Journal of Electronic Imaging 9(2), 178–193 (April 2000).

FLIR ATR using location uncertainty

Gang Liu
Robert M. Haralick
University of Washington
Intelligent Systems Laboratory
Department of Electrical Engineering
Seattle, Washington 98195-2500
E-mail: {gliu,haralick}@ee.washington.edu

Abstract. *A model-based FLIR ATR algorithm is described. It utilizes boundary contrast for target detection and recognition. Boundary contrast is related to the location uncertainty at target boundary points. A polygon model is used for deriving target centroid location uncertainty caused by the boundary point location uncertainty. The significance of the work lies in the sound mathematical models used in deriving the relationship between contrast and location uncertainty for the boundary points and the relationship between boundary point location uncertainty and centroid location uncertainty. Experiment results show significantly improved performance in detection, recognition, and localization.* © *2000 SPIE and IS&T.* [S1017-9909(00)01602-0]

1 Introduction

In the ATR application, the sensory images are usually taken in uncontrolled environment where the conditions of the image forming process may vary greatly, e.g., in outdoor scenes with unknown numbers and types of targets, complicated terrain background, significant amount of clutter, all kinds of weather conditions, different season of the year and different time of the day, occlusion and camouflage of the targets, etc. The great complexity in the image forming process has made ATR a distinctively difficult problem where the input image is characterized by extremely high variability in the content and quality. This is particularly true with the forward-looking infrared (FLIR) imagery.

Compared with the first generation sensors, the second generation FLIR sensors give much better image with higher resolution, higher signal-to-noise ratio, and higher stability. The better image quality, especially increased number of pixels on targets, reveals more structure of the target appearance in the image, and allows the design of knowledge- and model-based algorithms. The situation is further improved by dramatically improving computing power which strongly supports sophisticated algorithms. There has been significant improvement in the observed performance of the ATR systems since the late 1980's.

Since the second generation FLIR sensor has approached the optical limits of sensor resolution,[1] improvement in ATR performance should be expected mainly from advancement in ATR algorithm instead of from further improved image quality. Model-based approaches to the ATR problem are the most promising in developing high performance systems.

The ability of low-level feature extraction algorithms is restricted by their limited information input from the local image neighborhood. High-level information can be very useful to overcome this restriction. This is typically done in the model-based approaches. In these approaches, either the target models are directly utilized in the low-level information extraction modules, or hard decisions about the information content in the imagery are avoided on low-level measurements until they are combined, via the use of target models, into medium- or high-level evidence. In doing this, weak but consistent evidence in the low quality imagery can be recovered with significantly improved reliability.

In this article, we describe an algorithm that computes a target saliency measure by properly utilizing the target boundary contrast evidence. The two major pieces in developing the algorithm are the development of the relationship between boundary contrast and boundary point location uncertainty and the development of the model for target centroid location uncertainty induced by the boundary point location uncertainty. In experiments, the developed algorithm shows improved detection performance and more impressive improvement in recognition performance. The localization performance is also good.

2 Literature Review and Motivations

2.1 *Traditional Algorithms*

The traditional FLIR ATR algorithms assume that targets in the scene appears as blob-like regions with roughly homogeneous gray scale values, significantly brighter or darker than the background.[2–4] For detection, the main effort is made to segment out these target regions from the rest of the scene. Shape recognition techniques are applied on the segmented targets to perform recognition. Bhanu and Holben[5] use pixel intensity and edge strength in a relaxation procedure for thresholding ship targets from FLIR images. The joint relaxation penalizes inconsistency and ambiguity in the segmentation process. Pham *et al.*[6] use grayscale morphological operations to enhance the targets to get better thresholding results. Ernisse *et al.*[7] use the

Paper 20-026 received Mar. 17, 2000; accepted for publication Mar. 23, 2000.
1017-9909/2000/$15.00 © 2000 SPIE and IS&T.

hit/miss filter for region of interest (ROI) identification and the difference of Gaussian filter for segmentation and feature extraction. Thresholds are determined by a genetic algorithm. Recognition is performed by a multilayer perceptron (MLP) neural network. Kitrosser[8] evaluated a region-growing based algorithm of its detection and recognition performance. Markham[9] gives a comparison of three classes of segmentation algorithms. Haralick and Shapiro[10] give a survey on classical segmentation techniques.

Due to the poor quality of the input imagery, the earlier stages of detection and segmentation fail to produce a satisfactory result. This causes misdetections, false alarms, only parts of the targets getting segmented, and background clutter getting segmented as part of the targets. ATR performance from the traditional algorithms is unsatisfactory.

2.2 *Model-Based Algorithms*

In the model-based approach, it is believed that appropriate models can be of significant help in the information recovery in the early stages of processing. Friedland and Rosenfeld[11] uses simulated annealing to optimize a cost function that is a linear combination of low and high level confidence of the shapes of interest. The low level component is model-free and tries to improve contour smoothness and edge sharpness. The high level computes the similarity of the contour to shapes of interest. The optimization procedure puts more emphasis on the low level in the earlier stage and shifts over to the high level when the confidence in a match increases.

Correlation based algorithms assume that some invariant features can be reliably extracted from the two-dimensional (2D) appearance of the targets of interest. Fazlollahi and Javidi[12] derive an optimal receiver based on known target shape and intensity but unknown contrast. Uenohara and Kanade[13] use the K–L decomposition for invariant feature extraction. Ben-Arie and Rao[14] use nonorthogonal expansion for template matching. Wu and Bhanu[15] uses the Gabor grid to represent targets.

There are many favorable reports of correlation based algorithms. However, when confronted with great variation in the target appearance, the invariance features required by these algorithms cannot be extracted robustly or even do not exist anymore. This causes significant performance degradation.

In the event of having high quality imagery which allows reliable prediction of the appearances of targets, complex three-dimensional (3D) target models and the heat emission and reflection properties of the material are utilized by Stevens and Beveridge.[16] They combined bore-sight color, range, and IR images in a target-pose optimization procedure. Target detection is accomplished using the color image only. Recognition and pose estimation are accomplished at the same time by using all three images. The IR image is used only in regard of the target boundary.

At the other end of the image quality spectrum, the 2D appearance of the target within its boundary is highly unpredictable. Various research results[17–20] suggest that the most reliable visual cue for the targets in this kind of imagery is the existence of relatively high contrast between the inside and outside of the silhouettes of the targets. This contrast is in general higher than the contrast to be expected in the background. This is reliable because of the difference in the materials, hence physical properties, between the targets and the natural background. Although the actual appearance of these materials in the FLIR imagery may vary greatly in different imaging conditions, noticeable difference should exist between them within each image. To make use of this, the aspect of and approximate distance to the targets are assumed to be known. Targets are modeled by their silhouettes.[17–19] This modeling takes the stand point that only reliable information obtained by means of reliable assumptions, although small in amount, should be used. Other sources of information are ignored due to their unreliability.

2.3 *Neural Networks in ATR*

As in all other fields of signal processing applications, neural networks receive a significant amount of attention in the ATR algorithm development. Hamilton and Kipp[18] feed eleven features into an MLP network to perform preliminary target detection. Ernisse *et al.*[7] and Franques and Kerr[21] use MLP networks for target recognition. Li *et al.*[22] report favorable recognition performance of a six-layer convolutional neural network in comparison with a number of other ATR algorithms on a single large-sized test data set. Roth[23] and Rogers[24] give surveys of the research efforts in neural nets in ATR. Neural networks in ATR is still a very active area of research

Popular as it is, the neural networks can mainly be regarded as nonlinear classifiers or as tools for nonparametric modeling of data. For this reason, on the abstract level, neural networks in ATR can largely be considered as a way of implementation, rather than an entirely different approach to the problem.

2.4 *Multisensor Data Fusion in ATR*

Multisensor data fusion[2,16,25] combines the sensory inputs from independent sensors for the same scene and let them complement each other in the recoverable information they provide. The commonly used sensory modalities for fusion in the ATR applications include FLIR, TV, LADAR, SAR, and MMW radar. Nandhakumar and Aggarwal[26] consider thermal physical properties of materials under direct sun light. The information in thermal and visual images is combined for target material identification. Stevens and Beveridge[16] present results on applying a 3D object recognition algorithm with pose estimation on the Fort Carson data set[27] of bore-sight range, thermal, and color imagery. In addition to the data from imaging sensors, various kinds of auxiliary information, such as the time of the day, weather conditions, geographical location, etc., are also subject to fusion.

Currently, the theory of multisensor data fusion has not yet been well developed. Further study into this area is needed before a theory is established. The theory needs to indicate how to optimally perform the fusion, and to predict the amount of performance improvement due to the use of fusion.

2.5 *Performance Characterization in ATR*

Haralick[28,29] suggests an approach that studies how the imperfection and perturbation in the ideal input will affect the output of the algorithm. Analytical error propagation[30] through computer vision algorithm modules is the main

tool used in this approach. Although a number of standard algorithm modules have been studied,[31] the performance evaluation of a general computer vision algorithm sequence in general real applications is still not practical.

Currently, most researchers report the performance of algorithms using the experimental result on some test data set,[6,8,16,22,32–34] in terms of some performance measures such as detection rate, false alarm rate, and correct classification rate, or some variations of these. Although the notion of these objective and quantitative measures are widely used, their exact definitions and the ways their values are computed, have not been standardized. We use the methodology proposed in Ref. 35 to setup a maximal constrained one-to-one correspondence between ground-truth and declared targets. Performance measures based on this correspondence are accurate and consistent.

2.6 *Motivations for This Work*

We are concerned with the most adverse application scenario, where thermal conditions of targets and the surrounding area change so much from scene to scene that virtually no assumption can be reliably made on the appearance inside the target boundary. The only reliable visual cue for the targets is the existence of some contrast between the inside and outside of the silhouettes of the targets. Boundary models for the targets are suitable in this situation.

In this situation, a robust FLIR ATR algorithm should try to infer target existence by accumulating the evidence of contrast along the target boundary. This introduces three immediate problems to be solved. The first problem is that targets are 3D objects and will have infinite numbers of different 2D silhouettes due to its relative pose to the imaging sensor. Practically, although the silhouettes of interest can be reduced to a finite number, the amount of computation for processing all silhouettes can be very heavy when the number of targets of interest is large. Some decision tree type algorithm[17] can be used to reduce the amount of computation by considering only a subset of silhouettes at each level. The computational load then seemed to be acceptable to the US military research labs.

The other two problems in designing boundary contrast based FLIR ATR algorithms are how the contrast along the boundary is to be combined into a single evidence measure for the target's existence, and how the contrast at each boundary point should be represented to facilitate the evidence accumulation. This article presents an algorithm which results from properly addressing these issues. The issue of optimal estimation of contrast concerns the optimal gradient operator design problem and is beyond the scope of this article.

