Gabor Wavelet Representation for 3-D Object Recognition

Xing Wu and Bir Bhanu, *Fellow, IEEE*

***Abstract*— This paper presents a model-based object recognition approach that uses a Gabor wavelet representation. The key idea is to use magnitude, phase, and frequency measures of the Gabor wavelet representation in an innovative flexible matching approach that can provide robust recognition. The Gabor grid, a topology-preserving map, efficiently encodes both signal energy and structural information of an object in a sparse multiresolution representation. The Gabor grid subsamples the Gabor wavelet decomposition of an object model and is deformed to allow the indexed object model match with similar representation obtained using image data. Flexible matching between the model and the image minimizes a cost function based on local similarity and geometric distortion of the Gabor grid. Grid erosion and repairing is performed whenever a collapsed grid, due to object occlusion, is detected. The results on infrared imagery are presented, where objects undergo rotation, translation, scale, occlusion, and aspect variations under changing environmental conditions.**

I. Introduction

MODEL-BASED object recognition in real-world outdoor situations is difficult because a robust algorithm has to consider multiple factors such as: i) object contrast, signature, scale, and aspect variations; ii) noise and spurious low resolution sensor data; and iii) high clutter, partial object occlusion, and articulation. Current approaches use shape primitives, silhouettes and contours, colors, and invariant object features for matching. The performance of these methods is acceptable when objects are well defined, have high contrast, and are at close ranges. However, the recognition results generated by these approaches do not gracefully degrade and produce high false alarms when competitive clutter and object shape distortion are present in the input data [4], [6]. To improve the recognition performance under multiscenarios and varying environmental conditions, model of sensors, atmosphere, and background clutter are helpful in addition to the geometric model of an object. Using only a minimum set of object models and sensor model, multiscale Gabor wavelet representation of objects and a flexible matching mechanism described in this paper can potentially help to improve the recognition performance under real-world situations.

The use of Gabor wavelet representation to recognize three-dimensional (3-D) objects is motivated by the fact that as compared to popular edge-based representations, it is a rich multiresolution representation with sound theoretical basis and can be efficiently implemented. It allows the use of Gabor magnitude, Gabor frequency, and Gabor phase to localize objects, to recognize objects under different scales, and to estimate precise pose under operational constraints (such as occlusion, high clutter, etc.) for automatic target recognition (ATR). It can be used to detect interesting features such as contours and periodic patterns. The representation is less sensitive to small perturbations of contours that are commonly used for matching objects in infrared images. The Gabor filter responses, constituting the core of the representation, are expected to be robust to misalignment in the spatial and frequency domains, and are less sensitive to partial occlusion at least for some frequencies and orientations.

A. Definition of the Problem

The goal of the research presented in this paper is to use a model-based recognition paradigm to recognize 3-D rigid objects with varying appearances, signatures, and possible partial occlusion in highly cluttered sensor data. Distortions involved in most ATR scenarios are induced by 2-D projection of a 3-D object, sensor noise, object occlusion, and articulation. Also for infrared images, the signature of an object will change with changes in environmental conditions, like the time of the day, air temperature, vehicle operating conditions, etc. All these lead to distortion between object models and their images collected by a sensor. Gabor wavelet filters have potential to resolve these problems because they are less sensitive to minor viewpoint variations and can tolerate small local shape distortions caused by the above factors. The Gabor wavelet representation can help to reduce/correct these various errors.

Given a series of two-dimensional (2-D) intensity images as input, which may contain instances of the modeled objects, the system will then detect and identify each of the objects and determine its pose (or aspect), or it may report that none of the modeled objects are present in the input image. The recognition process verifies a given object hypothesis that uses both global and local image information produced by detection and indexing algorithms. Generally, there is more than one object hypothesis for a given region-of-interest (ROI) in a test image, and the hypotheses may consist more than one object *class* and *pose*.

In this research, we are working with objects, such as tank, jeep, truck, high mobility multipurpose wheeled vehicle (HMMWV) and armored personnel carrier (APC), etc. Nine to fifteen aspects and two to three viewing depression angles (angles between the horizon and the line of sight) are defined and modeled for each object type. The actual number of required

Manuscript received November 1, 1995; revised September 19, 1996. This work was supported by DARPA under Grant MDA972-93-1-0010.

The authors are with the College of Engineering, University of California, Riverside, CA 92521-0425 USA (e-mail: bhanu@shivish.ucr.edu).

Publisher Item Identifier S 1057-7149(97)00464-8.

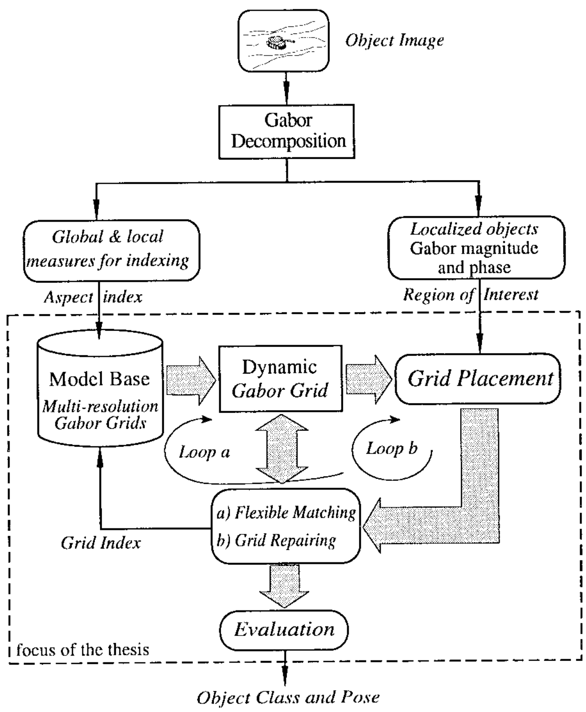


Fig. 1. Object recognition is an iterative process of matching.

TABLE I
RELATED APPROACHES FOR MODEL REPRESENTATION AND MATCHING

Approaches	Advantages	Drawbacks
Dynamic link architecture (Lades [19])	+ Gabor filters for local pattern structure + tolerate distortion + noise-tolerant	- no multi-scale matching - unguided deformation - no occlusion - no indexing support
Hierarchical correlation (Rosenfeld [22])	+ noise-tolerant + fast	- rigid template matching - no rotation/scale handling - no occlusion
Pattern Tree (Burt [12])	+ tolerates global distortion + robust local matching + efficient representation + partial occlusion	- no rotation/scale handling - no indexing support
Multi-Resolution Elastic Matching (Bajcsy [2])	+ tolerates local distortion + coarse-to-fine matching	- expensive unguided deformation - no occlusion handling - no indexing support
Stereo Matching (Sanger [24], Fleet [15])	+ use of Gabor phase to guide matching + coarse-to-fine matching + tolerates distortion	- 1-D matching - no rotation/scale handling - no indexing support
Geometric Invariants (Forsyth [16])	+ 3-D indexing + large aspect range	- requires explicit structural features - limited number of useful invariants
Multi-scale Steerable Filters (Ballard [3])	+ steerable filters for local pattern structures + handles rotation + supports indexing	- scale and scene clutter sensitive - no occlusion - excludes boundary points
Linear combination of models (Ullman [25])	+ generate new aspect using existing aspects of model + 3-D indexing	- sensitive to scene clutter - sensitive to distortion

aspects is computed theoretically and verified experimentally by considering factors such as the size of the object, the range to the object, the depression angle, the quality of the sensor data, variations of object appearances, and the design of Gabor wavelet filters [9].

B. Our Approach

In our model-based object recognition approach, 3-D models of objects are obtained from Gabor wavelet decomposition of a series of viewer-centered 2-D images of the objects at various aspect and depression angles. These models are generated off-line by either using available real data or by simulating geometric models of objects and sensor models. Although all models corresponding to one object are closely related, each model is treated as an independent model for recognition.

Both object models and images are represented by the magnitude and phase responses of multiscale Gabor wavelet filters (Gabor decomposition). The magnitude responses measure the localized signal energy and the phase responses encode the relational structure of an intensity pattern. *Gabor magnitude* is used for matching between an object and a model, based on local energy patterns. *Gabor frequency* is used to estimate the scale variation of a given object from the model. *Gabor phase* is used to evaluate the matching result in terms of average local image displacement between the model and the object. Objects are recognized when they successfully match with a specific model based on distinctive local features in the Gabor wavelet representation.

The general scheme of our system is depicted in Fig. 1.

The focus of the paper is model-based object matching. Object *detection* and *indexing* (which are not the focus of this paper) detect the ROI of an object, and generate hypotheses for a particular aspect of an object, respectively. Objects are detected using tuned Gabor filters [9], [10]. Global and local measures used for indexing are based on the axis of the least moment-of-inertia of the ROI and other local phase-based measures, respectively [10], [11].

Given the ROI of the image, initial matching is performed between hypothesized object models, represented by Gabor grids, and Gabor decomposition of the object image using a grid-placement algorithm to quickly find the location and scale of the object in the input image. Then, flexible matching is performed by allowing deformation of the model about this location. The precise alignment between the model and the object is obtained after performing Gabor phase-based evaluation. The matching results for all hypothesized object models are evaluated to select the best match. This is shown as the *Loop a* in Fig. 1. When object occlusion is present, a grid-repairing process starts to detect and remove subgrid portions that correspond to the occluded parts of the object, and matching is performed using this dynamic (*repaired*) Gabor grid. This is the *Loop b* shown in Fig. 1. Note that Gabor decomposition of an input image is a noniterative computation. Object model is represented by Gabor grid (for details see Section II-B) instead of an image. Gabor decomposition of an object image is matched with model Gabor grid to recognize an object in an image.

The reasons our approach is important for model-based object recognition are as follows:

1) It does not require any image segmentation to extract objects from the background.
2) It does not need explicit shape features and contours which cannot be reliably detected.

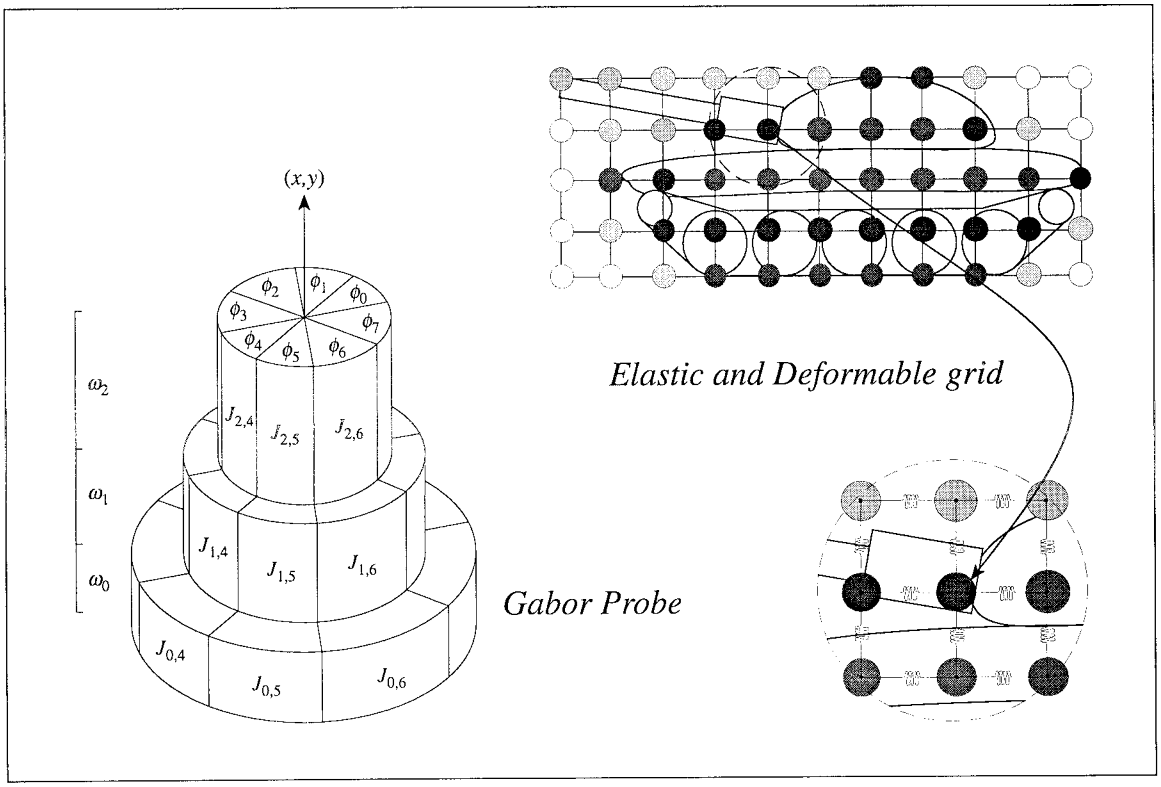


Fig. 2. Example of a Gabor grid. A vector of Gabor wavelet coefficients, which is called a *Gabor probe*, is stored at each node of this grid. Edges between nodes represent geometric constraints between probes, and can be deformed like a spring during matching.

