Minimum Description Length Understanding of Infrared Scenes

Aaron D. Lanterman

Center for Imaging Science (http://cis.wustl.edu)
Electronic Systems and Signals Research Laboratory, Campus Box 1127
Washington University, St. Louis, Missouri 63130-4899
adl@essrl.wustl.edu, http://essrl.wustl.edu/~adl

ABSTRACT

Our pattern theoretic approach to automatic target recognition for infrared scenes combines structured and unstructured representations: rigid, 3-D faceted models for known targets of interest and flexible, simply connected shapes to accommodate the unknown "clutterers" that the algorithm may encounter. The radiant intensities of both kinds of targets form nuisance variables which are incorporated into the parameter space. Statistical inference proceeds by simulating hypothesized scenes and comparing them to the collected data via a likelihood function. For a given target pose, we derive closed-form expressions for estimates of the thermodynamic variables via a weighted least-squares approximation.

Since the number of objects in the scene, both rigid and flexible, is unknown and must be estimated, the parameter space is a union of subspaces of varying dimension. Without constraints on the model order, scene descriptions may become too complex. We apply Rissanen's minimum description length (MDL) principle, which offers a mathematical foundation for balancing the brevity of descriptions against their fidelity to the data. For continuous parameters, the description length involves the log-determinant of the empirical Fisher information matrix. The relationship of Rissanen's MDL to Schwartz's application of Laplace's method of integration and to the Cramer-Rao bound are discussed.

Examples of likelihood surfaces and associated complexity penalties are given for synthetic tank data. In these experiments, the minimum description length approach correctly deduces the number of thermodynamic parameters.

Keywords: pattern theory, minimum description length, automatic target recognition, infrared, FLIR

1. INTRODUCTION

Pattern theory[1,2] provides a mathematical foundation for representing knowledge in complex systems.[3] As codified by Ulf Grenander, the theory emphasizes modeling the underlying objects of inference and brings the methods of traditional Bayesian inference to bear on the complicated parameter spaces needed to represent complex scenes.

For the past five years, our group has applied Grenander's theory to the automatic detection, location, and recognition of ground-based vehicles observed via a forward-looking infrared (FLIR) camera.[4–7] In these studies, three-dimensional, CAD shape-models represent the objects of interest; a likelihood function encapsulates the way the objects are seen by the sensor; and a random sampling algorithm based on jump-diffusion processes facilitates searching over complicated parameter spaces involving unknown numbers of targets.[8,9]

In our previous efforts, two assumptions were made: 1, that the radiant intensities of the objects in the scene was known *a priori*; and 2, that the scene consisted soley of objects contained in the ATR system's target library. The thrust of the present work is to relax those assumptions. Our approach, as always, will be to address them within the pattern-theoretic framework by extending previously employed representations and extending algorithms to perform inference with these new representations.

To tackle problem 1, the vast variability of the FLIR signature of a target under varying environmental and operational conditions, we have extended the parameter space to include parameters describing the thermodynamic state of the targets.[10,11]

Problem 2 is an issue since, no matter how much detail is placed into a particular ATR system, there may always be elements in the scene that the system doesn't know. These elements may appear similar to objects from the library (causing false alarms) or may hinder the ability to recognize or determine the pose of objects that are in the scene (causing missed detections, misclassifications, and estimation errors). We call such elements *clutter*. Clutter may be the most vexing problem facing designers of ATR systems; at the time of this writing, it seems to be generating

a great deal of discussion at conferences and workshops, in the literature, and in the military's calls for research proposals.

Our approach to clutter begins with the observation that the definition of what is a *target* and what is *clutter* is entirely dependent on the particular scenario. Clutter, man made or natural, can be just as interesting, just as complex, and just as *structured* as the targets of interest. One cannot possibly hope to provide shape models for every type of clutterer the system may encounter. Our approach is to introduce generic shape models into the scene representation,[12] flexible models which can bend and mold to accommodate whatever rock, building, lake, or other object the system may encounter. We may not necessarily want to discard such objects immediately; although an object may not match a template from the library, its presence may still be of interest to the user.

Introducing such flexible representations presents a delimma. Why would the algorithm, then, want a rigid target library at all? In fact, the algorithm might be able to do a better job of fitting a nonparametric shape - which can bend to accommodate every nook and cranny of noise - to a tank than the actual tank model in its library. This is a manifestation of the classic problem of model-order estimation. We must imbue our algorithms with an urge to favor simple representations over complex ones. Any particular image may support several equivalent representations. David Mumford has championed[13] the application of information-theoretic notions such as Rissanen's minimum description length principle[14] to pattern-theoretic problems. In this view, estimates of the configuration underlying the data are "descriptions" of the data; if there are several descriptions compatible with the observed data, we select the most parsimonous. The MDL or "minimum complexity"[15,16] principle provides a philosophy for balancing the brevity of descriptions against their fidelity to the data.

As an example of description length computation, consider a simply connected shape on the lattice.[12] It sufficies to define an initial point and a series of directions defining the border. This suggests a complexity measure which is asymptotically proportional to the length of the boundary. This "chain-length" code was suggested by by Leclerc[17] for applying the MDL principle to general image segmentation. A code based on boundary length is the basic method of boundary regularization in the objective functional proposed by Mumford and Shah.[18,19] Notice this is not the only possible technique; other flexible models, such as parametrically defined snakes and balloons,[20] would have other associated description lengths.

The present paper focuses on the description of the thermodynamic and pose parameters associated with rigid targets.

2. MODEL ORDER SELECTION CRITERIA

Since its introduction by Fisher, the method of maximum-likelihood has proven an effective method of vector parameter estimation when the dimension of the parameter space is fixed. When the dimension of the parameter space itself needs to be estimated, maximum-likelihood techniques tend to be "greedy," consistently picking the models of greatest complexity to yield overly tight fits to the data. Bayesian approaches can also suffer from such pathological behavior. For instance, in the maximum a-posteriori estimation of auditory nerve discharge rates,[21] the MAP estimate always picks the model order to be equal to the number of data points.