Intuitively, some form of weighted average seems to be appropriate for the combination of boundary contrast into a single target saliency measure. The higher the contrast along the boundary, the more likely the target is present. The reason for a weighted average instead of a simple average, such as the arithmetic or geometric average, is that generally boundary points are not equally informative of the existence of the target. When only a subset of the boundary points is used to represent the target, different subsets of the same number of boundary points carry different amount of evidence for discriminating this target from other types of targets and from the background clutter. For example, a good polygon approximation algorithm can use a small number of points to fairly well represent the boundary of a compact 2D shape, while randomly picked points will not do as well. The shape is even worse represented if the chosen points are clustered at unimportant places on the boundary. This is because the contribution of the boundary points to the detection of the target significantly depend on each other. The relative importance of a particular boundary point should be decided by the entire target boundary. A mathematically sound way of modeling the relative importance of the boundary points is needed. This issue is addressed in Sec. 3.

It is an established principle in signal and image processing that prominent features for any detection problem should both have high signal-to-noise ratio (SNR) and be well localized in its domain. This is also true for the target boundary points. The boundary points of high-contrast targets in the FLIR imagery can be localized with high precision, while low contrast boundary sections will cause high uncertainty in the boundary point location along the segments. This will weaken the evidence for the target's presence. Notice that although we make direct use of the localization property of boundary points, boundary point detection is not carried out. Boundary point detection is highly prone to error due to the low image quality. The need and desire to avoid this type of hard decision on the noisy low level measurements are among the driving forces toward developing model based ATR algorithms.

When the location of the boundary points has uncertainty, a good measure of the overall uncertainty is the uncertainty in the target centroid location. When the majority of boundary points have high location uncertainty due to their low contrast, the centroid uncertainty will be high, indicating low evidence for the target's presence. On the other hand, when many boundary points have high contrast and therefore low location uncertainty, the centroid can have a low uncertainty, showing strong evidence for the target's presence. Therefore, using the uncertainty in the centroid is a good way of combining the boundary contrast information into a single target saliency measure.

We use polygons to model target shapes, where the polygon vertices are specified by the target boundary points. The centroid location is a function of the vertex locations. In Sec. 3, we derive the formulas for computing the centroid location uncertainty as a function of the vertex location uncertainty.

The relationship between contrast and location uncertainty at boundary points is established using the edge detection framework, where the edgel location is determined as the zero-crossing of the second directional derivative in the direction of the gradient. The random perturbation in the zero-crossing location estimate and in the gradient orientation estimate contribute to the uncertainty in the edgel location. In Sec. 4 we derive precisely how this uncertainty is influenced by the contrast.

3 Target Centroid Location Uncertainty

In using boundary contrast for target detection and recognition in the FLIR imagery, we try to combine the contrast evidence along the target boundary into a single saliency measure for the target's presence. In studying the relation-

ship between a polygon's centroid location and those of its vertices, we discover that the relative contribution of the vertices to the location uncertainty of the centroid, which can be considered as a measure of the geometric stability of the polygon, can be used to model the relative importance of the vertices. Since the location uncertainty of the target boundary points is closely related to the contrast across the target silhouette (see Sec. 4), the relative importance of the polygon vertices derived from the polygon centroid uncertainty model can be used in combining the contrast at the boundary points into a single saliency measure for the target's presence. The centroid uncertainty is an inverted saliency measure for the target's presence—the more salient a target is, the smaller centroid uncertainty it has.

3.1 *Relationship Between a Polygon's Vertices and its Centroid*

Let the silhouette of the target of interest be represented by a simple polygon (A simple polygon is one whose sides do not cross each other. Within this work, we are only interested in simple polygons.) in the 2D plane. It is parameterized by an ordered list of its vertices. The vertices of an N-side polygon are denoted by $P_1,\ldots,P_N$. Either clockwise or counter-clockwise ordering can be used. It does not affect the result in this work.

Let O be the origin of the coordinate system, and let the coordinates of the N vertices be $(x_1,y_1),\ldots,(x_N,y_N)$. Let $x=(x_1,\ldots,x_N)^T$ and $y=(y_1,\ldots,y_N)^T$ be the vectors of the x and y coordinates of the vertices. The superscript T denotes the matrix transposition. x and y are given by the target silhouette. The centroid of the target polygon is fully determined and is denoted by $Q(x,y)=[Q_x(x,y),Q_y(x,y)]^T$.

In finding the centroid of the polygon, we use triangles. Consider connecting each vertex with the origin using straight line segments. These N line segments and the N sides of the polygon form N triangles $\{\Delta OP_iP_{i+1}, i=1,\ldots,N\}$ where the notation of $P_{N+1}=P_1$ is used. Regardless of whether the polygon is convex or not and whether the origin is inside the polygon or is on the outside, the centroid and the signed area of triangle ΔOP_iP_{i+1} are $((x_i+x_{i+1})/3,(y_i+y_{i+1})/3)$ and $(x_iy_{i+1}-x_{i+1}y_i)/2$, respectively. The centroid of the polygon is computed as the weighted sum of the centroids of the triangles, each weighted by its signed area, divided by the sum of the signed area of all the triangles. Hence, the centroid of the polygon is found to be

$$Q_x(x,y)=\frac{K_x}{3S},\quad Q_y(x,y)=\frac{K_y}{3S}, \tag{1}$$

where

$$K_x=\sum_{i=1}^{N}(x_iy_{i+1}-x_{i+1}y_i)(x_i+x_{i+1}), \tag{2}$$

$$K_y=\sum_{i=1}^{N}(x_iy_{i+1}-x_{i+1}y_i)(y_i+y_{i+1}), \tag{3}$$

$$S=\sum_{i=1}^{N}(x_iy_{i+1}-x_{i+1}y_i). \tag{4}$$

Here the notation of $x_{N+1}=x_1$ and $y_{N+1}=y_1$ is used.

3.2 *Uncertainty in the Vertex Locations*

As will become clear in Sec. 4, the contrast at the target boundary is closely related to the location uncertainty of the boundary points. This location uncertainty is modeled here as a zero-mean additive perturbation on the polygon vertices. When the vertex P_i is affected by this perturbation, its perturbed location $\hat{P}_i=(\hat{x}_i,\hat{y}_i)^T$ is given by $\hat{P}_i=P_i+\Delta P_i$,

$$\begin{pmatrix}\hat{x}_i\\ \hat{y}_i\end{pmatrix}=\begin{pmatrix}x_i\\ y_i\end{pmatrix}+\begin{pmatrix}\Delta x_i\\ \Delta y_i\end{pmatrix}. \tag{5}$$

Let this perturbation on P_i be called the input perturbation. We assume that ΔP_i and ΔP_j are independent for $i\neq j$, but Δx_i and Δy_i can be correlated for $i=1,\ldots,N$.

Let $\Delta x=(\Delta x_1,\ldots,\Delta x_N)^T$ and $\Delta y=(\Delta y_1,\ldots,\Delta y_N)^T$. We have $\mathbf{E}(\Delta x)=\mathbf{E}(\Delta y)=0$ and

$$\mathbf{Cov}(\Delta x,\Delta x)=\mathrm{Diag}(\sigma_{x,1}^2,\sigma_{x,2}^2,\ldots,\sigma_{x,N}^2), \tag{6}$$

$$\mathbf{Cov}(\Delta y,\Delta y)=\mathrm{Diag}(\sigma_{y,1}^2,\sigma_{y,2}^2,\ldots,\sigma_{y,N}^2), \tag{7}$$

$$\mathbf{Cov}(\Delta x,\Delta y)=\mathbf{Cov}(\Delta y,\Delta x),$$

$$=\mathrm{Diag}(\sigma_{xy,1}^2,\sigma_{xy,2}^2,\ldots,\sigma_{xy,N}^2). \tag{8}$$

The diagonal elements of these matrices are specified by the boundary point location uncertainty model and will be given by the edgel location characterization in Sec. 4.

3.3 *Uncertainty in the Centroid Location*

Being a deterministic function of $\{P_i\}$, the centroid Q is affected by the perturbation on $\{P_i\}$. Let the perturbation be denoted by

$$\Delta Q=Q(x+\Delta x,y+\Delta y)-Q(x,y) \tag{9}$$

or, in terms of its elements,

$$\Delta Q_x=Q_x(x+\Delta x,y+\Delta y)-Q_x(x,y), \tag{10}$$

$$\Delta Q_y=Q_y(x+\Delta x,y+\Delta y)-Q_y(x,y). \tag{11}$$

Within this section, this perturbation is called the output perturbation. The quantity we are interested in is $\mathrm{Cov}(\Delta Q)$, namely the covariance of the output perturbation. This is a full measure for the target centroid location uncertainty, the trace of which is to be used as an inverted saliency measure for the target's presence.

In order to compute the centroid location uncertainty just defined, we need to express $\mathrm{Cov}(\Delta Q)$ in terms of known or computable quantities. Specifically, these are the

locations of the unperturbed vertices and the covariance of the input perturbation, which are given and computable, respectively.

Let $Q_x(x+\Delta x, y+\Delta y)$ be approximated by the first order terms in the Taylor series expansion.

$$Q_x(x+\Delta x, y+\Delta y) \approx Q_x(x,y) + A^T \cdot \Delta x + B^T \cdot \Delta y, \tag{12}$$

where $A = \partial Q_x(x,y)/\partial x$ and $B = \partial Q_x(x,y)/\partial y$ are $N \times 1$ real vectors, which can be computed using Eqs. (A1) and (A2). Then

$$\Delta Q_x \approx A^T \cdot \Delta x + B^T \cdot \Delta y, \tag{13}$$

$$\mathbf{E}(\Delta Q_x) \approx 0, \tag{14}$$

$$\mathbf{Var}(\Delta Q_x) \approx A^T \mathbf{Cov}(\Delta x, \Delta x) A + 2A^T \mathbf{Cov}(\Delta x, \Delta y) B + B^T \mathbf{Cov}(\Delta y, \Delta y) B. \tag{15}$$

Similarly, we obtain

$$\mathbf{E}(\Delta Q_y) \approx 0, \tag{16}$$

$$\mathbf{Var}(\Delta Q_y) \approx C^T \mathbf{Cov}(\Delta x, \Delta x) C + 2C^T \mathbf{Cov}(\Delta x, \Delta y) D + D^T \mathbf{Cov}(\Delta y, \Delta y) D, \tag{17}$$

$$\mathbf{Cov}(\Delta Q_x, \Delta Q_y) \approx A^T \mathbf{Cov}(\Delta x, \Delta x) C + A^T \mathbf{Cov}(\Delta x, \Delta y) D + B^T \mathbf{Cov}(\Delta y, \Delta x) C + B^T \mathbf{Cov}(\Delta y, \Delta y) D, \tag{18}$$

where $C = \partial Q_y(x,y)/\partial x$ and $D = \partial Q_y(x,y)/\partial y$ are $N \times 1$ real vector, which can be computed using Eqs. (A3) and (A4). With these values computed, we obtain approximately

$$\mathbf{E}(\Delta Q) = \mathbf{0}, \tag{19}$$

$$\mathbf{Cov}(\Delta Q, \Delta Q) = \begin{pmatrix} \mathbf{Var}(\Delta Q_x) & \mathbf{Cov}(\Delta Q_x, \Delta Q_y) \\ \mathbf{Cov}(\Delta Q_y, \Delta Q_x) & \mathbf{Var}(\Delta Q_y) \end{pmatrix}, \tag{20}$$

where all terms have been obtained in terms of the locations of the unperturbed vertices and the covariance of the input perturbation. The validity of this derivation result has been verified in a validation experiment. In that experiment, random perturbations are added to a test polygon and the sample covariance matrix of the perturbed centroid location is compared with the one computed from the above equation. The comparison is done by means of a statistical test based on the multivariate normal distribution.[36,37] The details of the validation experiment are omitted here.