3) It can tolerate significant amount of object distortions due to viewing geometry, scale, aspect and environmentally induced deviations.
4) It allows us to recognize objects at different scales since we can estimate the scale of an object using the multiscale Gabor wavelet representation of an object model.
5) It allows precise estimation of scale/pose alignment between an object and the model by making use of local Gabor phase-based measures.

C. Related Approaches and Our Contribution

The related approaches for model representation and matching are summarized in Table I. Note that indexing is the process to hypothesize which object (and its aspect) exists in an image. Not every recognition system has indexing subsystem, but indexing is necessary when the number of models is large. The most closely related work is the dynamic link architecture technique by Lades, *et al.* [19]. The key differences between our approach and this technique are summarized in Table II.

The main contribution of this paper is to use Gabor wavelet representation to recognize 3-D objects under scale, rotation, translation, and significant distortions in shape and appearance, and under real-world changing environmental conditions. The principle is to use magnitude, phase, and frequency measures of Gabor wavelet representation in an innovative flexible matching approach that can provide robust recognition. The key features of the approach are as follows.

TABLE II
OUR APPROACH FOR OBJECT RECOGNITION VERSUS THE DYNAMIC LINK ARCHITECTURE

Features	Dynamic Link Architecture by Lades et al. [19]	Our Approach This paper
Approach	2-D object recognition	3-D object recognition
Representation	uses Gabor magnitude only.	uses Gabor magnitude, phase and frequency.
Occlusion	cannot be handled.	can handle up to 50% object occlusion.
Scale variations	cannot be handled.	can handle up to 3.5 octaves $(\sqrt{2})^7$ scale variations.
Distortion criteria	simple evaluation criteria.	comprehensive evaluation criteria.
Grid deformation	fixed annealing temperature	varying annealing temperature
Placement of model template	initial placement of the template by heuristics.	initial placement of the template by speedup grid-placement algorithm.
Imagery used	tested with simple images *without* background.	tested with both real visible and infrared images acquired under changing weather conditions.

1) Gabor magnitude, phase and frequency are used in model Gabor grid placement, flexible model matching, handling target scale variation, and evaluation of matching for precise alignment.
2) The distortions in shape and signature of objects are measured by the geometric constraint of the model representation.
3) Several evaluation criteria are used to measure the performance of matching and recognition;
4) Object occlusion is handled by grid erosion and repairing.
5) Object signature variation with changing environmental conditions is handled by simulating object models and anticipating performance degradation in the real world.

II. Gabor Wavelet Representation

Gabor wavelet representation is the set of Gabor functions that are self-similar and differ only by a quadrature phase shift, dilation, and rotation. Gabor functions are joint spatial and frequency domain measures, and are localized transformations in both domains. Gabor functions have many degrees of freedom that allow their spatial and spectral characteristics to be optimally adjusted to a specific visual requirement. Gabor wavelet filters have been used to solve a variety of image processing and computer vision problems [9]–[11], [13], [15], [17], [19]–[21], [26].

A. Gabor Functions and Gabor Wavelet

The general form of a 2-D Gabor function [14] is given as

$$
\begin{aligned}
G(x,y) &= \exp\left\{-\pi[(\frac{x'}{\sigma})^2 + (\frac{y'}{\alpha\sigma})^2]\right\} \\
&\quad \cdot \exp\{j\,[u\,(x-x_i) + v\,(y-y_i)]\} \qquad (1)\\
u &= \omega_k\,\cos\theta_l\,,\ v=\omega_k\,\sin\theta_l \\
\begin{bmatrix} x' \\ y' \end{bmatrix} &= \begin{bmatrix} \cos\phi_G & \sin\phi_G \\ -\sin\phi_G & \cos\phi_G \end{bmatrix} \begin{bmatrix} x-x_i \\ y-y_i \end{bmatrix}
\end{aligned}
$$

where $(x_i,\, y_i)$ is the spatial centroid of the elliptical Gaussian window whose scale and aspect are regulated by σ and α, respectively. ω_k and θ_l ($k,\ l\ \in\ \mathcal{N}$) are the modulation frequency and direction, respectively, and $(u,\,v)$ are the frequency components in x and y directions, respectively. The scale σ controls the size of the filter as well as its bandwidth, while the aspect ratio α and the rotation parameter ϕ_G (generally set equal to θ_l) control the shape of the spatial window and the spectral bandwidth.

Multiple Gabor kernels with various frequencies and orientations, which cover the whole spectral domain, can be organized to sample an image into bandpass energy channels—image decomposition [9]. These are localized transforms when compared with Fourier transform, such that the extent of one local measure is limited to a small neighborhood defined by the size of the filter kernel.

By representing Gabor wavelet filters as a set of self-similar and dilated quadrature pair $G_{\vec{\psi}_{k,l}} = \{\,G^+_{\psi_{k,l}},\ G^-_{\psi_{k,l}}\,\}$, the log-polar sampling in the frequency domain generated by the 2-D wave propagation vector $\vec{\psi}_{k,l}$ is given as

$$\vec{\psi}_{k,\,l} = \omega_k\,e^{i\theta_l},\ \text{where}\ \ \omega_k = \rho^k\,\omega_0, \quad \text{and} \quad \theta_l = l\,\theta_0 \qquad (2)$$

where k is the frequency index of the wavelet ($k=0,\ \cdots\ K$), l is the orientation index of the wavelet ($l\ =\ 0,\ \cdots\ L$), and ρ is the scaling factor of the wavelet. In a biologically inspired scheme, different Gabor functions in the wavelet representation have sizes distributed in logarithmic steps, such as one octave or half-octave. Also, the modulation frequency increases proportionally with the reduction in scale

$$\sigma_k = \frac{1}{\rho}\,\sigma_{k-1}, \quad \omega_k = \rho\,\omega_{k-1} \quad \text{and} \quad \sigma_k\,\omega_k = \sigma_{k-1}\,\omega_{k-1}.$$

By introducing a new parameter called *bandwidth-frequency ratio* $\lambda\ =\ 2\pi/(\sigma_k\omega_k)$, the wavelet filter kernel's frequency and orientation bandwidth can be defined as

$$\Delta\omega_k = \frac{2\pi}{\sigma_k} = \lambda\omega_k, \quad \Delta\theta_l = 2\ \arctan(\frac{\pi}{\alpha\sigma_k\omega_k}).$$

Since we want to be able to recognize objects at different scales and orientations, we chose filters to maximize responses for seven scales and eight orientations. Corresponding to each filter, we have a magnitude and a phase response. The high- and low-frequency Gabor filters behave like edge detectors and lowpass filters, respectively.

In this paper, Gabor wavelet filters $G_{\vec{\psi}_{k,l}}$ are defined by seven logarithmically spaced center frequencies (filter bands) and eight orientations for each filter band. Thus, we sample the frequency domain by 56 bandpass channels. These filters are indexed by $k \in \{0,\cdots,6\}$ (frequencies) and $l \in \{0,\cdots,7\}$ (orientations). The reason one expects 56 different Gabor filters treated as being of equal importance should be considered a good representation for an object is that we do not know the scale and orientation of objects. As a result, we treat the 56 Gabor filter responses equally. In the "grid placement" process (see Section III-A), we also estimate the object scale using these Gabor filter responses.

Other parameters of the filters are

$$\alpha = 1,\ \ \rho = \sqrt{2},\ \ \omega_0 = \pi/16,\ \ \theta_0 = \pi/8\ \ \text{and}\ \ \lambda = \pi/4.$$

The size of the Gabor filters varies with the change of center frequencies (see (2) and subsequent equations in Section II-A). The range of the center frequencies is selected according to the size of the object, so that local information can be adequately represented by the wavelet. It is between $\pi/2$ and $\pi/16$ in our experiments. Using half-octave ($\sqrt{2}$) modulation frequency ratio creates a 50% overlap between filter frequency and orientation bandwidth. Although the representation is overcomplete, it helps to represent an object in a smoothly varying manner with different scales and tolerate increased aspect distortions. Note that the filters exceed the spatial dimension of the objects for most frequency ranges and locations on the grid. Consequently, the representation is inherently a measure of object-plus-background phenomena. The responses of the filters depend upon both the object and background. Lacking an explicit segmentation phase to distinguish object from nonobject signal, most of what is measured depends upon both.

The bandwidth-frequency ratio λ also determines the number of aspects needed for recognizing an object. For $\lambda = \pi/4$, aspects are needed for every 44°. A simplified proof (assuming a periodic pattern) is given below.

Let S be one of the *local* planar surface patch of an object, and S_A and S_B be the area of the surface patch seen by the viewer in location A and B, respectively. Let the azimuth angle between A and B be β, and the elevation angle for these two aspects is α. Then, we have following relationship between S_A and S_B:

$$S_B = S_A\cos\beta\cdot\cos\alpha.$$

Suppose S contains a periodic pattern of frequency Ω_1. Due to the foreshortening effect, at location B the frequency of this periodic pattern will change to Ω_2. Then, Ω_2 can be estimated approximately as

$$\Omega_2 \approx \frac{\Omega_1}{\cos\beta}.$$

Here, we ignore the distortion due to the rotation in the elevation direction.

To answer the question of how many model aspects are needed to represent an object in order to cover all viewing parameters, we can simply take a look at the relationship of the two signals in the frequency domain. If the highest frequency channel in our Gabor wavelet is tuned to frequency Ω_1 (since high frequency channel has wide bandwidth), then the periodic pattern in S_A will have a peak response at Ω_1.

Correspondingly, the periodic pattern in S_B will have a peak response at Ω_2 using the same Gabor filter. So, the maximal distortion between these two signals with respect to peak response in frequency is $\Delta\Omega_1/2$. Then, we have the following relation:

$$\cos\beta \approx \frac{\Omega_1}{\Omega_2} \approx \frac{\Omega_1}{\Omega_1 + \Delta\Omega_1/2}$$

where $\Delta\Omega_1$ is the bandwidth of the Gabor filter, and

$$\Delta\Omega_1 = \lambda\Omega_1$$

here λ is the bandwidth-frequency ratio defined in the paper. Therefore, we have

$$\cos\beta = \frac{1}{1+\lambda/2}.$$

From the above equation, we can see that the number of object aspects needed to cover all viewing aspects is independent of the filter's center frequency and the frequency of the periodic pattern. This claim is restricted only to local structure of the object, since Gabor representation is localized. Thus, it is the *bandwidth-frequency ratio* λ that determines the number of aspects. For $\lambda = \pi/4$, model aspects are required for every 44°.