The challenges of model order estimation were addressed in 1974 in Akaike's influential paper.[22] Four years later, alternative approaches were proposed by Schwartz[23] and Rissanen.[24] Schwartz took a Bayesian approach of integrating out nuisance parameters, and employed Laplace's method of approximating integrals. Adopting a quite different tactic rooted in coding theory, Rissanen proposed the Minimum Description Length principle, which seeks to minimize the number of bits needed to describe the data over the available models. Although his 1978 paper may be the most ubiquitously cited reference for the origins of MDL, the concept of estimation-via-coding was planted a decade earlier in the computer science literature by Wallace and Boulton.[25] Wallace later refined his ideas and termed them Minimum Message Length.[26] The trio of technical reports by Baxter, Hand, and Oliver[27–29] purveys the clearest, most thorough discussion of the similarities and differences between the works Schwartz, Rissanen, and Wallace that we are aware of.

Most Bayesian inference procedures are based on minimizing an expected cost function. Wallace's MML criteria[26,30,31] crafts Bayesian inference procedures without having to specify an explicit cost function (such as squared error or probabilities of detection and false alarm). Rissanen, on the other hand, goes to great lengths to avoid what he considers "subjective" priors and staunchly resists Bayesian interpretations of his criterion. Prefering to make as few assumptions on the parameters as possible, he represents them using a "universal" coding scheme for the integers.[14,32–35] Rissanen's philosophical objections to the notion of prior knowledge nonwithstanding, there is nothing

to prevent devoted Bayesians from assimilating Rissanen's MDL techniques into their own agenda. Considering the rich Bayesian flavor of many pattern theoretic representations, we follow this path in Sec. 2.2. In this sense, our mindset is closer in spirit to that of Wallace, even though the practical choices we make (mostly concerning parameter truncation) follow Rissanen's reasoning.

2.1. Schwartz's application of Laplace's method

Consider the multihypothesis testing problem of determining which model $m \in \mathcal{M}$ generated a given data set y. Suppose each model has equal *a priori* probability. For each model, we need the likelihood and prior densities denoted by $p_l(y|x)$ and $p_p(x)$. (All densities and logdensities are conditioned on a particular candidate model m. In what follows, we suppress the notation of an explicit model index and the uniform *a priori* probability of the model.) The Bayesian procedure proceeds by integrating out the nuisance variable x to find the probability of y under the model:

$$p(y) = \int_x p_l(y|x)p_p(x)dx. \tag{1}$$

In a few specialized cases (such as auditory-nerve discharge rate estimation[21]), the Bayesian integral (1) can be performed analytically. In many cases, however, the integral is formidable, and the approximation technique employed in Schwartz[23] becomes attractive.

Let $L(y|x) = \ln p_l(y|x)$, $P(x) = \ln p_p(x)$, $H(y,x) = L(y|x) + P(x)$, and $\hat{x}(y) = \arg\max_x H(y,x)$. Suppose the model has dimension d. If the posterior density is highly peaked, we can approximate the integrand via Laplace's method (see p. 96 of[36])

$$p(y) = \int_x \exp[H(y,x)]dx \approx \exp[H(y,\hat{x})]\frac{(2\pi)^{d/2}}{\sqrt{\det I_H(y:\hat{x})}}, \tag{2}$$

where

$$I_H(y:\hat{x}) = \left[-\frac{\partial^2}{\partial x_i \partial x_j}H(y,x)|x=\hat{x}\right]. \tag{3}$$

In terms of logarithms, we have

$$\ln p(y) = H(y,\hat{x}) + \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\det I_H(y:\hat{x}). \tag{4}$$

Clarifications and extentions of Schwartz's technique are investigated by Poskitt.[37] Laplace's method has also found application in the study of the information-theoretic asymptotics of Bayes methods[38] and in computing probabilities of detection in low-noise ATR scenarios.[39]

2.2. Rissanen's Minimum Description Length

The overall approach is to describe the observed data y with a two-stage code in which we encode the MAP estimate $\hat{x}$ and then encode y under the model determined by $\hat{x}$. For discrete data y, given x, Shannon's theory tells us we can construct a code with length $\lceil -L(y|x) \rceil$. (For convenience, we will drop the notation for "next largest integer" in the remaining discussion.) Throughout this paper, we will use natural logarithms, so information is expressed in terms of "nats." (One can easily convert from "nats" to "bits" by dividing by $\ln 2$.) For real data y with likelihood density $p_l(y|x)$, we can truncate the data to a desired precision δ_y and replace the density with the probability $p(y|x)\delta_y^d$, yielding a code length $-L(y|x) - d\ln\delta_y$. Notice the $d\ln\delta_y$ term depends on neither the data y nor the parameter x. Hence, the precision δ_y is usually of minor interest and often dropped. In the literature, $-L(y|x)$ is often referred to as a "code length" even when y is real-valued, with the understanding that any desired precision can be achieved. We adopt these conventions here.

Once we've encoded y given x, we also need to encode the parameter x. For discrete parameters, the code length for encoding the parameter x is just $-P(x) = -\ln p_p(x)$. But if x is real, it must be truncated as well. Unlike the truncation of the data, the truncation of the parameters is a significant issue.

Expanding the loglikelihood as a function of the truncated value in a Taylor series around $\hat{x}$ (to second order) yields

$$-H(y,x_{tr}) \approx -H(y,\hat{x}) - \frac{1}{2}(\hat{x}-x_{tr})^T \left.\frac{\partial^2}{\partial x_i x_j}H(y,x)\right|_{x=\hat{x}} (\hat{x}-x_{tr}). \tag{5}$$

In his earliest work,[24] Rissanen took the approach of Wallace and Boulton,[25] who treated the quantization error $\hat{x} - x_{tr}$ as a uniformly distributed random vector and chose the precision to minimize the expected value of (5). (This is the approach taken by D. Michal for multiple target detection[40] with radar antenna arrays.) However, in all later papers, Rissanen adopts a worst-case minimax approach which picks the precision to minimize the maximum of (5) over the quantization region: "Instead of the maximum value we could calculate the mean increase by assuming some distribution [uniform in Wallace and Freeman] for the deviation of [the parameter] from the center of its enclosing rectangle. *Using the maximum value has the advantage, however, that it is independent of such distributions.*"[14] [Italics added.]