The 2×2 covariance matrix $\mathbf{Cov}(\Delta Q, \Delta Q)$ is a full characterization of the location uncertainty of the centroid. It describes how the likely position of the centroid is spread out in the 2D domain. A proper scalar measure of the uncertainty is the trace of the matrix $\mathbf{Var}(\Delta Q_x) + \mathbf{Var}(\Delta Q_y)$, i.e., the sum of the two diagonal elements.

In the ATR algorithm, the locations of the unperturbed vertices are given by the locations of the target silhouette boundary points relative to the target centroid. The covariance of the input perturbation is determined by the results obtained in Sec. 4. The trace of the output covariance matrix is inversely related to the saliency of the target's presence. A strong target appearance will result in a small trace of the centroid covariance.

4 Uncertainty in Edgel Location Measurement

Localization is an important aspect of edge detector performance. Our emphasis is on studying the relationship between the contrast and the edgel location uncertainty so that a better way of combining target boundary contrast can be done by means of location covariance propagation.

Local neighborhood gradient based edge operators, e.g., Refs. 38 and 39, search the maximum in the derivative of the image intensity surface along the direction of the gradient, or, equivalently, the zero-crossing in the second directional derivative in that direction. Although it has been argued that the gradient orientation estimate is usually quite noisy,[40,41] the analysis of the localization error of edge detectors has yet remained a one-dimensional formulation. The image intensity profile along the cutting plane in the direction of the gradient estimate is studied and the radial localization error along that direction is examined. The error in the orientation of the cutting plane has not been incorporated in the error analysis of edge localization until recently. Marimont and Rubner[42] are the first to combine orientation and radial localization distributions to compute the edge location probability for edge pixel detection.

To discriminate single edge pixels from edge features which usually consist of multiple edge pixels, we use the term edgel to denote edge elements which are just single edge pixels. The orientation of the edgel is the gradient orientation, and the location of the edgel is the zero-crossing in the second directional derivative.

In this section, we incorporate the gradient orientation estimation error into the study of the edgel location estimation. The edgel localization error is then a two-dimensional vector as opposed to the scalar used in previous analyses. This results in a more accurate characterization of the relationship between contrast and location uncertainty.

The derivations in this section closely parallel in the line of reasoning to that of Marimont and Rubner.[42] We became aware of their work only after we finished our derivation and experiments. Also, our work is different from Ref. 42 in that we do not use the resulting edgel location uncertainty in low level edge pixel detection. As has been stated, this is highly prone to error due to the noisy nature of local neighborhood operations. Instead, we explicitly compute the edgel location uncertainty as indexed by the signal-to-noise ratio (SNR). The location uncertainty along the target boundary is combined in the polygon centroid uncertainty model, hence we avoid making hard decisions at the pixel level. The edgel location distribution plays more of the role of the medium for conveying the contrast into another form which is more appropriate for evidence accumulation.

4.1 *Background*

Here we restate some facts concerning edgel measurements in gradient based methods. The underlying assumption is that the noise can be modeled as zero-mean white additive Gaussian.

- The orientation estimate $\hat{\theta}$ from the full angular range arc tangent function (atan2) follows the von Mises distribution.[43]

$$f_\Theta(\theta|\theta_0,\kappa)=\frac{1}{2\pi I_0(\kappa)}e^{\kappa\cos(\theta-\theta_0)}$$
$$0<\theta_0\leq 2\pi,\kappa>0,0\leq\theta<2\pi, \tag{21}$$

where θ_0 is the true underlying gradient orientation, κ is the squared output SNR, and $I_0(\cdot)$ is the modified Bessel function of the first kind and order zero.[44] This distribution is symmetric about θ_0. The concentration parameter κ controls how much the distribution is concentrated around θ_0. The bigger κ is, the more concentrated the distribution is. For large κ, the von Mises distribution can be well approximated by the normal distribution with mean θ_0 and variance $1/\kappa$. For any given gradient operator, κ is given by

$$\kappa=\kappa_0\Omega_{\text{in}}^2, \tag{22}$$

where Ω_{in} is the input SNR defined as ratio of the true step edge contrast over the standard deviation of the noise. κ_0 is a constant determined by the operator itself

$$\kappa_0=\frac{d_x^2+d_y^2}{\lambda^2}, \tag{23}$$

where λ^2 is the sum of squared elements of the gradient operator kernel; d_x and d_y are the two elements of the gradient computed by the operator on the unit-contrast step edge. In the 5×5 neighborhood with the true gradient orientation of 0°, κ_0 is 1.6 for the standard cubic facet model estimate and 3.8 for the integrated directional derivative gradient operator (IDDGO).[40,41,45]

- The edgel location $\hat{v}$ is determined in each neighborhood for the center pixel in that neighborhood. It is determined to be the zero-crossing of the second directional derivative in the direction of the gradient. The derivatives are computed by the IDDGO which is based on the cubic facet model.[38,45] When the image is noisy, even along the true gradient orientation, $\hat{v}$ is in general different from the true edgel location. The localization error from reasonable detectors has a distribution that is symmetric about 0.
- Using the cubic facet model, $\hat{v}$ is determined to be the inflection point of the cubic polynomial along the direction of the gradient orientation.

$$\hat{v}=-\frac{\hat{c}_2}{3\hat{c}_3}, \tag{24}$$

where $\hat{c}_2$and $\hat{c}_3$ are the estimated coefficients for the second and third order terms of the cubic polynomial. Due to the observation noise,

$$\hat{c}_2\sim n(c_2^*\Omega_{\text{in}}\sigma,\kappa_2^2\sigma^2),\quad \hat{c}_3\sim n(c_3^*\Omega_{\text{in}}\sigma,\kappa_3^2\sigma^2),$$

where c_2^* and c_3^* are the true value of the cubic polynomial coefficients; κ_2 and κ_3 are constants from the cubic facet model; and σ^2 is the variance of the observation noise. For the unit-contrast step edge in the 5×5 neighborhood with the true gradient orientation of 0°, the involved constant are

$$c_2^*=0,\quad c_3^*=-\tfrac{1}{12},\quad \kappa_2=0.1196,\quad \kappa_3=0.1179. \tag{25}$$

We can easily derive that

$$\hat{v}=-\frac{\kappa_2}{3\kappa_3}Z, \tag{26}$$

where Z is distributed as

$$\frac{n\left(\frac{c_2^*}{\kappa_2}\Omega_{\text{in}},1\right)}{n\left(\frac{c_3^*}{\kappa_3}\Omega_{\text{in}},1\right)}. \tag{27}$$

The term *normal quotient* is used to denote the family of distributions for the ratio of two independent normal random variables with unit variance. The probability density function (pdf) for this family has the form[46]

$$f_z(z|\mu_1,\mu_2)=\frac{1}{\pi}\frac{1}{z^2+1}\exp\left(-\frac{1}{2}\cdot\frac{(\mu_1-z\mu_2)^2}{z^2+1}\right)\times\left\{\exp\left(-\frac{(z\mu_1+\mu_2)^2}{2(z^2+1)}\right)+\sqrt{\pi}\frac{z\mu_1+\mu_2}{\sqrt{2(z^2+1)}}\operatorname{erf}\left(\frac{z\mu_1+\mu_2}{\sqrt{2(z^2+1)}}\right)\right\}. \tag{28}$$

For the step edge, $\mu_1=0$, $\mu_2=(c_3^*/\kappa_3)\Omega_{\text{in}}$, and the distribution is symmetric about 0. Studies have shown that the IDDGO gives superior estimate for the gradient orientation.[40,41,47,48] Hence it is used for gradient estimation.

4.2 *Edgel Location Estimation*

The edgel location estimation scenario is illustrated in Fig. 1. The polar coordinate system is used. The true and unobserved underlying gradient orientation is θ_0. The step edge discontinuity occurs at a radial distance of v_0 from the ori-

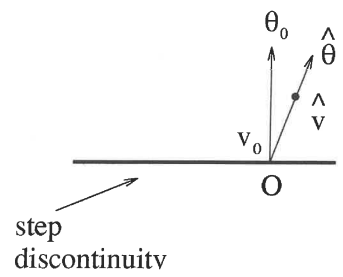
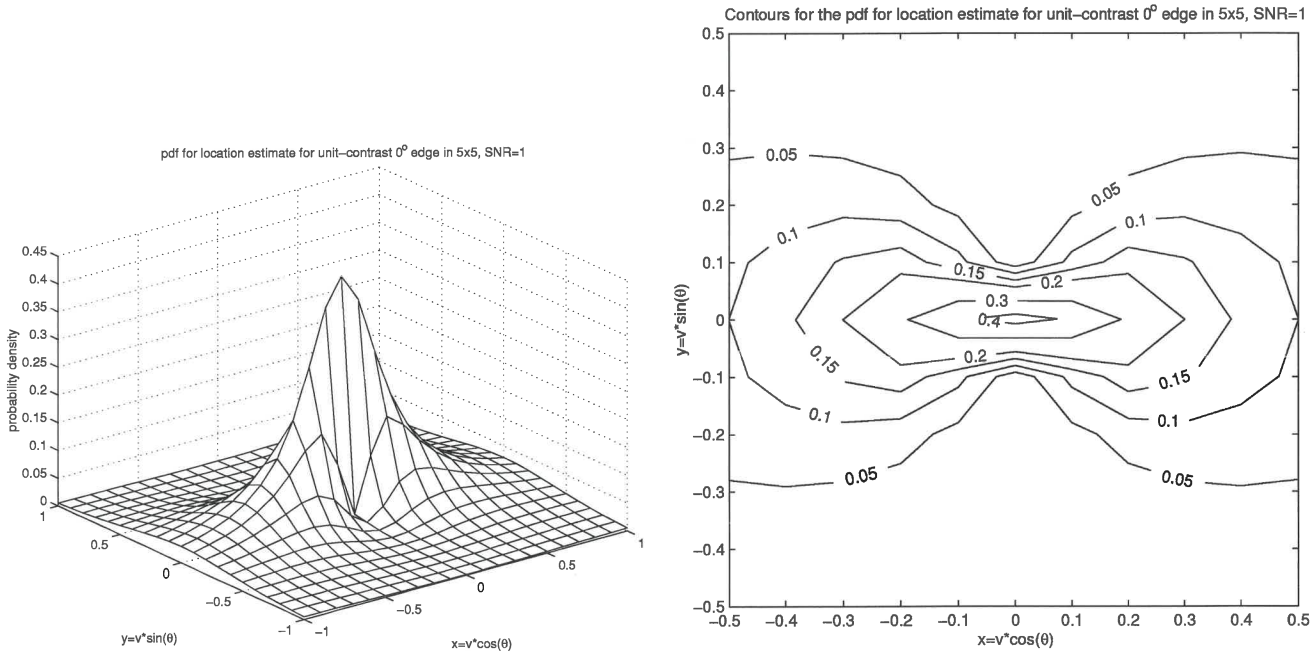


Fig. 1 Step edge detection scenario.