It is our intention that the wavelet cover the entire object. In order to recognize objects with all possible aspects, object model is sampled (for $\lambda = \pi/4$) at least for very 44° to cover the entire object.

In our experiments, the size of the filters varies from 11 × 11 to 89 × 89, and size of the correctly recognized objects varies from 256 pixels (16 × 16) to 16K (∼120 × 120) pixels. The kernel, 89 × 89, may seriously over sample the image. However, note that we do not know the scale of the object. Some filters will approximately match the size of the object within chosen ranges (256 pixels to ∼16K pixels). We do not discount responses where part of the kernel falls outside the object, since the scale of the object is unknown. It is possible to design a filter set and the associated recognition scheme for objects whose size may be smaller or larger than the filter set used in this paper. However, note that when objects are far from the sensor and have only few pixels, this and any model-based recognition approach will not be suitable.

B. Model Representation

The Gabor wavelet decomposition of an object image $I(\mathbf{x})$, obtained by convolving it with the complex Gabor wavelet filter kernels $G_{\vec{\psi}_{k,l}}$, is an iconic multiresolution template. To reduce the interpixel redundancy, subsampling this template forms a Gabor grid $\mathcal{G}_D$ that covers the object with $N \times M$ nodes (vertices V') in the *x* and *y* directions, and edges (E'), respectively. Thus, $\mathcal{G}_D = \{V',\ E'\}$.

1) Grid Nodes: Each node $v_j \in V'$ is a triple, - $v_j = \{\mathbf{x}_j,\, R_j^+,\, R_j^-\}$ where $\mathbf{x}_j$ is the image coordinates of grid node j (with respect to some normalized coordinate frame). Nodes are selected with fixed distance D_{space} from neighboring nodes for a model grid, $\mathbf{x}_j = (x_0 + nD_{space},\ y_0 + mD_{space})$, where ($n, m \in \mathcal{N}$). R_j^+ and R_j^- are vectors of length KL (where K is the total number of frequencies, and L is the total number of orientations of the wavelet) referred to as the sine and cosine parts of the *Gabor probe*

$$R_j^+[k,\, l] = (I * G_{\psi_{k,l}}^+)[\mathbf{x}_j], \quad \text{and} \quad R_j^-[k,\, l] = (I * G_{\psi_{k,l}}^-)[\mathbf{x}_j]. \tag{3}$$

where $(G_{\psi_{k,l}}^+,\, G_{\psi_{k,l}}^-)$ is a Gabor wavelet quadrature filter pair (*cosine* and *sine* components of a Gabor filter) with center frequency ω_k and modulation orientation ϕ_l. With frequency spacing Δ_ω and orientation spacing $\Delta_\phi = \pi/N_\phi$, ω_k and θ_l can be computed as

$$\omega_k = \omega_0 \cdot \Delta_\omega^k \ \text{ for } 0 \le k \le N-1$$

and

$$\phi_l = \phi_0 + l \cdot \Delta_\phi \ \text{ for } 0 \le l \le L-1.$$

ϕ_0 is taken as 0° and D_{space} is either 11 or 13 pixels for the experiments reported in this paper. (D_{space} is 11 pixels for all the results shown using real data with the exception of example shown in Fig. 11, for which D_{space} is 13 pixels. For all simulated data used in this paper, D_{space} is 13 pixels).

Note that the grid spacing is slightly larger than the width of the smallest kernel. We intentionally chose the spacing of the grid to be slightly larger than the width of the smallest kernel, since a single object model is used to recognize objects with different scales. When matching object and model with different scales, the grid edge needs to be scaled accordingly. Therefore, the smallest object that can be recognized by our model is constrained by the size of the highest Gabor filter center frequency and the spacing of our Gabor grid. This is discussed in detail in Section III-A (see (7)).

2) Grid Edges: The role of the graph edges $e_j \in E'$ is to represent neighborhood relationships and to serve as constraints during matching. They are interpreted as elastic links, such that an edge can be deformed like a spring to make a model probe match with the Gabor decomposition of a distorted object. The length between two nodes $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$ and angles between edges serve as initial constraints. Thus, the deformation can be measured and penalized immediately during matching.

Fig. 2 shows the Gabor grid and Gabor probe representation. The magnitude of the Gabor probe is used to measure the similarity between matched local features, while the phase of Gabor probe is used to fine tune the matching result. The extracted information (both signal energy and local pattern

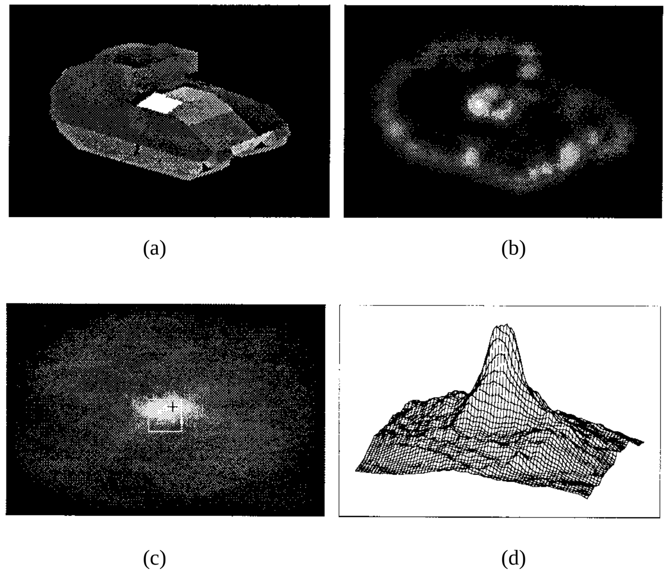


Fig. 3. Illustration of grid placement algorithm. (a) Object image. (b) Marginal magnitude of the object image Gabor decomposition. (c) Search region for object localization. (d) Similarity surface between the image decomposition and a specific model Gabor grid.

structure) associated with each probe spans a multiresolution neighborhood whose size equals the extent of the filter kernels. Each Gabor probe is pictured as a "Hanoi tower" in Fig. 2, which is a set of concentric multiscale disk platters, where the low-frequency channel is the large, thin platter, and high-frequency channel is the small, thick platter. It records Gabor wavelet decomposition of an object at a spatial location **x** with certain spectral extent $\Delta\omega_k$. It also represents the fact that the low-frequency channel extracts coarse image features in a large neighborhood, while high-frequency channel can extract fine localized features in a small neighborhood.

The grid may suggest that we are making localized measurements associated with object parts. However, note that the grid node records the localized features at each spatial location, and the grid-edge records the spatial relationship between nodes. We are not doing explicit parts-based object recognition like the work of Biederman [7] and others.

III. Flexible Model-Based Object Recognition

The flexible object recognition process includes i) *model grid placement*, ii) *flexible model matching*, and iii) *evaluation of matching*. Initial matching (model grid placement) between the Gabor magnitude response of an input image and a hypothesized model is performed to find the location of an object in an input image where the model grid is to be placed. Flexible matching performs model and image matching based on deformed model Gabor grid, and fine tunes it using Gabor phase information. Also, grid repairing is performed whenever a collapsed grid indicating the presence of an occluded object is detected. Finally, evaluation of matching is performed to select the best matched model by following the selected rules.

A. Model Grid Placement

The goal of model grid placement is to find the spatial location in the object image (ROI passed by the "localized objects" box in Fig. 1) where maximal similarity between a model grid and Gabor-based image features of an object is achieved. At this stage of the matching process, the hypothesized model Gabor grid G_{idx}, generated by the indexing process that potentially corresponds to the object aspect is positioned at $\mathbf{x}_{idx}$, scaled by s_{idx}, and rotated by ϕ_{idx} to reflect the pose and position of the potential object aspect A_{idx}, while the grid is kept rigid as in

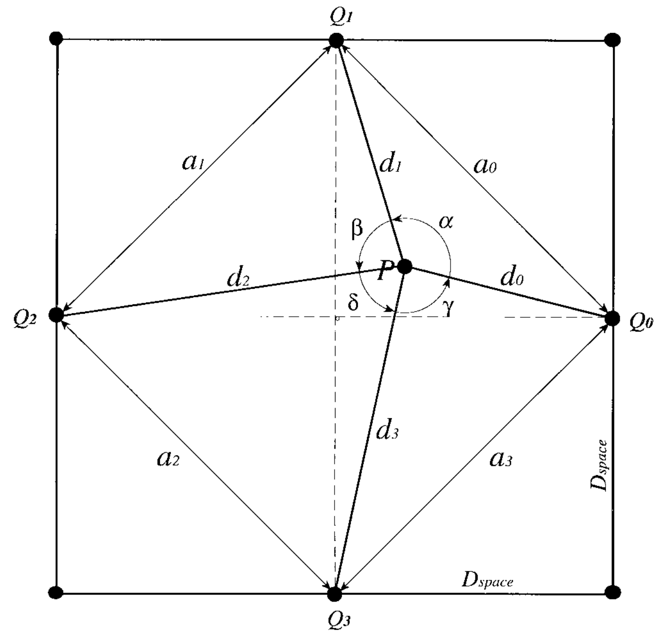


Fig. 4. Illustration of length and angular deformation of the Gabor grid. The deformed grid node *P* and edges with respect to the undeformed grid (in dashed lines) are shown.

$$G_{idx} = \{V_{idx}, E_{idx}\} \xrightarrow{\mathbf{x},s,\phi} G'_{idx} = \{V'_{idx}, E'_{idx}\}$$

where

$$\begin{aligned} v_j &= \{\mathbf{x}_j, P_j^+, P_j^-\} \xrightarrow{\mathbf{x},s,\phi} v'_j \\ &= \{\mathbf{x} + S_{s,\phi}(\mathbf{x}_j), S_{s,\phi}(P_j^+), S_{s,\phi}(P_j^-)\} \end{aligned} \tag{4}$$

$$e_{i,j} = d_{i,j} \xrightarrow{\mathbf{x},s,\phi} e'_{i,j} = s \cdot d_{i,j} \tag{5}$$

for all $v_j \in V_{idx}$ and $e_{i,j} \in E_{idx}$. The function $S_{s,\phi}()$, which performs a scaling and rotation operation on the grid nodes, is defined as

$$S_{s,\phi}(\mathbf{x}) = s\begin{bmatrix}\cos\phi & -\sin\phi \\ \sin\phi & \cos\phi\end{bmatrix}\mathbf{x}. \tag{6}$$

When the scale factor **s** is a power of Δ_ω ($s = \Delta_\omega^k$ for some $k \in \mathcal{N}$) and the orientation ϕ is a multiple of Δ_ϕ, the above transformation corresponds to deriving a new model grid $\mathcal{G}'_D$ at a given scale by scaling down edges of the Gabor grid by factor **s**, and shifting and rotating Gabor probe P_j at each node v_j from the corresponding frequency index ω_s and orientation ϕ_s.

$$\begin{cases} e_j & \mapsto \quad e_j/(\sqrt{2})^s, \\ P_j(\omega_k, \phi_l) & \mapsto \quad P_j(\omega_{(k-s)}, \phi_{l-s}). \end{cases} \tag{7}$$

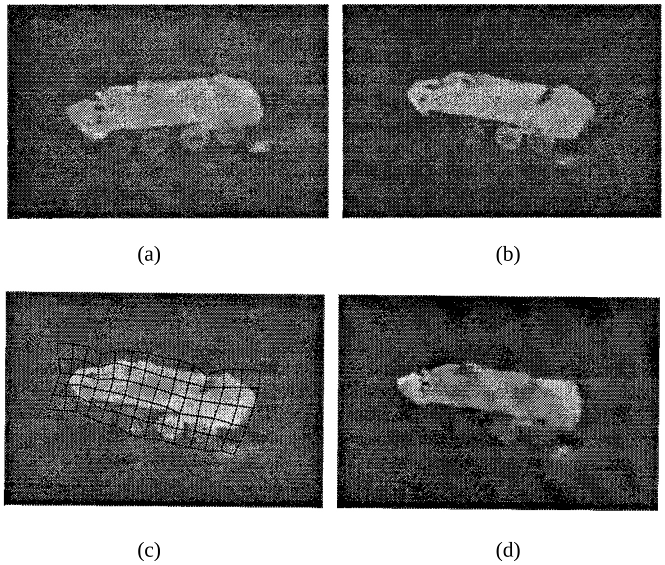


Fig. 5. Illustration of the quality of flexible matching. (a) Model image. (b) Object image. (c) Matching result. (d) Projected model.

