis and Interpretation of Range Images*, R. Jain and A. Jain, eds., Springer-Verlag, 1990.

[2] B. Bhanu, "Automatic target recognition: State of the art survey", *IEEE Trans. AES*, Vol. 22, No. 4, July 1986.

[3] B. Bhanu and R. Holben, "Model-based segmentation of FLIR images", *IEEE Trans. AES*, Vol. 26, No. 1, January 1990.

[4] N. Friedland and A. Rosenfeld, "Compact object recognition using energy-function-based optimization", *IEEE Trans. PAMI*, Vol. 14, No. 7, July 1992.

[5] W.E.L. Grimson, *Object Recognition by Computer*, MIT Press, 1990.

[6] R. Harr, "Training Collection at CECOM's Center for Night Vision and Electro-Optics", SPIE International Symposium on Optical Engineering and Photonics in Aerospace Sensing, April 1991.

[7] F. Incorporation and D. Dewitt, *Fundamentals of Heat and Mass Transfer*, Wiley, 1981.

[8] A. Kramer, D. Perschbacher, R. Johnston, and T. Kipp, "Relational template matching algorithm for FLIR automatic target recognition", SPIE Symposium on Aerospace and Remote Sensing, 1993.

[9] A. Margalit, I. Reed, and R. Gagliardi, "Adaptive detection of stationary optical and IR targets using correlated scenes", PhD dissertation, University of Southern California, December 1984.

[10] D. Nguyen, "An iterative technique for target detection and segmentation in IR imaging systems", 36th Conference on the Design of Experiments, University of Delaware, October 1990.

[11] H.V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, 1988.

[12] V.K. Rohatgi, *An Introduction to Probability Theory and Mathematical Statistics*, Wiley, 1976.

[13] R. Staudte and S. Sheather, *Robust Estimation and Testing*, Wiley, 1990.

[14] J.D. Tubbs, "A note on binary template matching", *Pattern Recognition*, Vol. 22, No. 4, 1989.