Rissanen considers the problem of truncation in multiple dimensions as partitioning the parameter space into parallelepipeds, noting that the quadratic term reaches its maximum when x falls at a corner of the parallelpiped. Consider a fixed maximum deviation

$$r = (\hat{x} - x_{tr})^T I_H(y : \hat{x})(\hat{x} - x_{tr}). \tag{6}$$

Rissanen chooses the quantizing parallelpiped as the maximum-volume rectangle[14] contained within the ellipsoid (6). This rectangle has edges parallel to the principal axes of the ellipsoid, with a maximum volume of

$$V(r) = \left(\frac{4r}{d}\right)^{(1/2)d} \frac{1}{\sqrt{\det I_H(y : \hat{x})}}. \tag{7}$$

For Rissanen's worst-case analysis, we seek the r which minimizes

$$\begin{aligned}
& -\ln V(r) p_p(x_{tr}) - L(y|x_{tr}) && (8)\\
= \; & -\ln V(r) - H(y, x_{tr}) && (9)\\
\approx \; & -\ln V(r) - H(y, \hat{x}) + \frac{r}{2} && (10)\\
= \; & -\frac{d}{2}\ln 4r + \frac{d}{2}\ln d + \frac{1}{2}\ln\det I_H(y : \hat{x}) - H(y, \hat{x}) + \frac{r}{2}. && (11)\\
& && (12)
\end{aligned}$$

Setting first derivatives equal to zero easily reveals the minimizing r to be $r = d$. Substituting $r = d$ into (12) yields the description length

$$-H(y, \hat{x}) + \frac{1}{2}\ln\det I_H(y : \hat{x}) + \frac{d}{2}(1 - \ln 4). \tag{13}$$

Although it follows straightforwardly by viewing Rissanen's ideas from a Bayesian angle, to the best of our knowledge, (13) has not explicitly appeared before in the literature.

2.3. Wallace's Minimum Message Length

Instead of Rissanen's parallelpipeds, Wallace and Freeman[26,31] consider quantizing in multiple dimensions using optimal quantizing lattices. For instance, in two dimensions, the optimal quantizing lattice forms a hexagonal grid. In three dimensions, the optimal lattice is a body-centered cubic lattice (p. 60 of[41]), whose Voroni regions (p. 34 of[41]) are *truncated octahedrons* (one of the Archimedean polyhedra). In higher dimensions, the optimal quantizing lattices are actually unknown, although bounds are available. Instead of working out a worst-case analysis as in Rissanen's later work, Wallace and Freeman consider the truncation error to be a uniformly distributed random variable. The resulting criterion has the same posterior and Fisher information terms as Schwartz's and Rissanen's criterion, with an additional term involving the efficiency of the quantizing lattice. Our group is currently investigating this third approach. Due to space constraints we will not consider it further here.

2.4. Penalized loglikelihood interpretation

As suggested by P. Green,[42] it is often fruitful to express model selection procedures via penalized loglikelihoods:

$$L_{pen}(x, y) = L(y|x) - C(x, y) \tag{14}$$

The penalties associated with Schwartz's and Rissanen's approaches are given by

$$C_R(x,y) = -P(x) + \frac{1}{2}\ln\det I_H(y:x) - \frac{d}{2}(\ln 2\pi),\ C_S(x,y) = -P(x) + \frac{1}{2}\ln\det I_H(y:x) - \frac{d}{2}(\ln 4 - 1) \tag{15}$$

In the applied literature, C is occasionally interpreted as a logprior; this is somewhat misleading since it includes the data. We prefer to think of C as an adaptive complexity penalty.

2.5. Relationship to the Cramer-Rao bound

Although they arrive from quite different viewpoints, Rissanen and Schwartz yield strikingly similar criteria. Both criteria have terms involving the posterior, a term which is one-half the log of the observed Fisher information, and a term which is linear in the number of parameters. Both techniques involve a Taylor series expansion, so the appearance of the Fisher information terms is perhaps not too surprising. This section focuses on the meaning of the Fisher information term. Recall the multivariate Cramer-Rao bound for random parameters given by

$$E_x\{E_{y|x}\{[\hat{x}(y) - x]^T[\hat{x}(y) - x]\}\} \geq [E_x\{E_{y|x}\{I_H(y:x)\}\}]^{-1}, \tag{16}$$

where we have iterated the expectation. The inner expectation on the right-hand side of corresponds to averaging the observed Fisher information, a prominent feature of both the Schwartz and Rissanen model order selection criteria, over the data. Viewed this way, the Cramer-Rao bound offers an interpretation of the Fisher information in the Schwartz/Rissanen penalty: it provides an estimate of the the accuracy of the MAP parameter estimates.

3. DATA LIKELIHOOD

Adopting the detector model of Snyder, Hammoud, and White,[43] we suppose the data $y(k), k \in \mathcal{D}$ collected by camera is Poisson distributed with mean $\lambda(k)$. The Poisson loglikelihood[44] of the data y given λ is

$$L_{CCD}(y|\lambda) = -\sum_{k\in\mathcal{D}} \lambda(k) + \sum_{k\in\mathcal{D}} y(k)\ln\lambda(k). \tag{17}$$

The loglikelihood could be generalized to accommodate optical blurring, readout noise, and background counts.[43,45] We neglect them here as a first-order approximation in order to generate the results of Sec. 4.3.

Of course, we are not interested in the ideal image λ directly, but in the underlying parameters x, the poses and shapes of objects of interest. Let $render : x \to \lambda$ represent the function which generates an ideal infrared image of the underlying objects with their associated poses, shapes, and thermodynamic states. This involves viewing objects through the effects of perspective projection and obscuration. We have developed rapid implementations of *render* for Silicon Graphics computers.

The sensor loglikelihood L_{CCD} and the *render* function provide a conceptual separation of the stochastic and deterministic aspects of the remote sensing process. Combining them yields the composite loglikelihood

$$L(y|x) = L_{CCD}(y|render(x)). \tag{18}$$

4. THERMODYNAMIC NUISANCE VARIABLES

We take an empirical statistical-approach to modeling target thermodynamics and construct prior distributions on the radiant intensities of target facets. By simulating a large number of radiance measurements, taken while varying environmental and internal heating parameters over reasonable ranges, we generate a population of radiance profiles to which we apply principal component analysis. Assuming a Gaussian model, the first few eigenvectors - here called 'eigentanks' - provide a parsimonous representation of the covariance. (For instance, the smallest model we currently employ is an M2 model that contains 95 regions. Without reducing dimentionality with the eigenvector or some other approach, we would have to deal with a 95 by 95 covariance matrix.) Our exploration of principal-component methods was motivated by the success of such methods in modeling face images[46,47] and hippocampi.[48,49]

We will need some notation. Suppose the tank is divided into I regions, with intensity assumed constant across each region, and that we are employing J basis functions. Let N_i be number of pixels in region i, as seen by the detector, $Y_i = \sum_{k\in R_i} y(k)$ be the sum of data pixels in region i, and λ_i be the *unknown* mean intensity of region i. We represent $\lambda_i = \sum_j \alpha_j \Phi_{ij} + m_i$, where the α_j's are expansion coefficients, Φ_{ij} is the value of eigentank j at region i, m_i is the mean of region i, and γ_j is the eigenvalue associated with eigentank j.