(a) pdf for 2-D location distribution

(b) level contour

Fig. 2 The pdf of the 2D Cartesian location. Underlying image data is the unit-contrast step edge in the 5×5 neighborhood, with 0° gradient orientation, and $\Omega_{in}=1$.

gin along the direction θ_0. The true input SNR is Ω_0. The chosen gradient operator gives the estimate for these as $\hat{\theta}$, $\hat{v}$, and $\hat{\Omega}$, respectively. The estimated edgel location is determined by $\hat{\theta}$ and $\hat{v}$. These estimates are random variables. Their statistical properties characterize the relationship between the edgel location uncertainty and the step edge SNR. This relationship is to be used along with the polygon centroid uncertainty model to combine target boundary contrast into a single target saliency measure.

Using the definition of conditional probability density, we have

$$f(\hat{\theta},\hat{v}|\,\theta_0,v_0)=f(\hat{v}|\,\theta_0,v_0,\hat{\theta})f(\hat{\theta}|\,\theta_0,v_0). \tag{29}$$

Using the IDDGO for gradient estimation and the zero-crossing of the second directional derivative along the direction of the gradient for edgel location, the two terms on the right-hand side of the equation can be approximated by the pdfs $fz(\cdot)$ and $f_\Theta(\cdot)$, respectively. $\hat{\Omega}$ obtained from the gradient and local noise estimate is used to substitute for Ω_{in}. The probability density function of the estimate for the edgel position in the 2D Cartesian space can be obtained from the above joint density function via standard random variable transformation.[49]

The pdf for the 2D Cartesian location estimate from the IDDGO based estimator is shown in Fig. 2. The underlying image data is a straight unit-contrast step edge in the 5×5 neighborhood, with $\theta_0=0°$, $\hat{v}_0=0$, and with an input SNR of 1.

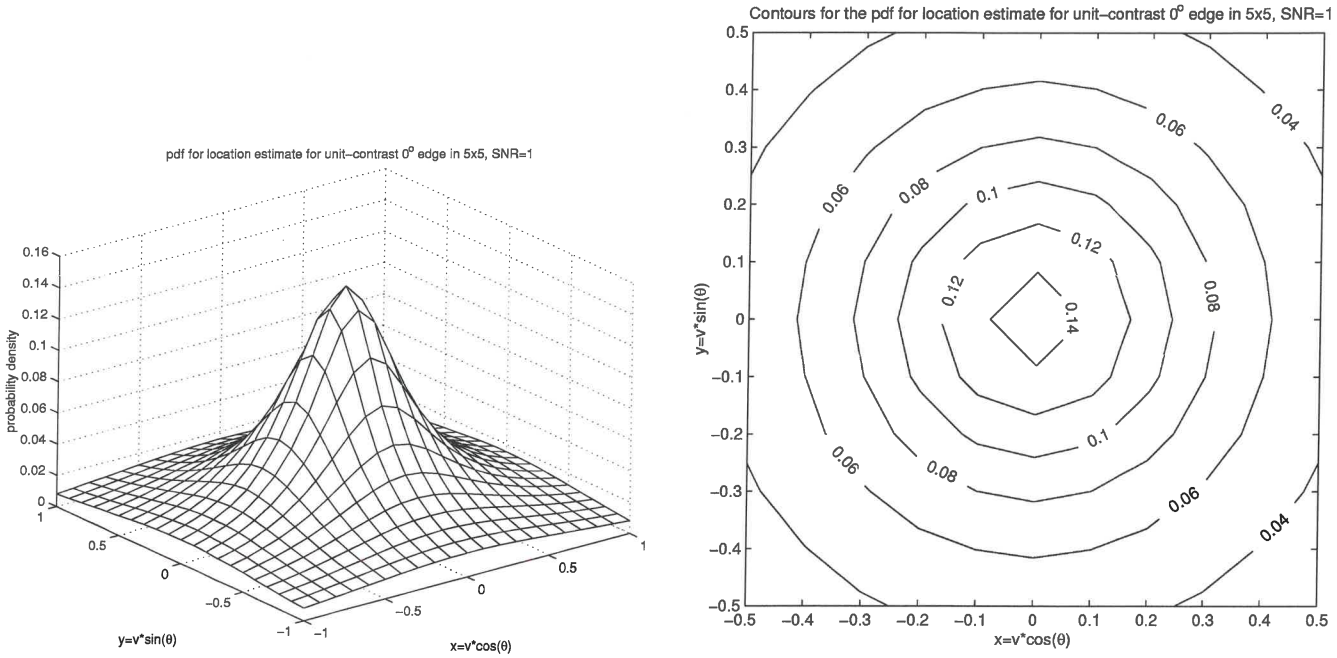
It can be seen from Fig. 2 that the distribution of the edgel location estimate is symmetric about the origin, which is the true edgel location. The distribution is not rotationally symmetric and is more concentrated along the true orientation of the gradient. As the SNR increases, the concentration is increased and the entire distribution shrinks toward the origin. For low SNR, as is the case shown in the figures, the variation in the orientation estimate is very large.

We note that the level contour in Fig. 2 has two very significant concavities at the direction perpendicular to the gradient orientation. This is due to the fact that the orientation estimate is more concentrated around the true orientation.

Since the 2D Cartesian location distribution is center-symmetric about the true edge location, the Cartesian coordinates computed from $\hat{\theta}$ and $\hat{v}$ is an unbiased estimate. Analytical derivation of the covariance matrix of this distribution is difficult, but the numerical values can be computed easily from the spatial spread of the pdf. The trace of this covariance matrix is a good measure of the edgel localization performance. When the SNR increases, the spread of the pdf shrinks and results in better edge localization performance. When there is no underlying edge structure, i.e., the SNR reduces to 0, there is no preference of any orientation for the nonexistent edgel. In this case, the 1D location estimate follows the Cauchy distribution. The pdf of the 2D Cartesian edgel location estimate is then rotationally symmetric. The pdf for the 2D location estimate in this case is shown in Fig. 3.

4.3 *Tabulating Edgel Location Uncertainty*

At a given input SNR, the pdf of the edgel location is used in the numerical evaluation of the edgel location covariance matrix. The edgel location uncertainty can be characterized by the trace or the larger of the two eigenvalues of the covariance matrix. These as functions of the input SNR are shown in Fig. 4.



(a) pdf for 2-D location distribution

(b) level contour

Fig. 3 The pdf of the 2D Cartesian location. Underlying image data is unit variance white Gaussian noise ($\Omega_{in}=0$).

It is interesting to notice that, in the very low SNR range, as the input SNR grows, the larger eigenvalue does not monotonically decrease as one might have expected. Rather, it increases first! However, there is an explanation to this behavior. The radial location estimation quality is insensitive to the initial small amount of increase in the SNR. The orientation estimate, however, is more responsive to the increased SNR and starts to concentrate toward the true orientation, while the radial location estimate essentially remains at the same precision level. The probability of the edgel is increased for locations around the true gradient orientation but with long radial distance from the true edgel location. Hence the increase in the uncertainty along the true gradient orientation is observed.

On the other hand, the decrease in the probability for locations around the direction perpendicular to the true gradient orientation is faster. The magnitude of the decrease in the location uncertainty caused by this is more than the increase in the gradient direction. As a result, the trace of the covariance matrix monotonically decreases with the increase in SNR.

For the 0° step edge, the correlation between the x and y elements of the edgel location is zero. For each SNR value of interest, we numerically compute the covariance matrix and tabulate the variance in the x and y directions.

Notice that, since the Cauchy distribution has infinite moments, the covariance matrix for the 2D location does not exist for $\Omega_{in}=0$. Practically, although the distribution is long tailed, most of the probability mass are concentrated in a finite neighborhood. The distribution truncated in this finite neighborhood is used for the covariance matrix evaluation.

The tabulated covariance matrices are used for computing the quantities $\mathbf{Cov}(\Delta x,\Delta x)$, $\mathbf{Cov}(\Delta y,\Delta y)$, and $\mathbf{Cov}(\Delta x,\Delta y)$ in Eqs. (6)–(8). The orientation θ and SNR at each boundary point are first estimated by the IDDGO. The SNR is used to index into the table made here and retrieve the two diagonal elements of the covariance matrix for that SNR level. This diagonal matrix is then pre- and postmultiplied by the rotation matrix for the orientation θ. The resulting matrix is the covariance matrix for the boundary point in question, and is used subsequently in the computation of the target centroid uncertainty.

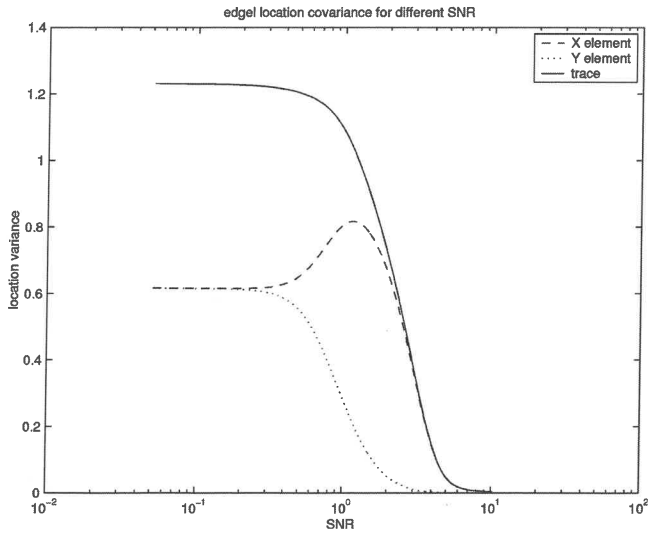


Fig. 4 Uncertainty of the edgel location estimate. Orientation estimated by the 5×5 IDDGO. Solid-trace, dashed-larger eigenvalue, dotted-smaller eigenvalue.

5 Experiment on FLIR ATR

This section describes the experiment conducted to evaluate the performance of the centroid uncertainty based FLIR ATR algorithm. We start by giving some implementation details of the algorithm that are not covered by previous sections, which mainly deal with the idea and theory behind the centroid uncertainty based target saliency measure. Specifically, in addition to an outline of the algorithm sequence, we briefly discuss the normalization of the centroid uncertainty measure, a generic thresholding procedure, and ROC operating point selection.

A few competing algorithms are then briefly described. The performance of these algorithms is compared with the proposed algorithm on the same set of test data. For the test data, it is unfortunate for us not to be able to gain access to some large set of real image data suitable for this study[22,32] because of the military classification on that data. We describe the generation of synthetic scenes in Sec. 5.3. The observed performance of the tested algorithms is reported in Sec. 5.4 along with some discussions.

5.1 *Centroid Covariance Based FLIR ATR Algorithm*

5.1.1 *Normalized centroid trace as inverted target saliency measure*

Without repeating the details in edgel location covariance estimation and polygon vertex-to-centroid covariance propagation, the proposed FLIR ATR algorithm works as follows. The IDDGO (we used the 5×5 neighborhood size in our experiments) is first applied to the input FLIR image. The gradient orientation and the edgel contrast along that orientation are computed for all pixels. For each target silhouette of interest, we compute its saliency measure at each pixel location in the image where the centroid of the target silhouette can be placed.