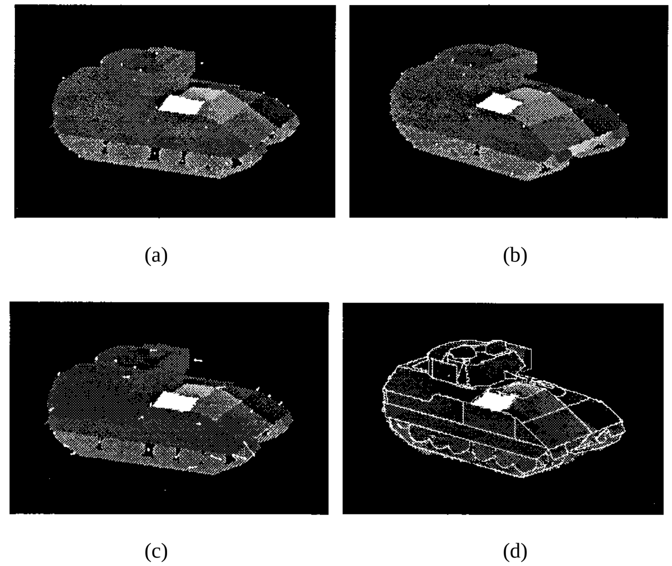


Fig. 6. Illustration of phased-based evaluation. By projecting points from object (b) to model (a), matching errors are estimated in terms of local image displacement (c) measured by their local phase difference between the corresponding point in (a) and (b). Object and the backprojected model (shown as boundary edges) are shown in (d) to illustrate how the matching has taken place. Note the front part of the model and the object.

In case $\mathbf{s}$ is not a power of ρ or ϕ is a not a multiple of Δ_ϕ, we can either i) round the scale factor $\mathbf{s}$ to $\mathbf{s}'$, which is the closest multiple of the frequency index ρ in (7), and let the subsequent flexible matching overcome this small scale distortion in Gabor decomposition (note that the grid edge will be scaled according to the exact scale factor $\mathbf{s}$.) or ii) implement a suitable interpolation scheme over scale and orientation. Object scale variation that can be handled by our recognition system is constrained by the number of center frequencies of the Gabor wavelet and the edge length of the model grid.

When the object scale is unknown, our multiscale representation of the model can help to estimate object scale using the grid placement algorithm given below. Since the object image decomposition by a Gabor filter with a specific frequency corresponds to a representation of an object at a specific scale, by computing the similarities of these representations between an object and a model grid, it allows us to estimate the object scale.

Since Gabor wavelet representation captures both coarse and fine information of an object, it is possible to combine this information to locate the existence of an object in an image [13]. By focusing our attention inside the ROI, we can safely assume that the high-magnitude response in the lower-frequency channel suggests that there is a possible interesting object in that region. We start the matching process by placing the center of the model Gabor grid at the location where the high Gabor magnitude response is obtained in the image. Now the correlation between the two is performed.

The location of the maximum correlation is used as the initial placement of a model grid for subsequent flexible matching. The key steps of the algorithm are given below.

Grid Placement Algorithm

1) Compute the *marginal* magnitude response of Gabor decomposition of the object image by summing up the magnitude responses from all filter orientations but with a common frequency

$$\overline{m}(x_i, y_j, \omega_k) = \sum_{\theta_l} m(x_i, y_j, \omega_k, \theta_l). \tag{8}$$

It generates the magnitude response for a specific frequency. At this stage in processing we want the greatest magnitude without regard for orientation.

2) Get an estimate of the object's spatial center ($\bar{\mathbf{x}}$) by computing the center of a rectangular bounding box b for which the *marginal* Gabor magnitude defined in (8) for the lowest filter frequency band is greater than the threshold t. In our experiments we have set $t = 1/3$ of the maximum magnitude. Note that this step of the algorithm is different from the "localized objects" box in Fig. 1.

3) Start the grid-placement algorithm in the area of the object image defined by $\bar{\mathbf{x}} \pm \Delta A/2$, where ΔA is based on the size of the image. The grid-placement algorithm computes the similarity (defined in Section III-B2) between all model grid nodes and the Gabor image decomposition at corresponding spatial locations with respect to the centroid of the model grid.

4) Select the local similarity peak computed by the grid-placement algorithm in the candidate region as the grid placement index, or expand the image region under consideration by $\pm\Delta$ if the similarity peak falls on the region boundary.

Note that the information regarding the region-of-interest of the image is given to the box labeled as "focus" box in Fig. 1. We know that something is of interest in this ROI but we do not know exactly where to place the model Gabor grid for matching. It may be near the center of ROI or some other location. This is why we need the model Gabor grid placement algorithm (and step 2) of this algorithm. As is traditionally done, ROI will also have a small number of background pixels all around it, and there is a potential object inside it.

An example is given in Fig. 3 where the image is of size 300×200 pixels. The grid-placement algorithm has detected the high magnitude response region using the *marginal* magnitude response shown in Fig. 3(b), and the computation is restricted to a small region, which is the potential center of the object, shown in Fig. 3(c). The size of this region $(\Delta A+1)\times(\Delta A+1)$ is 31×31 pixels. The global similarity peak between the model and the object shown in Fig. 3(d) is correctly detected inside the search region (marked as "+") shown in Fig. 3(c). The search region is dramatically reduced when it is compared to the whole image. In this experiment, responses from filters at eight different orientations for the lowest filter frequency band (see Section II) are used to generate the boundary of the search region $(\Delta A+1)\times(\Delta A+1)$, 56 filter responses of the whole filter set are used to compute the similarity surface in Fig. 3(d). Note that the approach is not limited to the case, as the example may suggest, when there is a single peak response toward the center of the image. If the bright spot (Fig. 3(c)) is located near the boundary, step 3) and step 4) of the algorithm will take care it. In step 3) there will be low similarity, and if the peak is located near the boundary (as the case will be depending upon how close the bright spot is to the boundary), in step 4), the image region under consideration (i.e, the size of ROI) will be expanded. The computational efficiency of the algorithm will depend upon the size of the region. In general, the process can be repeated for each location that has a high marginal magnitude response of the Gabor decomposition.

B. Flexible Model Matching

After object localization, flexible matching starts to verify the hypothesis for a model by moving nodes of the model Gabor grid locally and independently to find the best matched image probes. In this process, the 2-D image of an object with small aspect distortion from the model is matched by small, local elastic deformations of the model. When external forces are applied, an elastic model/object is deformed until an equilibrium state between the external forces and internal forces resisting the deformation is achieved. This equilibrium state can be described as

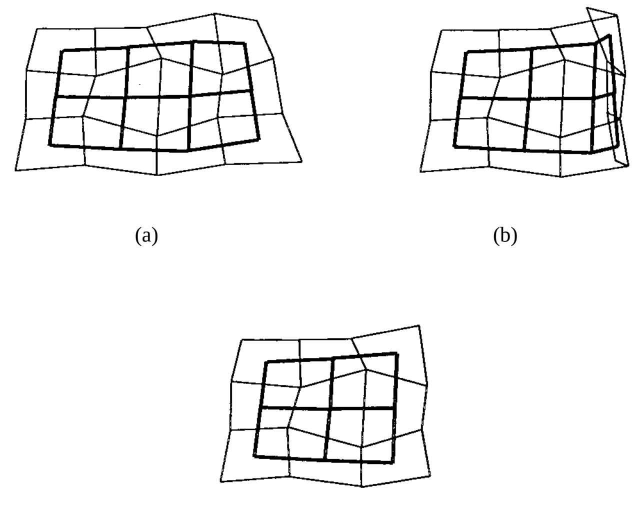
$$\mu\nabla^2\mathbf{u} + (\gamma+\mu)\frac{\partial\mathbf{u}}{\partial\mathbf{x}} + \mathbf{F} = 0 \tag{9}$$

where **x** is the coordinate of the object image representation, **u** is the displacement of the deformation, **F** is the external force, and μ and γ define the elastic properties of the model/object. To find the equilibrium state when the deformable model grid is matched with an image decomposition, the external forces in (9) are estimated by the extent of deformation **u** at which the best similarity between the object and the model is achieved. This can be expressed as an iterative process that minimizes a cost function $\mathcal{C}$ balanced between grid deformation $\mathcal{D}$ and local similarities $\mathcal{S}$. Therefore, we can rewrite (9) as a cost function of flexible matching

$$\mathcal{C} = \mu\sum_{i}^{N}\mathcal{D}(v_i) - \sum_{i}^{N}\mathcal{S}(P_i^I,\ P_i^M) \tag{10}$$

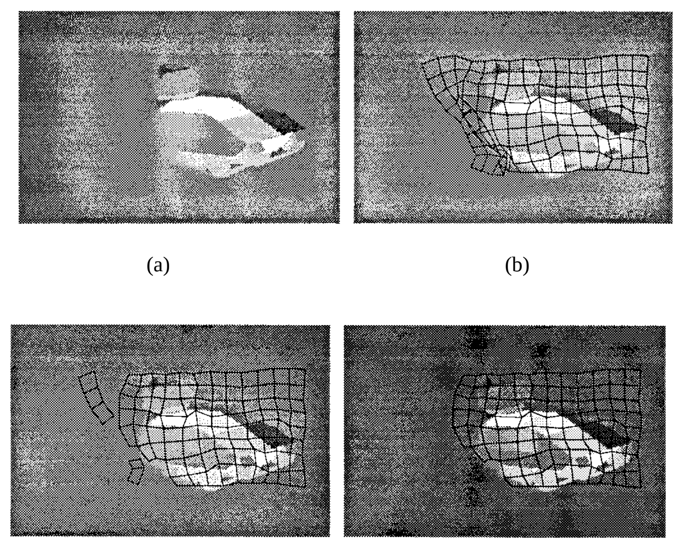
where N is the total number of grid nodes, μ is the elastic parameter that controls the grid deformation. v_i is a grid vertex, and P^I and P^M are the image and model Gabor decompositions, respectively. Note that for model Gabor nodes their values remain the same at all time, only the matched image probes change. During flexible matching the node positions change as the model matches with the image probes.

1) Grid Deformation: To compensate for small aspect variation between a model and an object, and changes induced in



(a) (b) (c)

Fig. 7. Repairing a collapsed grid. (a) Connected (thin line) and complementary grid (thick line) of a Gabor grid. (b) Collapsed grid. (c) Repaired grid after grid erosion.