4.1. Logposterior for rigid targets

Conditioned on the α_j's, $y(k) \sim Poisson(\lambda_i)$ for $k \in R_i$, yielding $Y_i \sim Poisson(N_i\lambda_i)$. Dropping terms independent of α, the logposterior is

$$H(\alpha|Y) = -\sum_i N_i\lambda_i + \sum_i Y_i \ln N_i\lambda_i - \sum_j \frac{\alpha_j^2}{2\gamma_j}. \tag{19}$$

Expanding λ in terms of eigentanks, we have

$$\begin{aligned} H(\alpha|Y) &= -\sum_i N_i(\sum_j \alpha_j\Phi_{ij} + m_i) + \sum_i Y_i \ln[N_i(\sum_j \alpha_j\Phi_{ij} + m_i)] - \sum_j \frac{\alpha_j^2}{2\gamma_j} && (20)\\ &= -\sum_i N_i \sum_j \alpha_j\Phi_{ij} - \sum_i N_i m_i + \sum_i Y_i \ln N_i && (21)\\ &\quad + \sum_i Y_i \ln(\sum_j \alpha_j\phi_{ij} + m_i) - \sum_j \frac{\alpha_j^2}{2\gamma_j}. && (22) \end{aligned}$$

Again dropping terms independent of α_j yields:

$$H(\alpha|Y) = -\sum_i N_i \sum_j \alpha_j\Phi_{ij} + \sum_i Y_i \ln(\sum_j \alpha_j\Phi_{ij} + m_i) - \sum_j \frac{\alpha_j^2}{2\gamma_j}. \tag{23}$$

4.2. Necessary conditions for a maximizer

Setting the derivative w.r.t. α_j equal to zero, we find the necessary conditions:

$$-\sum_i N_i\Phi_{ij} + \sum_i \frac{\Phi_{ij}}{\sum_k \alpha_k\Phi_{ik} + m_i} Y_i - \frac{\alpha_j}{\gamma_j} = 0, \ \forall j. \tag{24}$$

This is a set of coupled nonlinear equations for which no direct closed-form solution is evident. While we have derived an expectation-maximization algorithm for their solution, we are drawn to a computationally more efficient weighted least-squares approximation.

4.3. Weighted least-squares approximation

Here we approximate the posterior by replacing the Poisson loglikelihood (conditioned on the α's) using a weighted least-squares approximation:

$$H(\alpha|Y) \approx -\sum_i \frac{(N_i\lambda_i - Y_i)^2}{2Y_i} - \sum_j \frac{\alpha_j^2}{2\gamma_j}. \tag{25}$$

Incorporating $\lambda_i = \sum_j \alpha_j\Phi_{ij} + m_i$ yields

$$H(\alpha|Y) \approx -\sum_i \frac{[N_i(\sum_j \alpha_j\Phi_{ij} + m_i) - Y_i]^2}{2Y_i} - \sum_j \frac{\alpha_j^2}{2\gamma_j}. \tag{26}$$

Taking the derivative w.r.t. α_j as usual, we find J necessary conditions:

$$\begin{aligned} -\sum_i \frac{[N_i(\sum_k \alpha_k\Phi_{ij} + m_i) - Y_i]N_i\Phi_{ij}}{Y_i} - \frac{\alpha_j}{\gamma_j} &= 0, \ \forall j && (27)\\ -\sum_k \alpha_k \sum_i \frac{N_i^2\Phi_{ik}\Phi_{ij}}{Y_i} - \sum_i \left[\frac{N_i^2\Phi_{ij}m_i}{Y_i} - N_i\Phi_{ij}\right] - \frac{\alpha_j}{\gamma_j} &= 0, \ \forall j. && (28) \end{aligned}$$

Fortunately these are linear equations, conveniently expressed in matrix form:

$$\left(\begin{bmatrix} \sum_i \frac{N_i^2 \Phi_{i1}^2}{Y_i} & \cdots & \sum_i \frac{N_i^2 \Phi_{iJ} \Phi_{i1}}{Y_i} \\ \vdots & & \vdots \\ \sum_i \frac{N_i^2 \Phi_{i1} \Phi_{iJ}}{Y_i} & \cdots & \sum_i \frac{N_i^2 \Phi_{jJ}^2}{Y_i} \end{bmatrix} + diag \begin{pmatrix} 1/\gamma_1 \\ \vdots \\ 1/\gamma_J \end{pmatrix} \right) \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_J \end{bmatrix} = \begin{bmatrix} \sum_i (N_i \Phi_{i1} - \frac{N_i^2 \Phi_{i1} m_i}{Y_i}) \\ \vdots \\ \sum_i (N_i \Phi_{iJ} - \frac{N_i^2 \Phi_{iJ} m_i}{Y_i}) \end{bmatrix}. \quad (29)$$

For a given target pose, these equations allow us to compute MAP estimates for the α_j's in closed form, which we denote as $\hat{\alpha}_j$.

4.4. Empirical Fisher information for thermodynamic variables

The elements of the observed Fisher information matrix are given by

$$-\frac{\partial}{\partial \alpha_j} \frac{\partial}{\partial \alpha_j} H = \sum_i \frac{\Phi_{ij} \Phi_{il}}{(\sum_k \alpha_k \Phi_{ik} + m_i)^2} Y_i + \frac{1}{\gamma_j} \delta_{jl}, \quad (30)$$

where δ is a Kronecker delta, and we evaluate (30) at the MAP estimate $\hat{\alpha}$ (see p. 98 of[50]). In implementation, we compute $\log \det I(Y : \alpha)$ by summing the log-eigenvalues of $I(Y : \alpha)$; this avoids underflows which might occur in attempting to compute $\det I(Y : \alpha)$ directly. We have found that such underflows can occur when a large number of terms in the eigenexpansion are employed.