At each such location, we first compute the contrast across the target boundary for each point on the target silhouette. This is done by projecting the edgel contrast estimate along the local gradient direction onto the direction perpendicular to the local segment of the target silhouette. (In deciding the orientation of the local segment of the target silhouette, we use a least-squares straight line fit for each overlapping five-point segments of the target boundary.) The contrast across the target boundary is used to index into the precomputed table of edgel location covariance matrix obtained in Sec. 4.3. The retrieved covariance matrices are first pre- and postmultiplied by the rotation matrix determined by the direction perpendicular to the local boundary segment. They are then used in the formula for computing the target centroid covariance matrix. The trace of this final matrix is normalized by a constant determined for each target silhouette. The constant is the trace of the centroid covariance matrix computed for the same target shape with zero-contrast across the boundary for all boundary points. It is the centroid uncertainty expected for a target silhouette placed in a pure noise background. The normalization is necessary for bringing the centroid uncertainty for different targets into the same range of the scale. It reduces the improperly biased preference of the unnormalized centroid uncertainty toward certain target shapes over others.

5.1.2 *Sequential thresholding procedure*

A target ID-location pair is used to represent a target candidate if the normalized uncertainty value computed for the said target and location is smaller than some specified threshold. A list of target candidates are obtained by examining the normalized centroid uncertainty of every target at every pixel location. Since we do not expect targets to overlap each other very significantly, some of the candidates are dropped from the list when the amount of overlap is too much. In doing this, we first sort the list of candidates in the ascending order of the normalized centroid uncertainty. The first in the sorted list is declared to be the first detected target and taken away from the list. Any target candidate whose centroid falls within the bounding box of the first target is deemed invalid and dropped from the candidate list. Any candidate whose bounding box covers the centroid of any declared target is also invalid and dropped from the list. After this is done, the first in the remaining list is declared as the second detected target. The procedure then repeats until the list becomes empty.

The effect of this thresholding procedure is that all declared targets do not cover each other's centroid. A fairly large amount of target overlapping is still allowed, to the extent that the centroid of any one is not covered by the bounding box of another. For a target with strong appearance, there will be only one target declared instead of multiple declarations of the same target at several locations around the true target location.

5.1.3 *Operating point selection*

All ATR algorithms have tuning parameters. Varying these parameters will usually give varied trade-offs between detection and false alarm rates. In characterizing the performance of ATR algorithms, it is customary to use the receiver operating characteristic (ROC) curve which gives the detection versus false alarm rates over some operation range of interest. In practical applications, an operating point on the curve needs to be chosen which reflects a particular compromise between detection and false alarm rates.

In this experiment, we use a linear cost function for operating point selection.

$$C=N_{MD}+kN_{FA}\,, \tag{30}$$

where k is a specified constant to balance the cost of a misdetection and that of a false alarm. N_{MD} and N_{FA} are the numbers of misdetections and false alarm, respectively. It is a common practice in the ATR application to assign more cost to a misdetection than to a false alarm. One reason is that a misdetected enemy target can potentially cause very serious damage. Another reason is that ATR systems often consist of cascaded subsystems. Subsequent processing modules incorporating other/fused information are likely to reduce the number of false alarms, while misdetected targets in early stages are very difficult to recover. Following this practice, in the experiments conducted here, we arbitrarily regard a misdetection as being twice as harmful as a false alarm. According to this, k is set to 0.5.

5.2 *Competing Algorithms*

This section gives a brief description of some FLIR ATR algorithms whose performance are also observed in the experiments and compared to that of the centroid uncertainty based algorithm.

Different target saliency measures are computed by these algorithms. The sequential thresholding procedure described in Sec. 5.1.2 is used to produce the final target declarations. Multiple thresholds are used for each algorithm to obtain the performance over the entire operating range. Notice that in the thresholding procedure, the candidate list is sorted in the descending order for the algorithms of Secs. 5.2.1, 5.2.2, and 5.2.3.

5.2.1 *The Maryland algorithm*

Der and Chellappa[19,50] of the Center for Automation Research at the University of Maryland proposed a model-based FLIR ATR algorithm for exactly the same application scenarios considered in this study. Targets are modeled by their 2D silhouettes. The image local contrast is measured by ''probes'' which are the differences in the intensity values of pairs of pixels at certain orientations and variable distances. Sample distributions of the probes for the background hypothesis are estimated locally in the image. The uniform distribution is assumed for the foreground hypothesis. A generalized likelihood ratio test is setup using those distribution models.

To increase the robustness of the algorithm, the probe values need to be thresholded at proper levels and the target saliency measure is reduced to the number of probes along the boundary that exceed the threshold. Instead of an arbitrarily chosen threshold, an adaptive procedure is used to identify the threshold yielding the smallest false alarm rate, which is determined using a binomial model. At each possible target location, this smallest achievable false alarm rate is computed for all targets. The target with the smallest false alarm rate wins the competition. Further, if its saliency measure is above the threshold, it is declared to be a target candidate.

5.2.2 *The likelihood ratio algorithm*

A variant of the Maryland algorithm was proposed[51] where no threshold on the probe value is applied. The probe values are used directly in a generalized likelihood ratio test for target detection and recognition. On the MURI data set, this algorithm performed slightly better than the Maryland algorithm.[51]

5.2.3 *Matched filter for FLIR ATR*

When the appearance of targets can be characterized by random segments of regions with smoothly varying intensity values, the matched filter idea for ATR is not likely to work well. This is because there is no single reliable appearance of the target to match. However, we still apply the simple matched filter algorithm on the test data set. On one hand, its performance can be used as a rough standard of what can be achieved by the class of simplistic general purpose detection algorithms. On the other hand, from its performance we can make a general judgment of the variation in the appearance of the targets in the test images, and hence have some idea of the difficulty level of the test data set.

In the simplest form, the matched filter for a particular target has two values α and β for the pixel positions in the inside and outside of the target boundary, respectively. Let the total rectangular area for the matched filter be denoted by S, and the areas inside and outside of the target boundary be denoted by S_{in} and S_{out}, respectively, with $S=S_{\text{in}}+S_{\text{out}}$. The matched filter is determined by

$$\alpha=\sqrt{\frac{S_{\text{out}}}{SS_{\text{in}}}}, \tag{31}$$

$$\beta=-\sqrt{\frac{S_{\text{in}}}{SS_{\text{out}}}}. \tag{32}$$

The reason for this particular choice of α and β is for the matched filter to have zero mean (so that the output is insensitive to the constant bias of the input) and unit energy.

$$S_{\text{in}}\alpha+S_{\text{out}}\beta=\sqrt{\frac{S_{\text{in}}S_{\text{out}}}{S}}-\sqrt{\frac{S_{\text{in}}S_{\text{out}}}{S}}=0,$$

$$S_{\text{in}}\alpha^2+S_{\text{out}}\beta^2=\frac{S_{\text{out}}}{S}+\frac{S_{\text{in}}}{S}=1.$$

The matched filters obtained in this way is most effective for detecting targets whose inside area and immediate background are rather homogeneous looking with some difference between the average gray values of the two.

Target saliency measures are computed by convolving the matched filters with the image data. The larger output value is, the stronger evidence there is for the corresponding target. The sequential thresholding procedure with the descending order for the candidate list is used to determine the final target declarations.

5.3 *Simulated FLIR Scene Generation*

To aid the development of algorithms, quite some effort has been made to provide real FLIR data to the research community, e.g., Ref. 27 and some others available from the Center for Imaging Science (http://www.cis.jhu.edu) at the Johns Hopkins University. However, due to the practical difficulties involved in gathering and supplying the data, the target types, imaging conditions, and scene complexity are very limited. Only in rare cases is there precise ground-truth information about the targets in the scene. The situation is made worse by the restriction on the accessibility of the data due to the military classification of the data sets. For example, we failed to get access to the data set used in Li *et al.*[22] and Chan and Nasrabadi.[32]

Nonetheless, large data sets with ground truth is essential in assessing the performance of computer vision algorithms. Due to the difficulty in obtaining such data sets, synthetic images are often used in studies.[34,52,53] We also take this alternative and generate simulated FLIR scenes by planting random target appearances into real FLIR scenes. It has the advantage of extremely low cost and readily available ground-truth information from the scene genera-

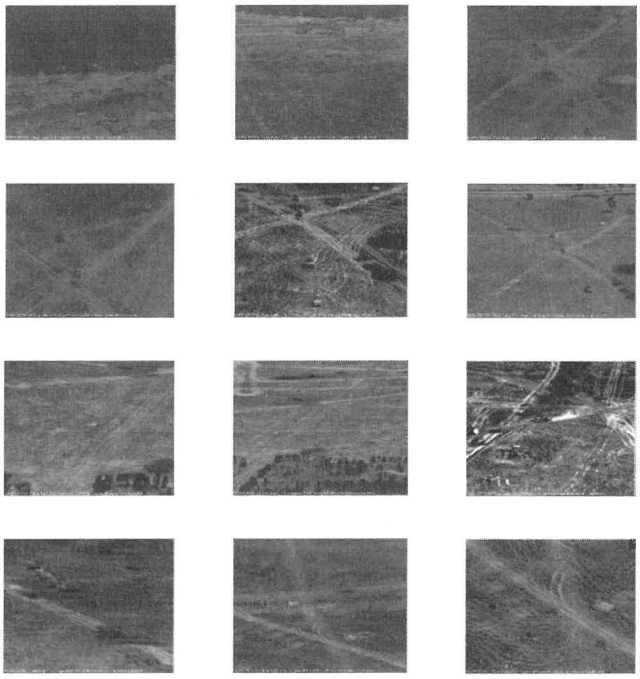


Fig. 5 Images in the Chinalake data set used as background scenes in the FLIR scene generation.

tion process. In what follows, we describe the procedure by which simulated FLIR scenes are generated.

5.3.1 *Scene background*

The FLIR scene generation procedure operates by choosing a background scene from a number of alternative real FLIR scenes and planting into that scene a random number of targets with random appearances at randomly picked locations. The ''land images'' in the NAWC Chinalake Presentation data set (http://www.cis.jhu.edu) contains seventeen FLIR scenes. These images contain our door scenes. The sizes of the images are around 512×400. The twelve images in this data set that are used in the scene generation are shown in Fig. 5.

5.3.2 *Insertion of targets*

From a library of target shapes, the procedure randomly picks ten to fifteen targets, simulates their intensity values, and selects locations in the background scene to place them so that their bounding boxes do not overlap each other.

The silhouettes of the thirty-six targets in the MURI FLIR data set are shown in Fig. 6. These are the targets to be planted into the background scenes. The target signature, or appearance, has to be decided. It is too simplistic and unacceptable to use a constant value for the entire interior of the target silhouette.

The best commercial software for target signature simulation is the physically reasonable infrared signature model (PRISM) running on the SGI workstations. The license for using this software is issued by the ThermoAnalytics, Inc. (http://www.thermoanalytics.com) under the authorization of the US Army Tank-automotive & Armaments Command (TACOM). We did not have access to this software. Instead, we use a procedure that randomly partitions the interior of the target shapes into unions of regions and assigns random intensity values for each of the regions.