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

Fig. 8. Occluded object (40%) and results after grid erosion. (a) Occluded object. (b) Initial matching. (c) Grid erosion. (d) Final result.

images as a result of varying environmental conditions such as time of the day and air temperature, grid deformation is allowed to find optimal and localized matches between model Gabor probes and Gabor decompositions of an input image. Since in this paper we are working with rigid objects, the Gabor grid has to be topology preserving. By grid deformation, we mean that the edges of the template are stretched or squeezed like a spring, whereas the information stored at each node (Gabor probe) is unchanged. When a node moves in a small tolerable extent, the Gabor probe associated with this node also moves, and the edges connected to this node are deformed. Note that the warping of the grid may imply a warping of the image neighborhood and warping the Gabor wavelet filters in a corresponding manner. Further, warping of the grid may indicate that the local image structure has warped, and hence the same Gabor filters actually measure different properties of the image. However, note that our representation can only handle image warping due to aspect changes to a certain degree. Under this restriction, the Gabor filters response is invariant to the small warping, since we are using 56 filters responses (with different orientation and center frequencies) instead of one.

To precisely estimate the deformation of the Gabor grid, two kinds of measures are used. Namely, they are *length* and *angular* deformations as shown in Fig. 4. They are computed by comparing the deformed grid with its original structure, which has fixed length D_{space} and a rectilinear grid. The deformation for a node v_k is then calculated as

$$\mathcal{D}(v_k) = \sum_{i=0}^{3} (d_i - D_{space})^2 + \sum_{i=0}^{3} \left(a_i - \sqrt{d_i^2 + d_{i+1}^2}\right)^2. \tag{11}$$

The first term in (11) measures the grid length deformation. It is zero for rigid or shear transformation. The second term measures the angular deformation of the grid, which is zero under a rigid transformation. In order to make both measurements compatible, they are in the form of distance measures that are proportional to the length of the grid. It can be seen from Fig. 4 that for the movement of any single grid node P only the similarity measure for P and deformations between P and its four connected neighbors ($Q_0, \cdots, Q_3$) need to be updated. Thus, update of the cost function (10) during matching can be executed in parallel.

2) Similarity Measurement: Given each Gabor probe given by (3) as a vector of Gabor wavelet decomposition of magnitude at a spatial grid location, the match of local features between probes from a model and an image corresponds to a search of maximum similarity between a model Gabor probe and an image probe. The similarity between two measured Gabor probes $\vec{P}$ (model) and $\vec{Q}$ (object) is computed as follows:

$$\mathcal{S}(\vec{P},\ \vec{Q}) = \frac{\vec{P}\cdot\vec{Q}}{||\vec{P}||\ ||\vec{Q}||}\ \min\left(\frac{||\vec{P}||}{||\vec{Q}||},\ \frac{||\vec{Q}||}{||\vec{P}||}\right)\min\left(||\vec{P}||, ||\vec{Q}||\right). \tag{12}$$

The first term in (12) measures the angle between vector $\vec{P}$ and $\vec{Q}$. The second term compares the length of the two vectors. The third term is used to make the similarity measurement proportional to the magnitude of the smaller of the two vectors. The rationale behind the similarity metric between two probes is as follows. It should be maximum when they point in the same direction (first term) and have about the same magnitude (second term). Since we want to minimize the similarity between a model probe and an image probe that may arise from the background in the image, the similarity should reduce by the minimum of the relative magnitudes of the two probes (second term), and it should further reduce by the lower of the magnitude of the two probes (third term).

In addition, we have to consider the differences between object appearances due to environmental changes such as time of the day and air temperature. As an example in visible and infrared images, changes in time will cause an object with a fixed viewpoint to have different signatures or absolute contrast. These changes will be primarily reflected in the third term in equation (12). Since we want to be able to recognize objects under varying environmental conditions, the third term in (12) is dropped, and a normalization factor η is introduced to provide comparable similarity measures for different object signatures obtained under varying environmental conditions. The environmental conditions invariant metric is

$$\bar{\mathcal{S}}(\vec{P},\ \vec{Q}) = \frac{1}{\eta}\ \frac{\vec{P}\cdot\vec{Q}}{||\vec{P}||\ ||\vec{Q}||}\ \min\left(\frac{||\vec{P}||}{||\vec{Q}||},\ \frac{||\vec{Q}||}{||\vec{P}||}\right). \tag{13}$$

In practice, we choose η to be

$$\eta = \min\left(\frac{||\vec{P}||}{||\vec{J}||},\ \frac{||\vec{J}||}{||\vec{P}||}\right)$$

$$\text{where } \vec{J} = \{\vec{J}_k :\ \max(\frac{\vec{P}\cdot\vec{J}_k}{||\vec{P}||\cdot||\vec{J}_k||}),\ \forall\ \text{probes}k\}. \tag{14}$$

Either (12) or (13) can be used in (10) depending upon the need for multiscenario recognition (see Section IV-A). For an object model obtained in a specific environmental condition, the metric given by (13) will provide approximately constant similarity as the environmental conditions change. Note that for (12) the similarity will increase as the contrast between the model and the image decreases. Likewise for (12), the similarity will decreases as the contrast between the model and the image increases.

The flexible matching algorithm that minimizes (10) based on simulated annealing [18] is given below.

Flexible Matching Algorithm

1) Use the location generated and the scale estimated by the grid-placement algorithm as the initial placement of the model grid on top of the image decomposition with the estimated scale factor.
2) For each grid node of the model (visited in random order), take a random step s. A move s for a node is valid and can be accepted if either
    a. the global cost $\mathcal{C}$ is reduced due to this move, or
    b. changes in cost $\Delta\mathcal{C}$ satisfies a probability $\exp(-\Delta\mathcal{C}/T)$, where T is the annealing temperature.
3) The matching terminates and produces a deformed model grid if either the matching reaches a desired

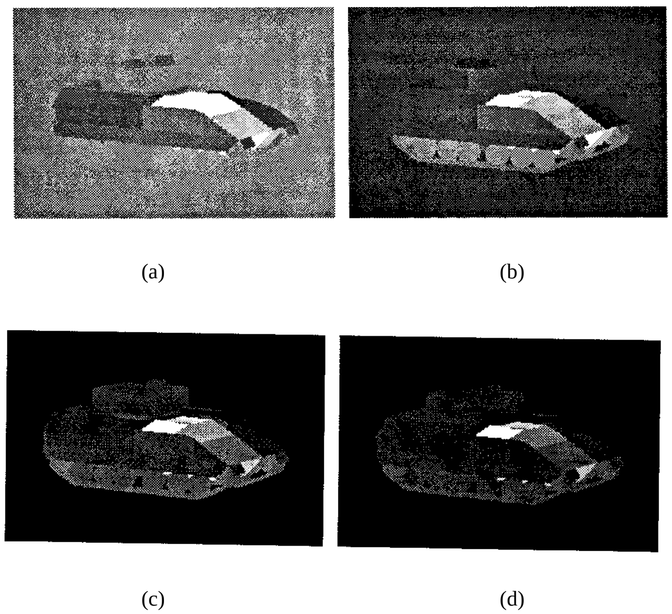


Fig. 9. Four object signature under varying environmental conditions depicted in Fig. 10. All objects have a 10° depression angle and different aspects. The time of the day at which the signatures are simulated is also shown. Changes in object signatures are clearly visible. (a) Frame 5 (11:00 am). (b) Frame 9 (3:00 pm). (c) Frame 17 (11:00 pm). (d) Frame 21 (3:00 am).

cost, or the annealing temperature is freezing. If neither condition is satisfied, continue previous step with temperature decreased by a cooling factor β.

The flexible matching process described above is controlled by three parameters, the elasticity parameter (μ), the annealing temperature (T) and the cooling factor (β). They are determined experimentally. μ controls the degree of deformation allowed for the Gabor grid. T controls the probability of finding the best matched local probes inside the Gabor decomposition of an object. The process starts with a high temperature and cools down generally at a constant rate β, until it is stable (the cost cannot be reduced any more). The annealing settles upon a locally optimal configuration. Underestimating the value of μ will derive a collapsed grid during matching. An overestimate of the value of μ will keep the process from generating the optimal matching result because the grid is too rigid.

For the experimental results reported in this paper, the annealing temperature T is set between 3 and 5, the elastic parameter μ is set between 0.8 and 2.5, such that the small number allows more grid deformation, and the larger number allows less grid deformation. The cooling factor β is generally set to 1.15.

To illustrate the quality of matching under image distortions, an example is given in Fig. 5, in which a matched model whose aspect is slightly different from the aspect of the object is *backprojected* onto the object image using the matched deformed Gabor grid. To backproject the model, the transformation is calculated based on the relationship between the model grid and the matched deformed grid, and the bilinear interpolation is used for gray-scale values.

3) Gabor Phase-Based Evaluation: Assume that the matched model and image are locally similar to each other only when a small shift $\Delta\mathbf{x}$ is made. Due to the fact that a shift of an image in the spatial domain corresponds to a phase shift in the frequency domain, we can estimate this shift $\Delta\mathbf{x}$ by $\Delta\phi/\omega_k$. It is approximately true for Gabor filters under conditions that most of the energy of the Gabor filter is contained within the Gaussian envelope and constant relative bandwidth λ (see Section II-A) is maintained [24]. Our Gabor filter design meets all these requirements. Therefore, Gabor phase-based evaluation allows precise model/image alignment which, in turn, allows the estimation of pose.

Since phase difference is restricted to $(-\pi, \pi]$, the maximum image displacement estimated is limited to $(-\pi/\omega_k, \ \pi/\omega_k]$. Therefore, in Gabor wavelet representation, low-frequency filters can estimate large displacement with less accuracy, while high-frequency filters can estimate small displacement with more accuracy. However, phase-wrapping is expected when large shape or signature distortions are present. Phase-based measures will not be reliable in these situations.

The phase difference at center frequency ω_k and orientation ϕ_l between a model and an image probe is given as

$$\begin{aligned}\Delta\theta(\omega_k, \phi_l) &= \theta_m(\omega_k, \phi_l) - \theta_i(\omega_k, \phi_l)\\ &= \arctan\left(\frac{R_m^-[k,l]R_i^+[k,l] - R_m^+[k,l]R_i^-[k,l]}{R_m^-[k,l]R_i^-[k,l] + R_m^+[k,l]R_i^+[k,l]}\right)\end{aligned} \tag{15}$$

where R^+ and R^- are the *cosine* and *sine* components of the Gabor probe (3), respectively. Thus, the translational displacement in the direction of ϕ_l can be estimated as

$$d(\omega_k, \phi_l) = \frac{\Delta\theta(\omega_k, \phi_l)}{\omega_k}. \tag{16}$$

We want to come up with an average displacement measurement by using all available filter bands. To overcome noise, the displacement estimates by different filter frequency bands with the same orientation are first combined [24] as

$$d(\phi_l) = \frac{\sum_k w(\omega_k, \phi_l) d(\omega_k, \phi_l)}{\sum_k w(\omega_k, \phi_l)} \tag{17}$$

where local magnitude responses of the model $m_m(\omega_k, \phi_l)$ and the image $m_i(\omega_k, \phi_l)$ (indexed by filter frequency ω_k) are used as weight

$$w(\omega_k, \phi_l) = \min\left[\frac{m_m(\omega_k, \phi_l)}{m_i(\omega_k, \phi_l)}, \ \frac{m_i(\omega_k, \phi_l)}{m_m(\omega_k, \phi_l)}\right]. \tag{18}$$

Then, the image displacement estimated [8] for the probe p is computed as the weighted sum of estimations over all filter orientations (L).

$$d_x(p) = -\frac{\sum_{ij} p(\phi_i, \phi_j) d(\phi_j) \sin\phi_i}{\sum_{ij} p(\phi_i, \phi_j) \sin(\phi_j - \phi_i)} \tag{19}$$

$$d_y(p) = \frac{\sum_{ij} p(\phi_i, \phi_j) d(\phi_j) \cos\phi_i}{\sum_{ij} p(\phi_i, \phi_j) \sin(\phi_j - \phi_i)} \tag{20}$$