4.5. Implementation on Silicon Graphics hardware

We have previously exploited the powerful real-time rendering capabilities of Silicon Graphics computers to simulate infrared images and simulate laser radar range images by reading the z-buffer values which result as a byproduct of the 3-D rendering process.[7,51] Here, we once again benefit from the flexibility of SGI's systems. To compute the N_i's and D_i's needed in Secs. 4.3 and 4.4, the number of pixels from each region appearing on the detector and the sum of the data values across each region, we first render the scene using region numbers instead of intensities. The rendered image is scanned, and each pixel is placed in the appropriate bin, adding one to the accumulating sum for N_i, while the data corresponding to that pixel is added to the appropriate D_i. We refer to this technique as "painting by numbers." Coefficients may be estimated using (29) and the empirical Fisher information computed using (30). With the coefficients estimated, the corresponding Poisson loglikelihood can be computed via $-\sum_i N_i \lambda_i + \sum_i D_i \ln N_i \lambda_i$ with $\lambda_i = \sum_j \alpha_j \Phi_{ij} + m_i$. Also, a hypothesized infrared image can be directly simulated in a second rendering, this time using λ_i's instead of region numbers.

5. EXAMPLE

Figure 1 features 140 by 140 images of a synthetic T62 superimposed on a real infrared background. The background is from a data set provided courtesy Dr. James Ratches of the U.S. Night Vision and Electronic Sensors Directorate. The camera position was chosen so that the tank appears in correct perspective when viewed against the background. The radiant intensities of the facets were generated from a principal-component eigentank model generated by Matthew Cooper,[10,11] with the first three coefficients of the T62's eigenexpansion set to nonzero values and remaining coefficients set to zero.

In all of the following experiments, the background was assumed to have an unknown, uniform intensity, adaptively estimated using its maximum-likelihood estimate, which in this case is just the average of the background pixels. This parameter would have an associated description length penalty; this penalty will be neglected in the following experiments. We are currently exploring the possibilities for encoding such background parameters.

We generated a data set from the tank at 45 degrees (second frame, top row of Fig. 1.) Given the correct position, Fig. 2 shows the loglikelihood varying with hypothesized orientation for this simulated data. From bottom to top, the lines correspond to loglikelihoods computing using the first 1, 3, 5, and 17 terms in the thermodynamic eigenexpansion. As expected, there is a sharp peak at the correct orientation of 45 degrees. In addition, using more terms in the expansion generates higher loglikelihoods, especially when the orientation is far from the correct value, since the larger number of parameters gives the model greater flexibility in trying to fit the tank to the background.

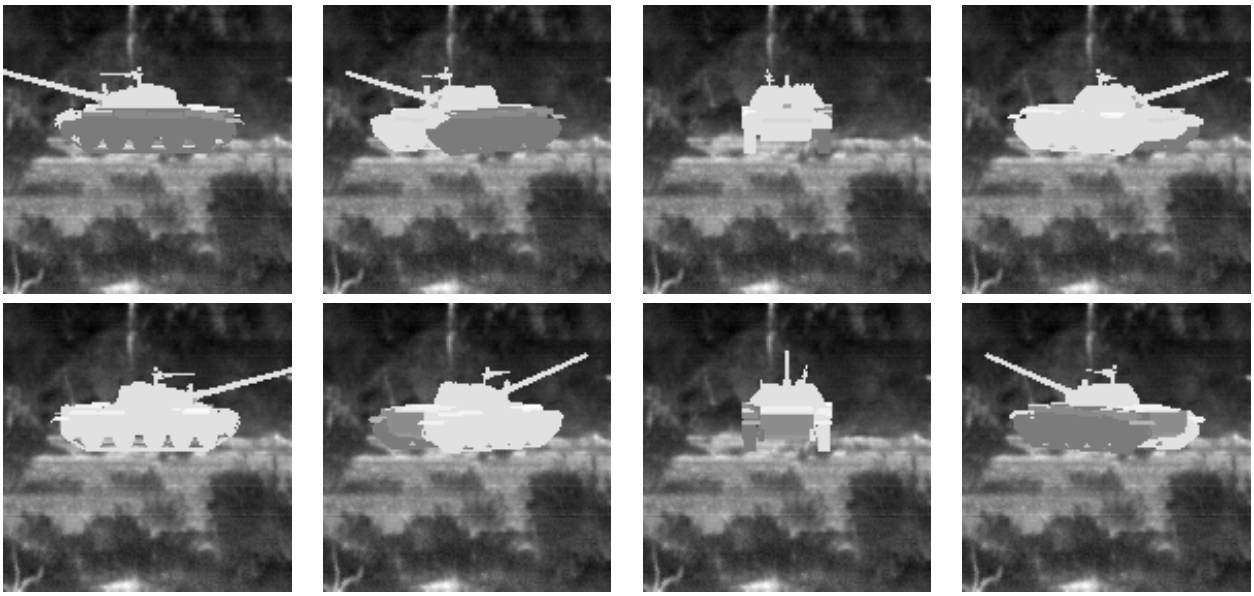

Figure 1. Synthetic T62 superimposed over a real infrared background, with orientation at 45 degree increments. Real infrared data courtesy Dr. Jim Ratches, Night Vision and Electronic Sensors Directorate.

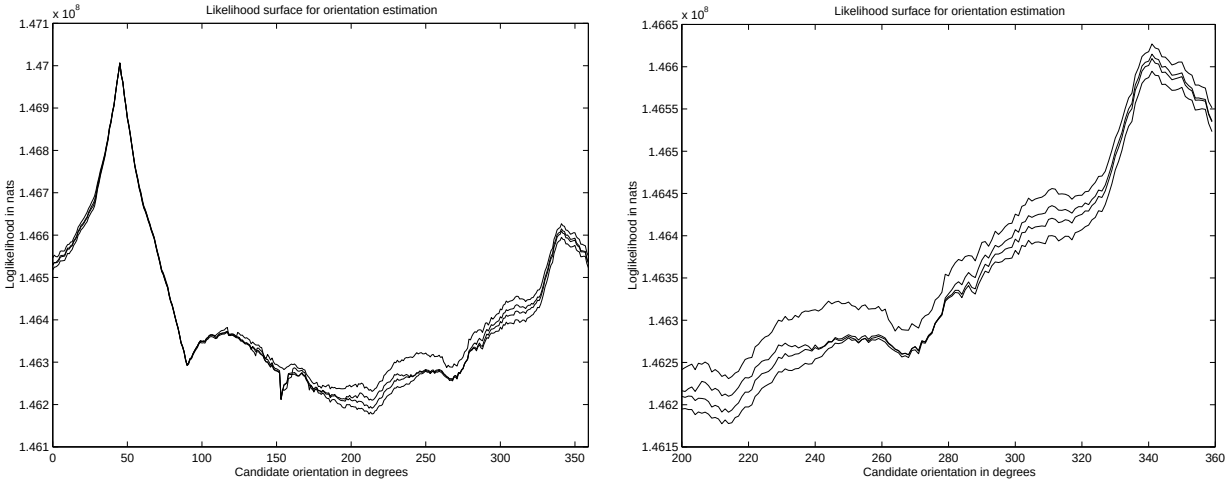

Figure 2. Left panel shows the loglikelihood varying with respect to orientation for data generated with a true orientation of 45 degrees. From bottom to top, lines correspond to likelihoods computed using 1, 3, 5 and 17 eigenvalues. Right panel shows zoom of a section of the left panel.