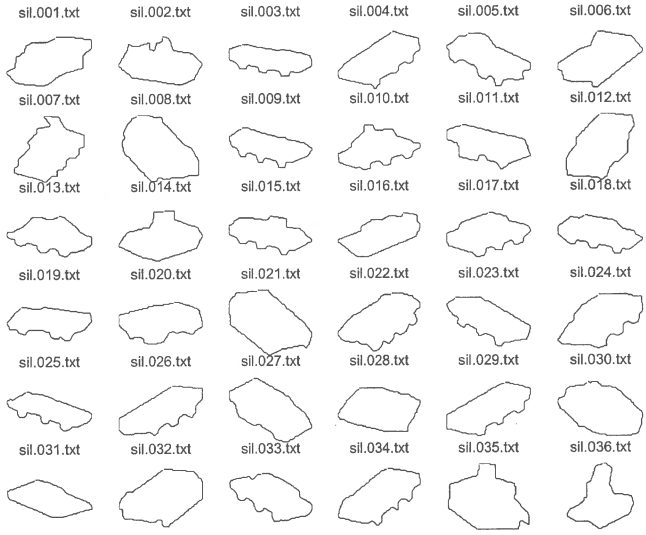


Fig. 6 Target silhouettes in the MURI FLIR data set.

It has been pointed out that in real FLIR scenes, the interior of the target usually will be divided into a number of regions of relatively smoothly varying intensity.[8] The partitioning of the target shape into the regions, however, is highly dependent on the thermal state of the target, and hence is highly variable. We generate random target appearance for each target instance by giving it a random partition of its interior region, and assigning a random intensity value with additive white Gaussian noise for each of the regions in the partition.

The random partition of the target region makes use of a segmented image of another FLIR scene. Shown in Fig. 7 is the boundary map of the regions obtained from an image in the MURI FLIR data set segmented by the region grow algorithm.[45] When a chosen target silhouette is placed at some location on this boundary map, the inside area of the silhouette is partitioned by the region boundaries. The placement of the silhouette on the boundary map is random. Hence we obtain a random partition of the area inside the target silhouette.

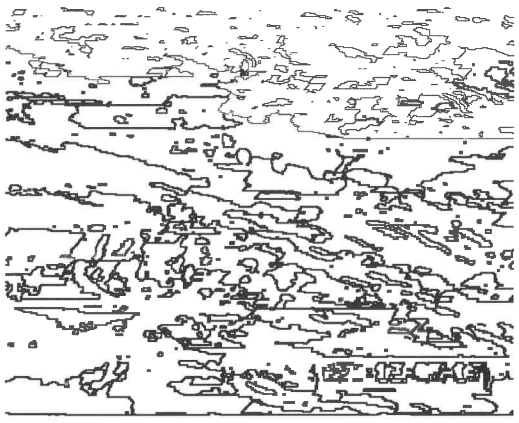
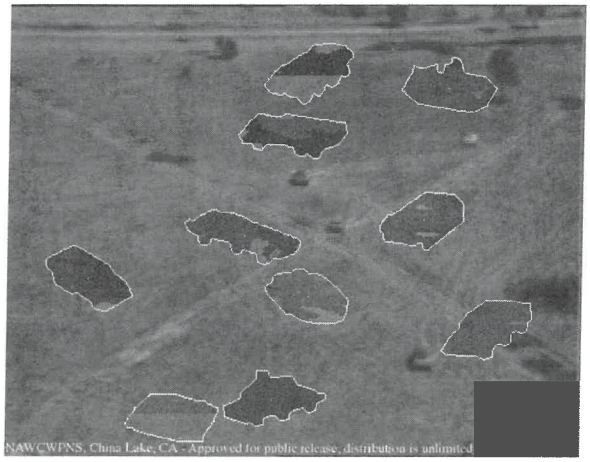


Fig. 7 Boundary of the segmented regions in the MURI FLIR image used for the random partition of target regions.

(a) simulated scene.



(b) with silhouttes overlaid.

Fig. 8 Example of simulated FLIR scene.

For each region in the partition of the target shape, we assign an intensity value for each pixel location in that region. The intensity value for each region is a white Gaussian random variable with a selected mean value and standard deviation. The mean value for each of the regions in the target shape is chosen independently. It is picked uniformly from an adaptively specified range $[a+b\cdot 10, a+b\cdot 50]$, where a is the average value of the background scene image within that region, and b is equally likely to take the values 1 and -1. This roughly gives a contrast on the order of 10–50 between the target and its immediate background. The standard deviation in the intensity value is chosen to be 10. As a result, the simulated target instances roughly have a signal-to-noise ratio of about 1–5.

Figure 8 shows an example of the simulated scene. In this scene, 10 targets are planted into the 6th image in Fig. 5. Because of the low contrast in the scene, some targets are quite hard to see even to the human eyes. To make it easier for the reader to study the image, another image is provided with the target silhouettes outlined. Notice that the inside area of many of the targets are broken into pieces of irregular shaped regions of homogeneous intensity values.

There are a total of 625 target instances in the 50 images generated for this experiment.

5.4 *Algorithm Performance and Discussions*

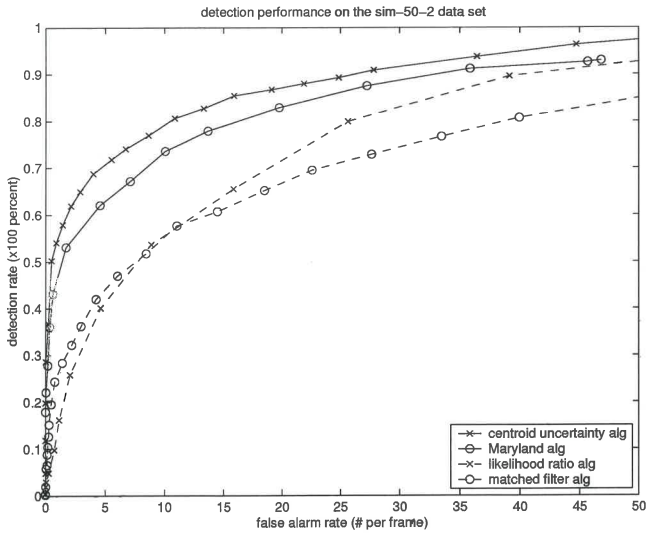
Each algorithm described in Sec. 5.2 is applied independently to the test data set described in Sec. 5.3. The ATR performance evaluation procedure described in Ref. 35 is carried out to compare the output with the ground truth.

In this procedure, a constrained maximal one-to-one correspondence between the ground truth and declared targets is established. A radial distance of 25 pixels is used in searching for the correspondence. From this correspondence, we determine the numbers of detected and misdetected ground-truth targets, correct declarations, false alarms, and the targets correctly recognized. The detection rate is the percentage of detected ground-truth targets among all 625 ground-truth targets. The false alarm rate is computed as the total number of false alarms divided by the number of frames (50). The unconditional recognition rate is the percentage of correctly recognized targets among all ground-truth targets. The conditional recognition rate is the percentage of correctly recognized targets among only the targets that are correctly detected. The root-mean-square (rms) error in the target centroid location for the correctly recognized targets is also computed to characterize the localization performance.

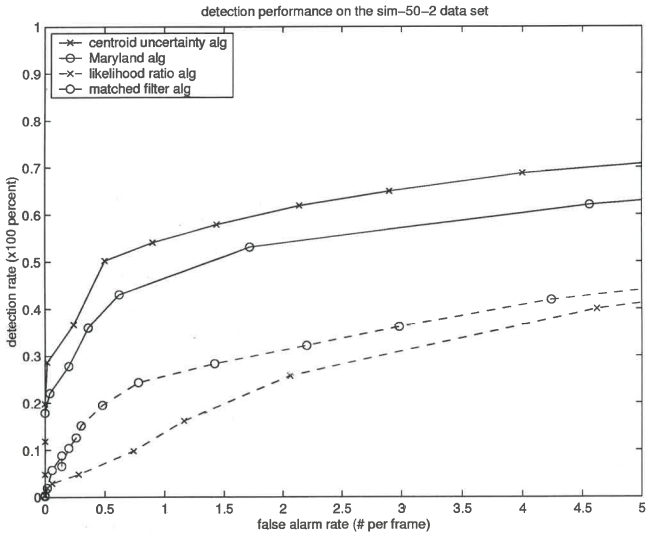
The ROC curve is used to present the detection performance of the tested algorithms. To give a picture of the detection performance over the whole operating range, as well as to show the details of the algorithm performance in the operating range of interest, two versions of the ROC curve are give in Fig. 9. The unconditional recognition and localization performance curves are given in Fig. 10.

5.5 *Discussions*

- From Figs. 9 and 10(a), we see that the proposed algorithm performs uniformly better than all other algorithms in terms of detection and recognition performance over the entire operating range. The recognition performance is especially impressive. For the low false alarm operating range, the localization performance of the algorithm is also the best. For larger false alarm ranges, the localization performance of the algorithm is also among the best.
- The Maryland algorithm has the second best detection performance. Its unconditional recognition rate is also relatively good for low false alarm ranges, but saturates at around a low 60% when the false alarm rate is larger than 10 per frame.
- Following the method outlined in Sec. 5.1.3 for the operating point selection, the least-costing operating point for the proposed algorithm is at detection rate of 65% with an average of 2.1 false alarms per frame. For the Maryland algorithm, the optimal operating point is at detection rate of 53% with an average of 1.7 false alarms per frame. At the optimal operating point, the unconditional recognition rates for the proposed algorithm and the Maryland algorithm are 61% and 48%, respectively. These numbers are relatively low

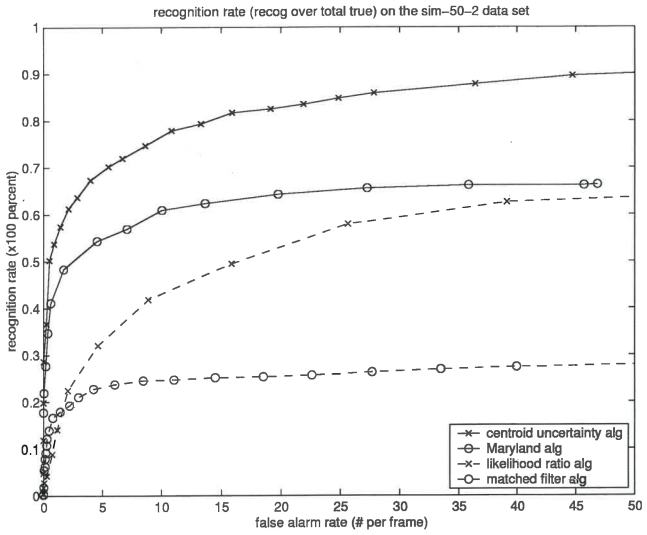


(a) full operating range

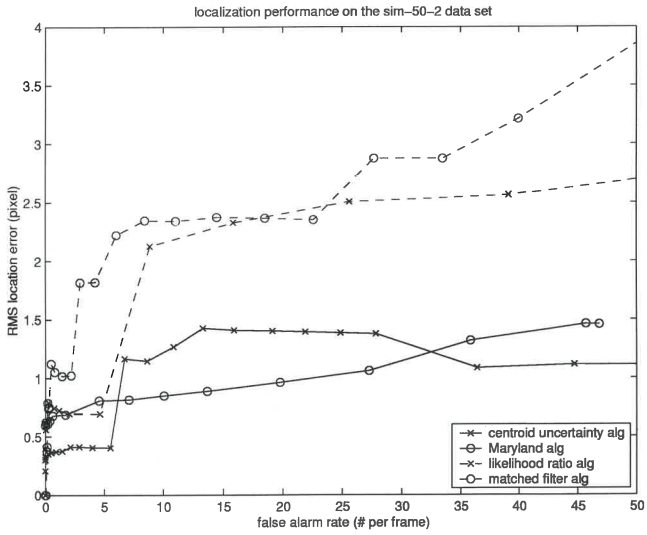


(b) low false alarm range

Fig. 9 Detection performance (ROC curves) of the ATR algorithms.