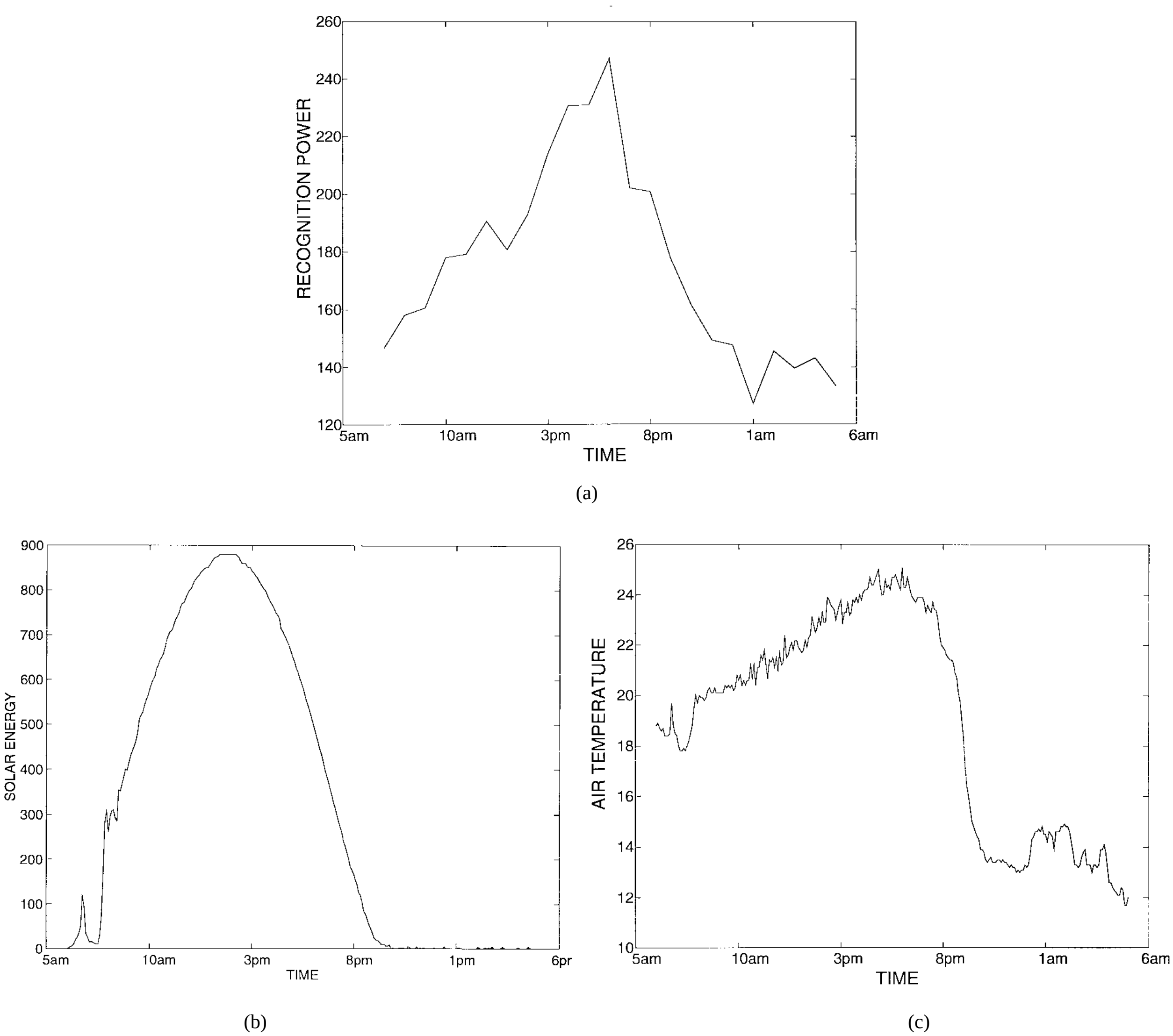


Fig. 10. Environmental parameter used to simulate the infrared object signature. (a) Solar energy and (b) air temperature changes for a period of 23 h are recorded on July 19, 1984, at Grayling, MI. (c) Recognition performance as measured by the "recognition power" for 23 infrared signatures of an object in a day under different weather conditions.

where ϕ is the filter orientation, $i, j = 0, \cdots, L-1, d(\phi_j)$ is obtained from (17), and $p(\phi_i, \phi_j)$ is computed as

$$p(\phi_i, \phi_j) = \min\left[\frac{d(\phi_i)}{d(\phi_j)}, \ \frac{d(\phi_j)}{d(\phi_i)}\right]. \tag{21}$$

Finally, the matching error for a probe p is estimated in terms of amplitude and direction as

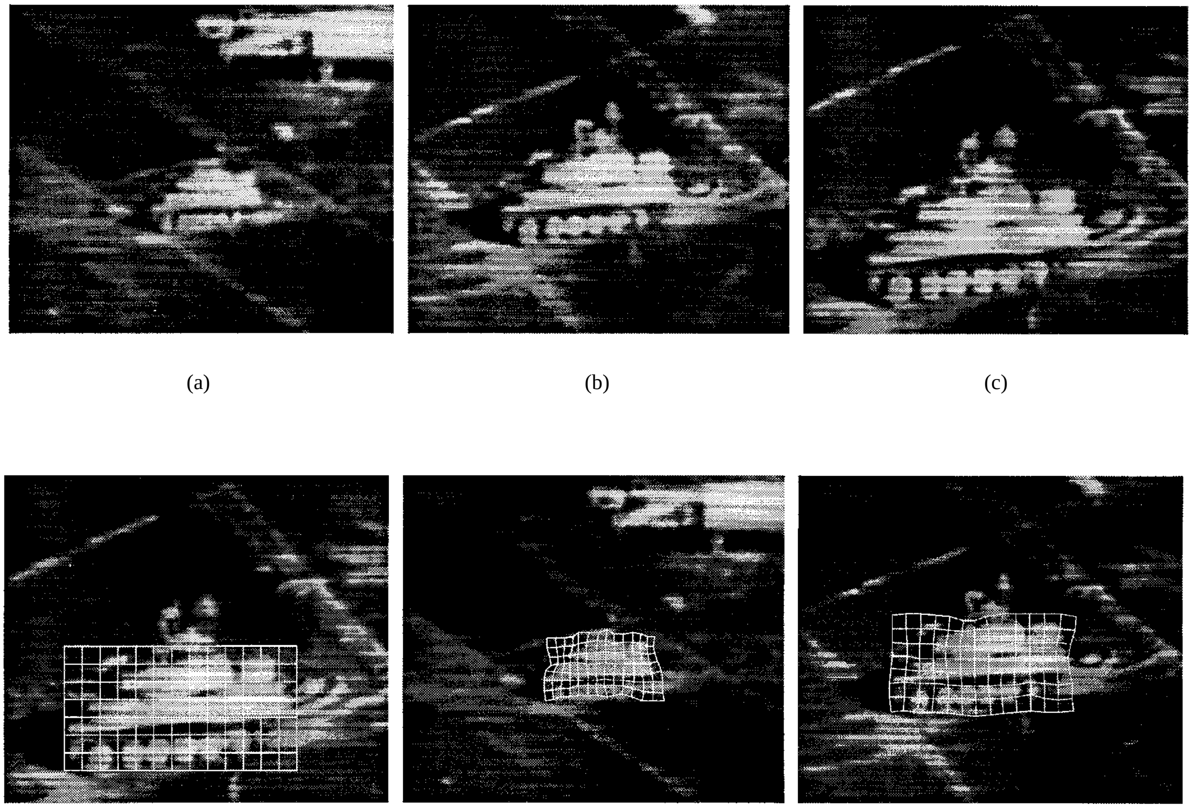
$$|\mathbf{d}(p)| = \sqrt{d_x(p)^2 + d_y(p)^2}, \tag{22}$$

$$\theta_{\mathbf{d}(p)} = \arctan\left[\frac{d_y(p)}{d_x(p)}\right]. \tag{23}$$

Due to the observation that grid nodes are not necessarily located at high Gabor magnitude response points, in our approach, points selected from the image with high Gabor magnitude responses are used and backprojected onto the matched model. More accurate phase measures can be obtained using these projected pairs than using grid nodes. An average of the local displacement estimated for all these points is used as the matching error δ for comparison. Similarly, the average pose can be estimated by using (23) for each probe.

Fig. 6 gives a quantitative illustration of how phase-based evaluation is used to estimate matching error. In this example, the aspect of the object and the model are 50° and 67.5°, respectively. Thirty-one points selected from the object image (Fig. 6(b)), which are local high Gabor magnitude responses, are backprojected onto the matched model (Fig. 6(a)) using the transformation computed by the matched deformed Gabor grid. The matching errors estimated for each of these points in terms of local image displacement are displayed by their direction and magnitude in Fig. 6(c). Many of these errors are associated with points in the front part of the object. This can be verified by overlapping the two images together as shown in Fig. 6(d). An average local image displacement of 3.4 pixels is obtained.

The success of phase-based evaluation is based on the fact that distinguishable features in the interior of the object, which are likely to be preserved under aspect and shape distortion,



(a) (b) (c)

(d) (e) (f)

Fig. 11. Object signatures in an image sequence and the matching results. The full-scale object in the image sequence (object 3) is selected as the model and matched with the other images in the sequence. (a) Object 1 (scale 52%). (b) Object 2 (scale 82%). (c) Object 3 (scale 100%). (d) Object 3 (model). (e) Object 1. (f) Object 2.

are available in the object image. Boundary points are sensitive to noise and background clutter and, therefore, they are not suitable for use in phase-based measures and we have not used them in this paper. However, since we are deriving models based on Gabor representation using sensor models and 3-D geometric models of the objects where boundary versus internal can be determined, this information can be identified and utilized in future research.

4) Recognizing Occluded Objects: Our approach for recognizing an occluded object can be described by *dynamic modification* of the Gabor grid through grid erosion and repairing processes performed during matching. The idea is to determine which subset of the model Gabor grid matches with the nonoccluded part of the object in the image. Since a Gabor grid encodes the localized signal energy and structural patterns of an object, the following two facts can be used to detect a potentially occluded object.

1) During flexible matching process, any collapsed grid results only due to object occlusion, since part of the model does not match with the exposed background corresponding to the occluded part of the object in the image.
2) A sub-Gabor grid corresponding to the occluded part has very low similarity measurement due to random matching with background clutter.

Thus, it is safe to assume that the part of the grid having a noncollapsed grid and high similarity matching result corresponds to the nonoccluded part of the object. Grid nodes from the hypothesized occluded part of the Gabor grid are discarded iteratively during dynamic grid refinement. The following processes are repeated until no more refinement is necessary: i) *model grid placement*, ii) *flexible matching*, iii) *grid repairing*, and iv) *new Gabor grid*. (See *Loop b* in Fig. 1). Fig. 7 shows how the connected and the complementary grids are used to detect and repair a collapsed grid. We define the *connected-grid* to be the same as its Gabor grid, and the *complementary-grid* is created by connecting the center of each block in the connected grid (the thick lines in Fig. 7(a)).

The *dynamic grid modification* process starts to erode the grid, cutting boundary nodes from the Gabor grid one by one, and calculating the grid placement again each time until either the collapsed and randomly matched part of the grid are removed, or the remaining grid moves to a new location without collapsing and random matching. The grid repairing algorithm is given below.