Fig. 3 further explores the effect of changing the number of terms in the eigenexpansion. Assuming a correct hypothesized orientation of 45 degrees, the left panel shows the Rissanen (top line) and Schwartz (bottom line) penalty terms associated with the given number of eigentanks. These complexity penalities are given by

$$C_S(\alpha, y) \quad = \quad \frac{1}{2}\left\{ \sum_j \ln \gamma_j + \sum_j \frac{\alpha_j^2}{\gamma_j} + \ln \det I_H(y:\alpha) \right\} \tag{31}$$

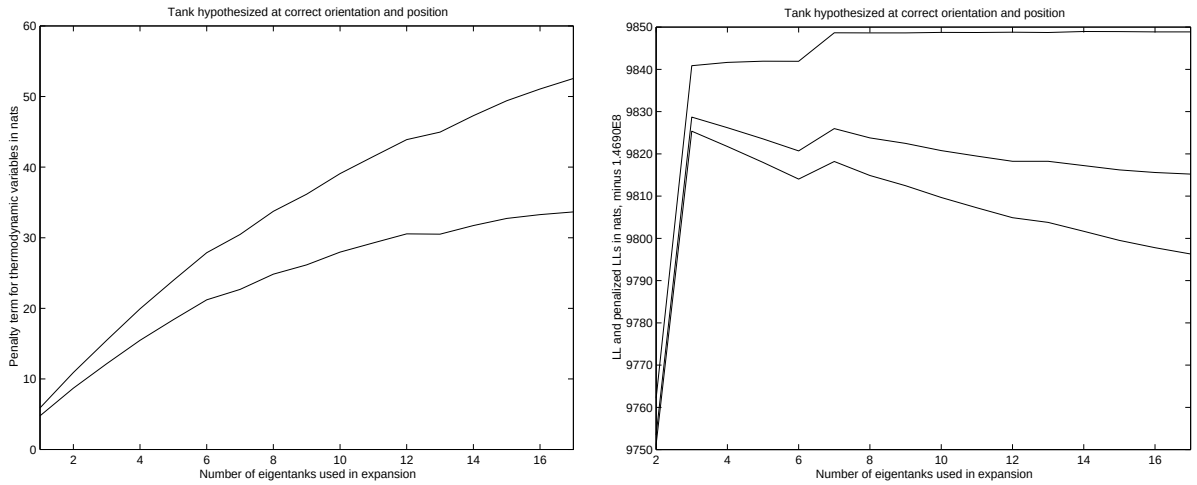

Figure 3. Assuming a correct orientation of 45 degrees, the left panel plots penalty term for thermodynamic nuisance variables with respect to number of eigentanks employed in the expansion. Top line corresponds to Rissanen's method of parameter truncation; bottom line corresponds to Schwartz's application of Laplace's method. From top to bottom, right panel shows: loglikelihood at 45 degrees with respect to number of terms in expansion; penalized loglikelihood using Schwartz's method; penalized loglikelihood using Rissanen's method.

$$C_R(\alpha, y) \quad = \quad \frac{1}{2}\left\{ d \ln 2\pi + \sum_j \ln \gamma_j + \sum_j \frac{\alpha_j^2}{\gamma_j} + \ln \det I_H(y:\alpha) - d[\ln(4) - 1] \right\} \tag{32}$$

where the elements of I_H are given by (30). Notice that in the Schwartz penalty, the $\ln 2\pi$ term from the prior cancels the $\ln 2\pi$ term from the Laplace approximation. The Rissanen penalty is harsher.

The effect of subtracting these penalty terms from the loglikelihood is shown in the right panel. The top line is the loglikelihood associated with a given number of terms; notice that it increases monotonically with the number of terms, as expected. The next line from the top represents the result of subtracting the Schwartz term; the bottom line is the result of subtracting the Rissanen term. For both types of penalties, the penalized loglikelihood increases rapidly at first, reaching a peak at three terms, and then slowly decreases. The maximum penalized-likelihood in each case is at three, which is as expected, since in our simulation we set the fourth and higher order terms to zero.

Now we explore how the penalty terms vary with respect to true orientation. To approximate the expected value of the penalty term under low-noise conditions, we computed the penalty term using noiseless data. Fig. 4 shows graphs of the Schwartz penalty term for the thermodynamic variables with respect to orientation (the graph for the Rissanen term would have the same shape, just vertically shifted). In the left panel only the first term in the eigenexpansion is used; in the right panel, the first three terms are used. Notice that between 90 and 200 degrees, the tank presents a relatively constant intensity to the detector. The number of nats used to represent the thermodynamic state is lower in this region. However, as the tank continues to spin, passing 250 degrees, we begin to see the other side, which shows more contrast within the tank. This additional information in the data is manifest in the higher number of nats in the parameter description length.

We conducted a similar experiment, this time plotting the penalty term for orientation vs. true orientation. For this, a uniform prior on tank orientation is assumed, and the observed Fisher information (the negative second derivative with respect to orientation) is computed numerically using a stepsize of two degrees. In computing the loglikelihoods, the first three terms in the eigenexpansion were used. Note again that the penalty term is higher at orientations where interesting features are apparent in the tank.