(a) unconditional recognition rate



(b) localization error

Fig. 10 Unconditional recognition rate (recognized targets divided by all ground-truth targets) and rms localization error of the ATR algorithms.

since they are restricted by the detection rates of the selected operating points. Without the restriction from the detection performance, the conditional recognition rates are 99% for the proposed algorithm and 91% for the Maryland algorithm.

- Due to the great variation in the target appearance in the test images, the matched filter algorithm fails miserably by achieving only a very low recognition rate of around 20% for low false alarm ranges and less than 30% for high false alarms ranges. This is an expected behavior, and it confirms the widely held opinion that matched filter is insufficient for the great variability in FLIR scenes. The target boundary model is more appropriate in this situation.
- Although in some limited-data occassions the likelihood ratio algorithm described in Sec. 5.2.2 performed slightly better than the Maryland algorithm, it is definitely not doing as well in this moderate-sized experiment. It shows that the robustness considerations in the Maryland algorithm pay off quite well in the presence of low target contrast and high noise levels.

6 Summary and On-Going Research

We have described a model-based FLIR ATR algorithm that makes use of boundary contrast for target detection and recognition. The relationship between contrast and boundary point location uncertainty is worked out in an edge detection framework without actually carrying out edge detection, which is highly prone to error due to the low quality FLIR imagery. The location uncertainty along the boundary is combined into a single measure of the centroid location uncertainty which is used as the key saliency measure for target detection and recognition. The significance of the proposed algorithm lies in the sound mathematical models used in combining the boundary contrast by means

of location uncertainty, as well as in relating the contrast to location uncertainty for target boundary points within the edge detection framework. The FLIR ATR algorithm based on this measure shows improved performance in detection, recognition, and localization, with the improvement in recognition performance being the most impressive.

Currently research is being conducted to further improve the performance of the algorithm. The efforts include developing optimal gradient operator for better boundary contrast and location estimation, using the MDL principle to improve the quality of facet parameter estimation in the gradient operator, and combining the proposed algorithm with region-based FLIR ATR algorithms. Also, the relationship between edgel contrast and location uncertainty is being utilized in developing a new edge detection technique using the Bayesian approach.

Acknowledgment

This work was supported by MURI Grant No. N00014-95-1-0521.

Appendix: Partial Derivatives in Centroid Covariance Expression

We now give the exact expressions for the partial derivatives involved in Eq. (20).

When the locations of the polygon's vertices are perturbed, K_x, K_y, and S in Eq. (1) are all perturbed. Let the perturbed quantities be $\hat{K}_x$, $\hat{K}_y$, and $\hat{S}$, respectively. Let ΔK_x, ΔK_y, and ΔS be the difference between the perturbed values and the ideal values, respectively.

$$\hat{S}=\sum_{i=1}^{N}(\hat{x}_i\hat{y}_{i+1}-\hat{x}_{i+1}\hat{y}_i)$$

$$=\sum_{i=1}^{N}[(x_i+\Delta x_i)(y_{i+1}+\Delta y_{i+1})-(x_{i+1}+\Delta x_{i+1})\times(y_i+\Delta y_i)],$$

$$\Delta S=\hat{S}-S$$

$$\approx\sum_{i=1}^{N}[(x_i\Delta y_{i+1}+y_{i+1}\Delta x_i)-(x_{i+1}\Delta y_i+y_i\Delta x_{i+1})],$$

$$\hat{K}_x=\sum_{i=1}^{N}(\hat{x}_i\hat{y}_{i+1}-\hat{x}_{i+1}\hat{y}_i)(\hat{x}_i+\hat{x}_{i+1})$$

$$=\sum_{i=1}^{N}[(x_i+\Delta x_i)(y_{i+1}+\Delta y_{i+1})-(x_{i+1}+\Delta x_{i+1})\times(y_i+\Delta y_i)]\cdot(x_i+\Delta x_i+x_{i+1}+\Delta x_{i+1}),$$

$$\Delta K_x=\hat{K}_x-K_x$$

$$\approx\sum_{i=1}^{N}\{(x_iy_{i+1}-x_{i+1}y_i)(\Delta x_i+\Delta x_{i+1})+(x_i+x_{i+1})\times[(x_i\Delta y_{i+1}+y_{i+1}\Delta x_i)-(x_{i+1}\Delta y_i+y_i\Delta x_{i+1})]\},$$

$$\hat{K}_y=\sum_{i=1}^{N}(\hat{x}_i\hat{y}_{i+1}-\hat{x}_{i+1}\hat{y}_i)(\hat{y}_i+\hat{y}_{i+1})$$

$$=\sum_{i=1}^{N}[(x_i+\Delta x_i)(y_{i+1}+\Delta y_{i+1})-(x_{i+1}+\Delta x_{i+1})\times(y_i+\Delta y_i)]\cdot(y_i+\Delta y_i+y_{i+1}+\Delta y_{i+1}),$$

$$\Delta K_y=\hat{K}_y-K_y$$

$$\approx\sum_{i=1}^{N}\{(x_iy_{i+1}-x_{i+1}y_i)(\Delta y_i+\Delta y_{i+1})+(y_i+y_{i+1})\times[(x_i\Delta y_{i+1}+y_{i+1}\Delta x_i)-(x_{i+1}\Delta y_i+y_i\Delta x_{i+1})]\}.$$

Extending the notation to have $x_0=x_N$, $y_0=y_N$, for $i=1,2,\ldots,N$, we obtain

$$\frac{\partial S}{\partial x_i}=y_{i+1}-y_{i-1},$$

$$\frac{\partial S}{\partial y_i}=-(x_{i+1}-x_{i-1}),$$

$$\frac{\partial K_x}{\partial x_i}=[2x_iy_{i+1}+x_{i+1}(y_{i+1}-y_i)]-[2x_iy_{i-1}+x_{i-1}(y_{i-1}-y_i)],$$

$$\frac{\partial K_x}{\partial y_i}=-(x_i+x_{i+1})x_{i+1}+(x_{i-1}+x_i)x_{i-1},$$

$$\frac{\partial K_y}{\partial x_i}=(y_i+y_{i+1})y_{i+1}-(y_{i-1}+y_i)y_{i-1},$$

$$\frac{\partial K_y}{\partial y_i}=-[2x_{i+1}y_i+y_{i+1}(x_{i+1}-x_i)]+[2x_{i-1}y_i+y_{i-1}(x_{i-1}-x_i)].$$

Now the expressions

$$\frac{\partial Q_x(x,y)}{\partial x_n}=\frac{\frac{\partial K_x}{\partial x_n}S-K_x\frac{\partial S}{\partial x_n}}{3S^2}=\frac{1}{3S}\cdot\left\{\frac{\partial K_x}{\partial x_n}-\frac{K_x}{S}\cdot\frac{\partial S}{\partial x_n}\right\},\quad\text{(A1)}$$

$$\frac{\partial Q_x(x,y)}{\partial y_n}=\frac{\frac{\partial K_x}{\partial y_n}S-K_x\frac{\partial S}{\partial y_n}}{3S^2}=\frac{1}{3S}\cdot\left\{\frac{\partial K_x}{\partial y_n}-\frac{K_x}{S}\cdot\frac{\partial S}{\partial y_n}\right\},\quad\text{(A2)}$$

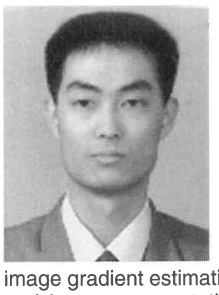
$$\frac{\partial Q_y(x,y)}{\partial x_n}=\frac{\frac{\partial K_y}{\partial x_n}S-K_y\frac{\partial S}{\partial x_n}}{3S^2}=\frac{1}{3S}\cdot\left\{\frac{\partial K_y}{\partial x_n}-\frac{K_y}{S}\cdot\frac{\partial S}{\partial x_n}\right\},\quad\text{(A3)}$$

$$\frac{\partial Q_y(x,y)}{\partial y_n} = \frac{\frac{\partial K_y}{\partial y_n} S - K_y \frac{\partial S}{\partial y_n}}{3S^2} = \frac{1}{3S} \cdot \left\{ \frac{\partial K_y}{\partial y_n} - \frac{K_y}{S} \cdot \frac{\partial S}{\partial y_n} \right\}, \tag{A4}$$

only involve the known locations of the unperturbed vertices, and hence are readily computable.