Grid Repairing Algorithm

1) Create the connected and complementary grids as graphs based on the deformed model Gabor grid.
2) Detect and remove collapsed and randomly matched grid nodes using the following rules:

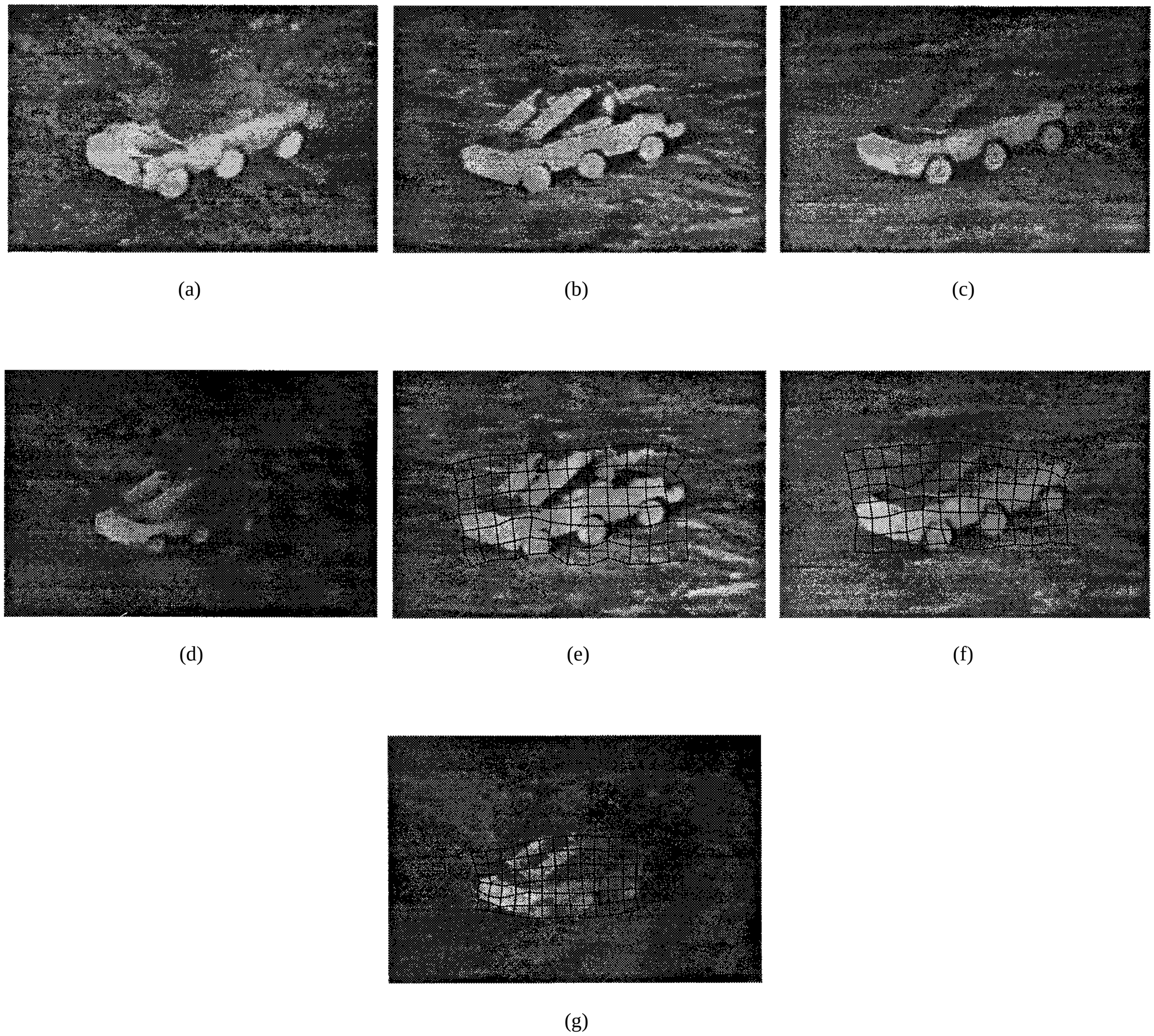


Fig. 12. Object model is matched with three objects having scale, aspect variations, and different signatures. The size of the images is 300 $\times$ 200 pixels. They are taken as the ROI from original images of size 512 $\times$ 512. (a) Model 1. (b) Object 1. (c) Object 2. (d) Object 3. (e) Second best match. (f) Best match. (g) Third best match.

a. A node is collapsed if the topological relationship between this node and its four-connected grid nodes in the connected or the complementary graph is broken. i.e., a collapsed node will result in "folding edges" in the Gabor grid.
b. A grid node that is randomly matched with background clutter will generally have small normalized similarity (in our experiments it is ≤ 0.3)

3) Remove collapsed grid nodes and corresponding edges according to their spatial locations and relationships. Remove isolated subgrids (nodes), and subgrids which are marked as randomly matched grid nodes.
4) Repeat the flexible matching process using this repaired Gabor grid, and evaluate matching result.

An example of occluded object recognition using synthetically generated infrared imagery obtained from the Physically Reasonable Infrared Signature Model (PRISM) infrared simulator [1] is given in Fig. 8. Fig. 8(a) shows an object with 40% occlusion. The initial matching resulted in a collapsed Gabor grid as shown in Fig. 8(b). Three subgrids that survived after grid erosion are shown in Fig. 8(c). After repairing, the remaining grid that matched with the nonoccluded part of the object is shown in Fig. 8(d). To determine which subgrid should survive, the *size* of the subgrid and the *average similarity* measurement of the subgrid are used to make a decision.

C. Evaluation of Matching

As we have seen, successful object recognition is based on matching of Gabor probes in a model grid with the probes obtained in an incoming image. Regardless of whether or not a corresponding object is in the model database, the process of matching always yields a best value for $\mathcal{C}$ in (10). Successful recognition tends to have small geometric distortions and high similarity measurements as defined by (10). However, a matching result for the correct object class may not be distinct enough when large object aspect variations and large changes in object signatures are present in the input data. In order to overcome the drawbacks of using only a single evaluation

TABLE III
STATISTICS OF RECOGNITION PERFORMANCE (AS MEASURED BY THE PERCENTAGE OF IMAGES CORRECTLY CLASSIFIED INCLUDING THE CORRECT POSE) FOR A TOTAL OF 207 IMAGES

Criteria used	Images	Failures	Percent Correct
$\mathcal{C}$	207	79	61.8%
$\varsigma+\varepsilon$	207	14	93.2%
$\varsigma+\varepsilon+\delta$	207	5	97.6%

ς: flexible matching cost, ε: dissimilarity measurement, δ: phase based average image displacement.

criterion (10), we introduce a set of comprehensive measures described below.

1) *Flexible matching cost* $\mathcal{C}$: It combines a measure of grid deformation and the similarity measure between probes. It is given by (10).
2) *Dissimilarity Cost* ε: It is defined as the difference between perfect matching and the actual matching results.

$$\varepsilon = \sum_{i}^{N} \left[1 - \bar{\mathcal{S}}(P_i^I,\, P_i^M)\right]^2 \tag{24}$$

where $\bar{\mathcal{S}}$ is the normalized similarity given by equation (13). The angle between two Gabor probe vectors is zero for perfectly matched probes, and the normalized similarity is 1. According to our experience, ε is less than 0.5 for a randomly matched probe pair.
3) *Displacement Cost* δ: It is defined as the average local translational displacement for all the matched Gabor probes.

To find the correct model after flexible matching, the results are evaluated based on the three criteria discussed above and the following rules in order.

1) For all matching results, sort each of the costs $\mathcal{C}$, ε and δ in descending order.
2) Select the model having both the lowest matching cost $\mathcal{C}$ and the smallest dissimilarity cost ε. If neither the values of $\mathcal{C}$ or ε for the top two matched models are distinguishable enough (by a predefined threshold), go to step 4).
3) Select the model having the smallest dissimilarity cost ε while its matching cost $\mathcal{C}$ and dissimilarity ε are both lower than a predefined threshold.
4) Select the model having the smallest displacement cost δ which is less than a predefined threshold.
5) Any matches that fail the above tests are rejected for recognition.

IV. OBJECT RECOGNITION EXPERIMENTS

A. Simulated Infrared Imagery

Synthetic infrared images are generated using PRISM [1], a well known infrared simulator on an SGI machine running IRIX4.0.5F. The main reason for using synthetic infrared signatures is to understand and to provide quantitative measures on how the recognition performance varies with changes in object signature caused by varying environmental conditions. We have investigated the effect of air temperature and solar energy on object signatures. The air temperature varied from 12°C to 26°C over a period of 23 h on July 19, 1984, at Grayling, MI (Fig. 10). A total of 18 models for one object are synthesized at 4 pm, with depression angles of 0° and 20°, and aspect angles from 0° to 180° with 22.5° separation. A total of 138 object signatures are generated for testing. They have depression angles of 10° and 30°, and aspect angles of 60°, 90°, and 120° and were obtained at one hour intervals over 23 hours. Four object signatures including background, with 60° aspect angle and 10° depression angle are shown in Fig. 9.

The *recognition-power*, defined as the difference of flexible-matching cost (10) between the best and second best matched models, is used to evaluate the performance of our recognition algorithm. The recognition performance shown in Fig. 10(c) is obtained from experiments on 23 infrared object signatures with three aspect angles and 0° depression angle. As expected, the recognition power is highest at the approximate time corresponding to the models. Also, the correlation between the curve of the recognition performance and the environmental parameter curves shown in Fig. 10(a) and 10(b) is clearly seen. It shows that as the contrast changes with the changes in time, recognition performance also changes. As a result, models to be used for matching signatures of objects in infrared images should be generated for specific environmental conditions. A model generated for a given environmental condition can accommodate only so much variation. For example, the model generated at 4 pm will be suitable (from approximately 8 am to 10 pm) when the recognition power is above a certain threshold (say, at 150 in Fig. 10(c)).

Table III shows the results of a total of 207 recognition experiments using the synthesized infrared models and images mentioned above These results include the following: i) 69 objects at depression angle 10° are recognized using object model sets at depression angle 0°; ii) 138 objects at depression angle 10° and 30° are recognized using object model sets at depression angle 20°. A success rate of 61.8% was achieved when only the flexible matching cost is used in matching. The

TABLE IV
RECOGNITION PERFORMANCE, AS MEASURED BY THE COST, DISSIMILARITY AND DISTORTION MEASUREMENTS, IN THE IMAGE SEQUENCE

Object	Grid Scale	Probe Scale	Cost	Dissimilarity	Distortion
Object 1	52%	50%	-78.7	1.65	47.6
Object 2	82%	71%	-93.4	1.38	28.3

TABLE V
RECOGNITION PERFORMANCE SHOWN AS CONFUSION MATRIX FOR SEVEN INFRARED TARGET SIGNATURES

	ASTRO	BTR_60	ZIL	SS_21R	SS_21L	TZM	SA_8
ASTRO	1						1
BTR_60		8			1		1
ZIL			3				1
SS_21R				3	1		
SS_21L				1	1		
TZM				1		2	
SA_8			1				7

performance is improved by using other evaluation criteria, and a successful recognition rate of 97.6% was achieved when the three evaluation criteria described in Section III-C (flexible matching, dissimilarity, and displacement) are used together. All the five matching errors are due to the foreshortening projection at 0° depression angle and 0° aspect.

B. Object Matching in an Image Sequence

Fig. 11(a)–(c) shows a sequence of tank images, where the wheel tracks of the tank on the ground form competing clutter. Both object and background clutter show high Gabor magnitude response in these images. Using the spatially tuned Gabor filters [9], the response of the periodic pattern such as the wheel track of the tank is enhanced, while the response of the background clutter is suppressed. The location of the object was correctly generated. Using these results, the flexible matching and tracking can start with a region-of-interest of the object and only a few object aspects. Table IV and Fig. 11(d)–(f) show the matching results. In this experiment, the full-scale image object 3 is used as the model to match with object 1 and object 2, with a grid scale of 52% and 82%, respectively. Using our multiscale model representation, the object scales are estimated and rounded to the closest probe scale which are 1/2 and $1/\sqrt{2}$, respectively. Flexible matching cost $\mathcal{C}$, dissimilarity ε, and grid deformation $\mathcal{D}$ are also given. This experiment illustrates the capability of combining spatial groupings of certain object features (periodic pattern in the detection phase) and object matching under scale, aspect, and image distortions.