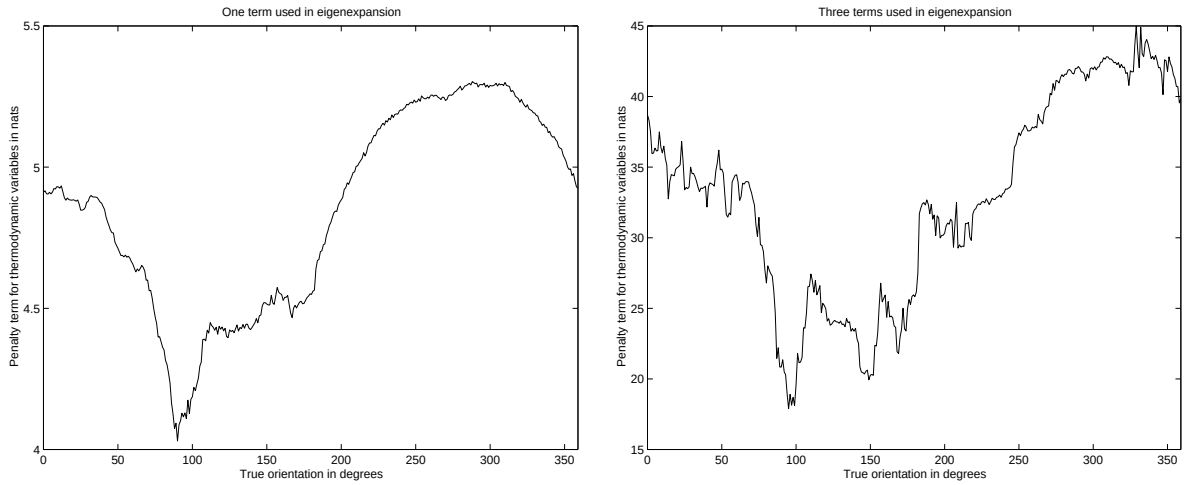

Figure 4. Schwartz penalty term for estimating thermodynamic nuisance parameters, assuming correct orientation is hypothesized, varying with respect to true orientation.

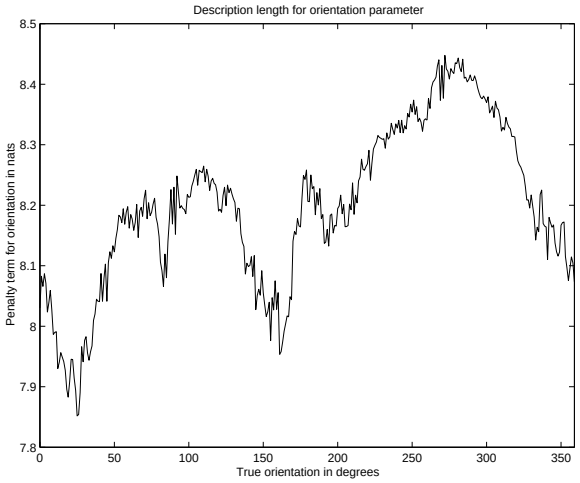

Figure 5. Penatly term (not including the $\ln 2\pi$ or $\ln(4) - 1$ terms associated with Schwartz's and Rissanen's methods, respectively) for estimating orientation parameter, varying with respect to true orientation. Loglikelihoods computed with first three eigenterms in expansion.

6. ACKNOWLEDGEMENTS

This work was supported by ARO DAAH04-95-0494, ONR N00014-94-1-0859, ONR/AASERT N00014-94-1-1135, and ARO/AASERT DAAH04-94-G-0209.

REFERENCES

1. U. Grenander, *Elements of Pattern Theory*, Johns Hopkins Univ. Press, 1996.
2. U. Grenander, *General Pattern Theory*, Oxford Univ. Press, 1994.
3. U. Grenander and M. I. Miller, "Representations of knowledge in complex systems," *Journal of the Royal Statistical Society B* **56**(3), pp. 549–603, 1994.

4. A. Lanterman, M. Miller, D. Snyder, and W. Miceli, "Jump-diffusion processes for the automated understanding of FLIR scenes," in *Automatic Object Recognition IV*, F. Sadjadi, ed., vol. SPIE Proc. 2234, pp. 416–427, (Orlando, FL), April 1994.
5. A. Lanterman, M. Miller, and D. Snyder, "Implementation of jump-diffusion processes for understanding flir scenes," in *Automatic Object Recognition V*, F. Sadjadi, ed., vol. SPIE Proc. 2485, pp. 309–320, (Orlando, FL), April 1995.
6. A. Lanterman, M. Miller, and D. Snyder, "Automatic target recognition via the simulation of infrared scenes," in *Proc. of the 6th Annual Ground Target Modeling and Validation Conf.*, pp. 195–203, U.S. Army TACOM, (Houghton, MI), August 1995.
7. A. Lanterman, M. Miller, D. Snyder, and W. Miceli, "The unification of detection, tracking, and recognition for millimeter wave and infrared sensors," in *Radar/Ladar Processing*, W. Miceli, ed., vol. SPIE Proc. 2562, pp. 150–161, (San Diego, CA), August 1995.
8. A. Lanterman, "Jump-diffusion algorithms for the automated understanding of forward-looking infrared scenes," Master's thesis, Washington University, St. Louis, MO, May 1995.
9. A. Lanterman, M. Miller, and D. Snyder, "General Metropolis-Hastings jump diffusions for automatic target recognition in infrared scenes," *Optical Engineering* **36**, pp. 1123–1137, April 1997.
10. M. Cooper, A. Lanterman, S. Joshi, and M. Miller, "Representing the variation of thermodynamic state via principle components analysis," in *Proceedings of the Third Workshop on Conventional Weapon ATR*, November 1996.
11. M. Cooper, U. Grenander, M. Miller, and A. Srivastava, "Accommodating geometric and thermodynamic variability for forward-looking infrared sensors," in *Algorithms for Synthetic Aperture Radar IV*, E. Zelnio, ed., vol. SPIE Proc. 3070, pp. 162–172, (Orlando, FL), April 1997.
12. A. Lanterman, M. Miller, and D. Snyder, "Representations of shape for structural inference in infrared scenes," in *Automatic Object Recognition VII*, F. Sadjadi, ed., vol. SPIE Proc. 3069, pp. 257–268, (Orlando, FL), April 1997.
13. D. Mumford, "Pattern theory: a unifying perspective," in *Proceedings 1st European Congress of Mathematics*, pp. 187–224, Birkhauser, 1994.
14. J. Rissanen, "A universal prior for integers and estimation by minimum description length," *The Annals of Statistics* **11**, pp. 416–431, 1983.
15. A. Barron and T. Cover, "Minimum complexity density estimation," *IEEE Trans. on Information Theory* **37**, July 1991.
16. Y. Yang and A. Barron, "An asymptotic property of model selection criteria," *IEEE Trans. on Information Theory* **44**, January 1998.
17. Y. Leclerc, "Constructing stable descriptions for image partitioning," *International Journal of Computer Vision* **3**, pp. 73–102, 1989.
18. D. Mumford and J. Shah, "Optimal approximations of piecewise smooth functions and associated variational problems," *Comm. in Pure and Appl. Math.* **42**, pp. 577–685, 1989.
19. J. Morel and S. Solimini, *Variational Methods in Image Segmentation*, Birkhauser, 1995.
20. S. Zhu, T. S. Lee, and A.Yuille, "Region competition: Unifying snake, region growing and Bayes/MDL for multi-band image segmentation," *IEEE Trans. Patt. Anal. Mach. Int.* , in review.
21. K. E. Mark and M. I. Miller, "Bayesian model selection and minimum description length estimation of auditory-nerve discharge rates," *Journal of the Acoustical Society of America* **91**, pp. 989–1002, February 1992.
22. H. Akaike, "A new look at the statistical model indentification," *IEEE Trans. on Automatic Control* **19**, pp. 716–723, December 1974.
23. G. Schwartz, "Estimating the dimension of a model," *Annals of Statistics* **6**, pp. 461–464, 1978.
24. J. Rissanen, "Modeling by shortest data description," *Automatica* **14**, pp. 465–471, 1978.
25. C. Wallace and D. Boulton, "An information measure for classification," *Computer Journal* **11**, pp. 195–209, August 1968.
26. C. Wallace and P. Freeman, "Estimation and inference by compact coding," *Journal of the Royal Statistical Society B* **49**(3), pp. 241–252, 1987.
27. J. Oliver and D. Hand, "Introduction to minimum encoding inference," Technical Report 205, Department of Computer Science, Monash University, 1994. available at `http://www.cs.monash.edu.au/~jono/Open_Uni/TR4-94.ps`.