References

1. J. A. Ratches *et al.*, "Aided and automatic target recognition based upon senory inputs from image forming systems," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(9), 1004–1019 (1997).
2. B. Bhanu, "Automatic target recognition: State of the art survey," *IEEE Trans. Aerosp. Electron. Syst.* **AES-22**(4), 364–379 (1986).
3. B. J. Schachter, "A survey and evaluation of FLIR target detetion/segmentation algorithms," in *Proceedings of DARPA Image Understanding Workshop*, pp. 40–57 (1982).
4. A. S. Politopoulos, "An algorithm for the extraction of target-like objects in cluttered FLIR imagery," IEEE Aerospace and Electronic Systems Society Newsletter, 23–37 (1980).
5. B. Bhanu and R. D. Holben, "Model-based segmentation of FLIR images," *IEEE Trans. Aerosp. Electron. Syst.* **AES-26**(1), 2–11 (1990).
6. Q. H. Pham *et al.*, "Sequential digital filters for fast detection of targets in FLIR image data," *Proc. SPIE* **3069**, 62–73 (1997).
7. B. Ernisse *et al.*, "Complete automatic target cuer/recognition system for tactical forward-looking infrared images," *Opt. Eng.* **36**(9), 2593–2603 (1997).
8. J. H. Kitrosser, "An evaluation of a region-growing algorithm performing detection for automatic target recognition," *J. Imaging Sci. Technol.* **38**(4), 311–320 (1994).
9. K. C. Markham, "Comparison of segmentation processes for object acquisition in infrared images," *Proc. IEEE* **136**(1), 13–21 (1989).
10. R. M. Haralick and L. G. Shapiro, "Survey: Image segmentation techniques," *CVGIP: Image Understand.* **29**, 100–132 (1985).
11. N. S. Friedland and A. Rosenfeld, "An integrated approach to 2D object recognition," *Pattern Recogn.* **30**(3), 525–535 (1997).
12. A. H. Fazlollahi and B. Javidi, "Optimum receivers for pattern recognition problems with nonoverlapping target and background noise," *Opt. Eng.* **36**(10), 2633–2641 (1997).
13. M. Uenohara and T. Kanade, "Use of Fourier and Karhunen-Loeve decomposition for fast pattern matching with a large set of templates," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(8), 891–898 (1997).
14. J. Ben-Arie and K. R. Rao, "A novel approach for template matching by nonorthogonal image expansion," *IEEE Trans. Circuits Syst. Video Technol.* **3**(1), 71–84 (1993).
15. X. Wu and B. Bhanu, "Gabor wavelet representation for 3-d object recognition," *IEEE Trans. Image Process.* **6**(1), 47–64 (1997).
16. M. R. Stevens and J. R. Beveridge, "Precise matching of 3-d target models to multisensor data," *IEEE Trans. Image Process.* **6**(1), 126–142 (1997).
17. A. Kramer *et al.*, "Relational template matching algorithm for FLIR automatic target recognition," *Proc. SPIE* **1957**, 29–37 (1993).
18. M. K. Hamilton and T. A. Kipp, "Model-based multi-sensor automatic target identification for FLIR fused with MMW," *Proc. SPIE* **2485**, 286–294 (1995).
19. S. Z. Der and R. Chellappa, "Probe-based automatic target recognition in infrared imagery," *IEEE Trans. Image Process.* **6**(1), 92–102 (1997).
20. S. Der *et al.*, "Analysis of probe methodologies for target recognition," *Proc. SPIE* **2755**, 46–57 (1996).
21. V. T. Franques and D. A. Kerr, "Wavelet-based rotationally invariant target classification," *Proc. SPIE* **3068**, 102–112 (1997).
22. B. Li *et al.*, "Experimental evaluation of neural, statistical and model-based approaches to FLIR ATR," *Proc. SPIE* **3371**, 388–397 (1998).
23. M. W. Roth, "Survey of neural network technology for automatic target recognition," *IEEE Trans. Neural Netw.* **1**(1), 28–43 (1990).
24. S. K. Rogers *et al.*, "Neural network for automatic target recognition," *Neural Networks* **8**(7-8), 1153–1184 (1995).
25. *Multisensor Fusion for Computer Vision*, J. K. Aggarwal, Ed., Springer-Verlag, New York (1991).
26. N. Nandhakumar and J. K. Aggarwal, "Integrated analysis of thermal and visual images for scene interpretation," *IEEE Trans. Pattern Anal. Mach. Intell.* **10**(4), 469–481 (1988).
27. J. R. Beveridge *et al.*, November 1993 Fort Carson RSTA data collection final report, Technical Report CS-94-118, Dept. of Computer Science, Colorado State University, Fort Collins, CO 80526 (1994).
28. R. M. Haralick, "Overview: Computer vision performance characterization," *Proc. IUW'94*, pp. 663–666 (1994).
29. R. M. Haralick, "Performance characterization protocol in computer vision," *Proc. IUW'94*, pp. 667–673 (1994).
30. R. M. Haralick, "Propagating covariance in computer vision," *Int. J. Pattern Recognit. Artif. Intell.* **10**(5), 561–572 (1996).
31. V. Ramesh, "Performance characterization of image understanding algorithms," PhD thesis, University of Washington (1995).
32. L. A. Chan and N. M. Nasrabadi, "Automatic target recognition using vector quantization and neural networks," *Opt. Eng.* **38**(12), 2147–2161 (1999).
33. J. G. Verly *et al.*, "Model-based system for automatic target recognition from forward-looking laser-radar imagery," *Opt. Eng.* **31**(12), 2540–2552 (1992).
34. D. Cyganski *et al.*, "ROC analysis of ATR from SAR images using a model-based recognizer incorporating pose information," *Proc. SPIE* **2487**, 190–200 (1995).
35. G. Liu and R. M. Haralick, "Assignment problem in edge detection performance evaluation," *Proc. CVPR'00* (2000).
36. T. Kanungo and R. M. Haralick, "Multivariate hypothesis testing for Gaussian data: Theory and software," Technical Report, Intelligent Systems Laboratory, University of Washington (1995).
37. T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York (1958).
38. R. M. Haralick, "Digital step edges from zero crossing of second directional derivatives," *IEEE Trans. Pattern Anal. Mach. Intell.* **6**(1), 58–68 (1984).
39. J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **8**(6), 679–698 (1986).
40. E. P. Lyvers and O. R. Mitchell, "Precision edge contrast and orientation estimation," *IEEE Trans. Pattern Anal. Mach. Intell.* **10**(6), 927–937 (1988).
41. O. A. Zuniga and R. M. Haralick, "Integrated directional derivative gradient operator," *IEEE Trans. Syst. Man Cybern.* **SMC-17**(3), 508–517 (1987).
42. D. H. Marimont and Y. Rubner, "A probablistic framework for edge detection and scale selection," *Proc. ICCV'98*, pp. 207–214 (1998).
43. K. V. Mardia, *Statistics of Directional Data*, Academic Press, New York (1972).
44. W. H. Press *et al.*, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed., Cambridge University Press, Cambridge (1992).
45. R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, Addison-Wesley Publishing Company, Reading, MA (1992).
46. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw–Hill, New York (1991).
47. G. Liu and R. M. Haralick, "Optimizing the integrated directional derivative gradient operator." Technical Report, Intelligent Systems Laboratory, University of Washington (2000).
48. V. Ramesh and R. M. Haralick, "An integrated gradient edge detector—theory and performance evaluation." *Proc. IUW'94*, pp. 689–695 (1994).
49. G. Casella and R. L. Berger, *Statistical Inference*, Duxbury Press (1990).
50. S. Z. Der, "Automatic target recognition using passive infrared and laser radar sensor," Technical Report CAR-TR-762, Center for Automation Research, University of Maryland (1995).
51. G. Liu and R. M. Haralick, "FLIR ATR algorithm using maximum likelihood ratio criterion," Technical Report, Intelligent Systems Laboratory, University of Washington (1996).
52. J. Yao and G. Lebreton, "Invariant-pattern recognition: smart algorithms for segmented two-dimensional patterns with an axis of symmetry," *Opt. Eng.* **38**(12), 2022–2028 (1999).
53. A. D. Lanterman *et al.*, "Kullback-leibler distance for quantifying clutter and models," *Opt. Eng.* **38**(12), 2134–2146 (1999).



Gang Liu is currently finishing his Ph.D. education with the Department of Electrical Engineering at the University of Washington in Seattle, WA, USA. Prior to joining UW, he received his master's and bachelor's degrees, both in electrical engineering, in 1995 and 1992 from Tsinghua University in Beijing, China and Xidian University in Xi'an, China, respectively. His research interests in computer vision include low-level feature extraction (including image gradient estimation, edge and line detection, corner detection and image segmentation), object representation, automatic target recognition, pattern classification and classifier combination, and algorithm performance evaluation.

Robert M. Haralick occupies the Boeing Clairmont Egtvedt Professorship in the Department of Electrical Engineering at the University of Washington. Professor Haralick is responsible for developing the gray scale co-occurrence texture analysis technique and the facet model technique for image processing. In high level computer vision he has worked on robust methods for photogrammetry and developed fast algorithms for solving the consistent labeling problem. He has developed shape analysis and extraction techniques using mathematical morphology, the morphological sampling theorem, and fast recursive morphology algorithms. In the area of document image understanding, Professor Haralick, along with Professor Ihsin Phillips, developed a comprehensive ground-truthed set of some 1600 document image pages most in English and some 200 pages in Japanese. He has also developed algorithms for document image skew angle estimation, zone delineation, word and text line bounding box delineation. His most recent research is in the area of computer vision performance characterization and covariance propagation. Professor Haralick is a Fellow of IEEE for his contributions in computer vision and image processing and a Fellow of IAPR for his contributions in pattern recognition, image processing, and for service to IAPR. He has published over 500 papers and has just completed his term as the president of the International Association for Pattern Recognition.

Journal of Electronic Imaging 9(2), 194–202 (April 2000).

Fundamental matrix from optical flow: optimal computation and reliability evaluation

Kenichi Kanatani
Gunma University
Department of Computer Science
Kiryu, Gunma 376-8515, Japan

Yoshiyuki Shimizu
Sharp, Ltd.
Osaka, 545-0013, Japan

Naoya Ohta
Gunma University
Department of Computer Science
Kiryu, Gunma 376-8515, Japan

Michael J. Brooks
Wojciech Chojnacki
Anton van den Hengel
University of Adelaide
Department of Computer Science
Adelaide, SA 5005, Australia

Abstract. *The optical flow observed by a moving camera satisfies, in the absence of noise, a special equation analogous to the epipolar constraint arising in stereo vision. Computing the "flow fundamental matrix" of this equation is an essential prerequisite to undertaking three-dimensional analysis of the flow. This article presents an optimal formulation of the problem of estimating this matrix under an assumed noise model. This model admits independent Gaussian noise that is not necessarily isotropic or homogeneous. A theoretical bound is derived for the accuracy of the estimate. An algorithm is then devised that employs a technique called renormalization to deliver an estimate and then corrects the estimate so as to satisfy a particular decomposability condition. The algorithm also provides an evaluation of the reliability of the estimate. Epipoles and their associated reliabilities are computed in both simulated and real-image experiments. Experiments indicate that the algorithm delivers results in the vicinity of the theoretical accuracy bound.* © *2000 SPIE and IS&T.* [S1017-9909(00)01202-2]

1 Introduction

Two distinct approaches are available for extracting three-dimensional (3D) information from motion images. One utilizes data in the form of *point correspondences*,[1–4] while the other exploits data in the form of *optical flow*, which is theoretically an instantaneous velocity field over the image frame. Under this latter approach, we can compute the 3D motion of the camera and the focal length, and hence the 3D structure of the scene, in an analytically closed form by regarding the image motion as infinitesimal.[5,6] A similar analysis can be carried out using the fundamental matrix,[7–9] but it requires many stages of complicated analysis and computation.

Optical flow analysis relies heavily upon the accuracy of the optical flow estimation. Usually, optical flow is determined by applying a differential equation called the *gradient constraint* to the gray levels of the images.[10–18] This has the advantage that the flow can be detected over the image frame via a single operation. However, this process involves many stages of approximation such as replacing differentials by finite differences, imposing smoothness constraints, and applying regularization methods. As a result, the accuracy of the detected flow is not high enough for 3D reconstruction of the scene. In order to obtain accurate flow, we need to trace individual feature points directly by the same means as those used for finite motion, such as template matching and spatio-temporal analysis. In this article, we assume that optical flow has been obtained with high accuracy for a limited number of salient feature points.

If the images are captured by a camera moving through a stationary scene, the associated optical flow satisfies, in the absence of noise, a motion equation of special form. This equation is analogous to the epipolar equation arising in stereo vision. The coefficients of the motion equation can be expressed in terms of a matrix, which plays the same role as the fundamental matrix in stereo analysis. For this reason, we call it the *flow fundamental matrix*. The main

Paper 99-048 received Aug. 31, 1999; revised manuscript received Mar. 10, 2000; accepted Mar. 14, 2000.
1017-9909/2000/$15.00 © 2000 SPIE and IS&T.