28. J. Oliver and R. Baxter, "MML and Bayesianism: Similarities and differences (introduction to Minimum Encoding Inference - part II)," Technical Report 206, Department of Computer Science, Monash University, 1994. available at `http://www.cs.monash.edu.au/~jono/TechReports/intro.2.ps`.
29. R. Baxter and J. Oliver, "MDL and MML: Similarities and differences (introduction to Minimum Encoding Inference - part III)," Technical Report 207, Department of Computer Science, Monash University, 1994. available at `http://www.cs.monash.edu.au/~rohan/PAPERS/intro.3.ps`.
30. A. O'Hagan, "Contribution to the discussion of the papers by Dr. Rissanen and Professors Wallace and Freeman," *Journal of the Royal Statistical Society B* **49**(3), pp. 256–257, 1987.
31. C. Wallace and P. Freeman, "Single-factor analysis by minimum message length estimation," *Journal of the Royal Statistical Society B* **54**(1), pp. 195–209, 1992.
32. J. Rissanen, "Stochastic complexity and modeling," *The Annals of Statistics* **14, no.3**, pp. 1080–1100, 1986.
33. J. Rissanen, "Stochastic complexity," *Journal of the Royal Statistical Society B* **49**(3), pp. 223–239, 1987.
34. J. Rissanen, "Fisher information and stochastic complexity," *IEEE Transactions on Information Theory* **42**, pp. 40–47, January 1996.
35. G. Qian and H. Kunsch, "Some notes on Rissanen's stochastic complexity," *IEEE Trans. on Information Theory* **44**, pp. 782–896, March 1988.
36. G. Polya and G. Szego, *Problems and Theorems in Analysis I*, Springer-Verlag, New York, 1972.
37. D. Poskitt, "Precision, complexity and bayesian model determination," *Journal of the Royal Statistical Society B* **49**(2), pp. 199–208, 1987.
38. B. Clarke and A. Barron, "Information-theoretic asymptotics of bayes methods," *IEEE Trans. on Information Theory* **36**, pp. 453–471, May 1990.
39. U. Grenander, A. Srivastava, and M. Miller, "Performance analysis of Bayesian object recognition in computer vision," to be submitted.
40. D. Michal, *Multiple Target Detection for an Antenna Array*, Doctoral Dissertation, Department of Electrical Engineering, Sever Institute of Technology, Washington University, St. Louis, MO, August 1993.
41. J. Conway and N. Sloane, *Sphere Packings, Lattices and Groups*, Springer-Verlag, New York, second ed., 1993.
42. P. Green, "Penalized likelihood," in *Encyclopedia of Statistical Sciences*, update volume, to be published; available at `www.stats.bris.ac.uk/~peter/abstracts/penlik.html`.
43. D. Snyder, A. Hammoud, and R. White, "Image recovery from data acquired with a charge-coupled-device camera," *Journal of the Optical Society of America A* **10**, pp. 1014–1023, May 1993.
44. D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*, Springer-Verlag, 2nd ed., 1991.
45. D. L. Snyder, C. W. Helstrom, A. D. Lanterman, M. Faisal, and R. L. White, "Compensation for readout noise in CCD images," *J. Optical Society of America A* **12**, pp. 272–283, 1995.
46. P. Hallinan, "A low-dimensional representation of human faces for arbitrary lighting conditions," *IEEE Conf. Comp. Vis. Patt. Rec., Seattle* , 1993.
47. A. Yuille, "Deformable templates for face recognition," *Journal of Cognitive Neuroscience* **3**, 1991.
48. S. Joshi, *Large Deformation Diffeomorphisms and Gaussian Random Fields for Statistical Characterization of Brain SubManifolds*, Doctoral Dissertation, Department of Electrical Engineering, Sever Institute of Technology, Washington University, St. Louis, MO, August 1997.
49. S. Joshi, U. Grenander, and M. Miller, "On the geometry and shape of brain sub-manifolds," *International Journal of Pattern Recognition and Artificial Intelligence: Special Issue on Processing of MR Images of the Human* , accepted November 1996, to appear 1998.
50. J. Lindsey, ed., *Parametric Statistical Inference*, Oxford Univ. Press, 1996.
51. J. Kostakis, M. Cooper, T. Green, M. Miller, J. O'Sullivan, J. Shapiro, D. Snyder, and J. O'Sullivan, "Multispectral active-passive sensor fusion for ground-based target orientation estimation," in *Automatic Target Recognition VII*, F. Sadjadi, ed., vol. SPIE Proc. 3371, (Orlando, FL), April 1998. to be published.