C. Second Generation Infrared Imagery

1) Examples of Single Object: Fig. 12 shows an example where the object undergoes scale, aspect, and signature variations.

In this experiment, the distortion values (with respect to the model) for the best matched object correspond to 105 m in viewing distance, 11° in depression angle, and 52° in aspect angle. Note that the best match is the closet match to the model.

2) Examples of Multiple Objects: In the first experiment, four object classes with a total of 16 object aspects are extracted from the second generation infrared images, as the ROI's of the object for recognition. The four object class types are ASTRO, BTR_60, SS_21R, and SA_8, which were captured at Grayling, MI, in October 1992. These 16 objects are shown in Fig. 13, where the first object signature in each row is used to obtain the object model, and the rest of the images are used as test cases. Note that images show shape distortion due to aspect, scale, and depression angle variations, signature changes, and background clutters.

The best-matched object is selected as the instance of an object model. Ten of the 12 objects shown in Fig. 13 are successfully recognized. The two failures are for ASTRO 2 (Fig. 13(c)) and SS_21R 1 (Fig. 13(j)). The ASTRO 2 (Fig. 13(c)) is recognized as SS_21R model (Fig. 13(i)) and SS_21R 1 (Fig. 13(j)) is recognized as ASTRO model (Fig. 13(a)). These failures are due to severe signature and scale distortions with respect to the model. Only the first two evaluation criteria (flexible matching cost and dissimilarity cost) are used to obtain the matching results. Gabor phased-based evaluation is not used in this experiment because of the nondistinct interior structure of the object in the sensed data.

In the second experiment to recognize multiple objects, a total of 48 infrared images, similar to the ones shown in Fig. 13 in structure and complexity, are selected; 15 of these images are used to build object models to recognize objects in the rest of the 33 images. The confusion matrix is given in Table V and the recognition performance is 76%. The reason for the errors is significant distortion in shape and signature variations. Note that for the results shown in Section IV-AC, no significantly occluded objects (like the ones presented in the next Section, IV-D) were present in the data set. The data included some minor occlusions like the one shown in Fig. 13(m)–(p).

D. Recognizing Occluded Objects

An occluded object (with 35% occlusion) is selected from the second generation infrared image database, which is identified as an ASTRO, shown in Fig. 14. Although the grid in the initial matching results, Fig. 14(c), is not collapsed, the similarity of those grid nodes which matched with background clutter is relatively low with respect to the nodes which matched with the nonoccluded part of the object. These randomly matched nodes are detected and removed from further consideration. The final matching result is shown in Fig. 14(d). To illustrate the performance of matching under occlusion and distortion, the edge boundaries of the matched object model are backprojected onto the occluded object using the repaired distorted Gabor grid (Fig. 14). The quality of the result can be observed by comparing Fig. 14(e) and 14(f).

Another example of occlusion with a truck (ZIL, 30% occluded) is shown in Fig. 15. We have carried out other

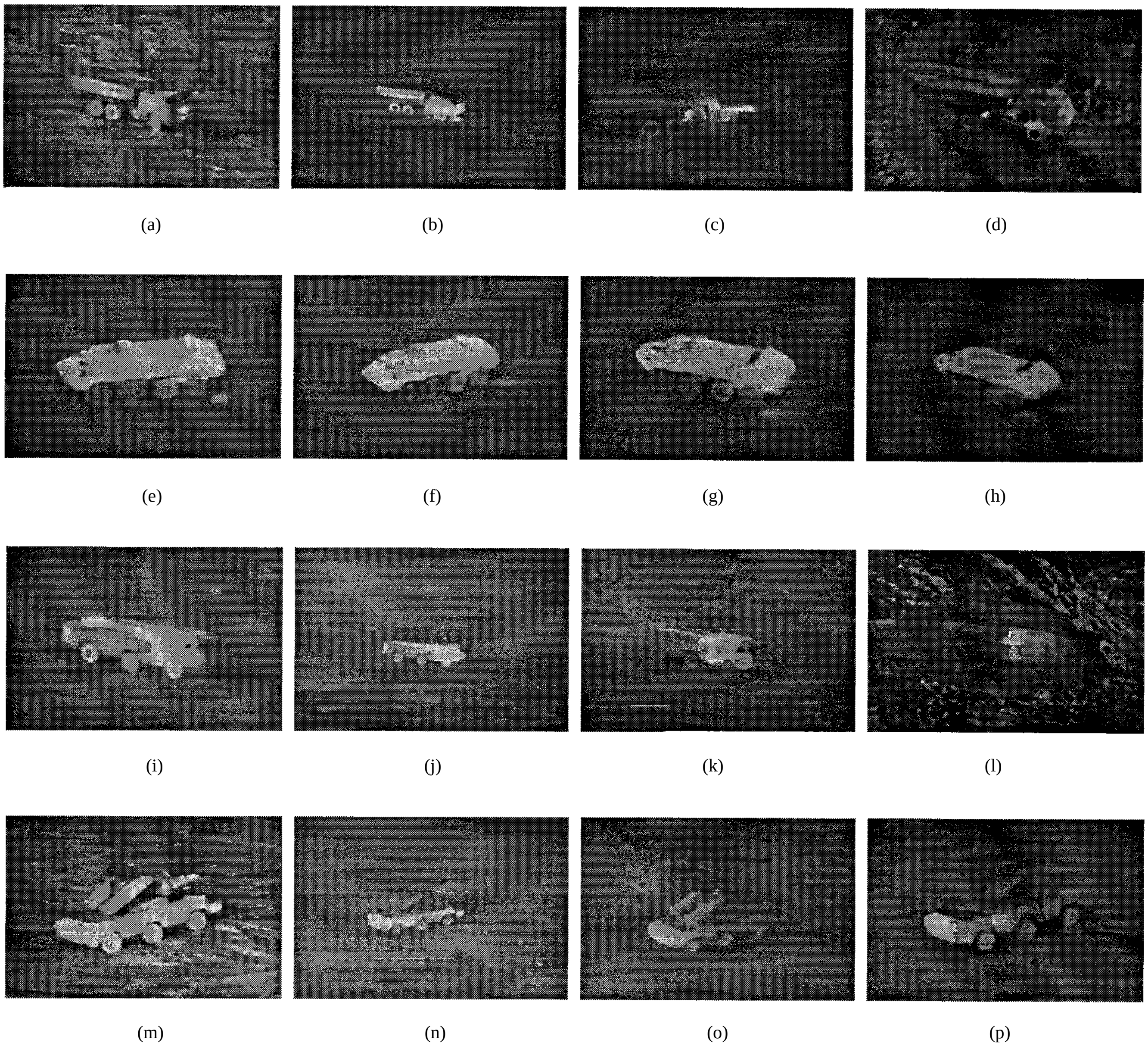


Fig. 13. Object model and images for ASTRO, BTR_60, SS_21R, and SA_8. (a) ASTRO model. (b) ASTRO 1. (c) ASTRO 2. (d) ASTRO 3. (e) BTR_60 model. (f) BTR_60 1. (g) BTR_60 2. (h) BTR_60 3. (i) SS_21R model. (j) SS_21R 1. (k) SS_21R 2. (l) SS_21R 3. (m) SA_8 model. (n) SA_8 1. (o) SA_8 2. (p) SA_8 3.

experiments with up to 50% occlusion and similar results were obtained.

V. Conclusions

We have shown that the multiscale Gabor wavelet-representation-based flexible matching technique that uses both Gabor magnitude and phase is a potentially robust method for object recognition under real-world conditions. The approach can tolerate variations of up to 20° in depression angle and 22.5° in aspect. The dynamic grid erosion and repairing allow the recognition of objects that may have up to 50% occlusion caused by natural objects or other man-made objects. Our approach can be used to recognize objects with the size varying from 256 pixels (16×16) to 16K pixels ($\sim 120 \times 120$) using a single scale object model and its multiscale Gabor wavelet representation. The scaling factor of our Gabor wavelet filters is $\sqrt{2}$. Since our wavelet filters consist of seven center frequency bands (corresponding to seven scales), so the scale variations that can be handled by our system is $(\sqrt{2})^7$. These scale changes from a given model are handled gracefully. In our experiments, the range for synthetic images has varied from 200 m to 800 m and, for real second generation infrared images, has been 180 m to 250 m (due to the availability of data). When the objects are far away and have only a few pixels, this approach, or for that matter any model-based recognition approach, will not be suitable. Although infrared signature of the object may change with changes in the environmental conditions, our approach can adapt to these changes to some extent. The use of both

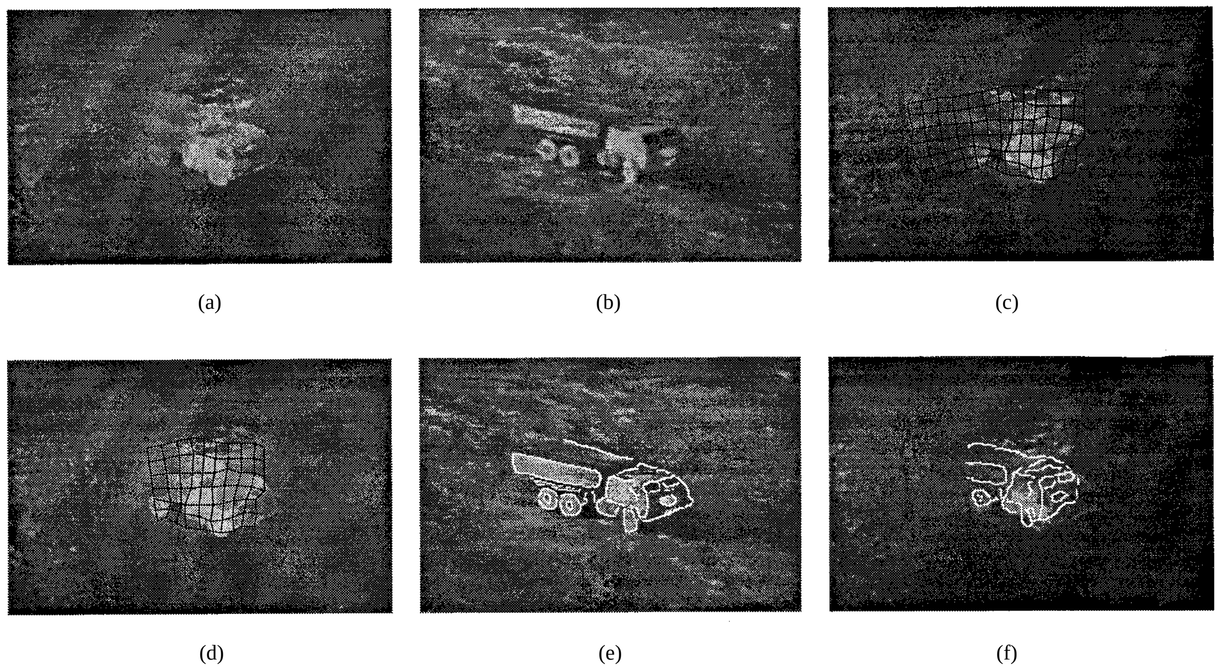


Fig. 14. Example of an occluded military truck (ASTRO) and the matching result. To illustrate the matching result, the edge boundaries of the model 14 are backprojected onto the occluded object 14 using the repaired distorted Gabor grid 14. (a) Object (ASTRO). (b) Object model. (c) Initial matching result. (d) Result of grid repairing. (e) Model edges. (f) Backprojected model.

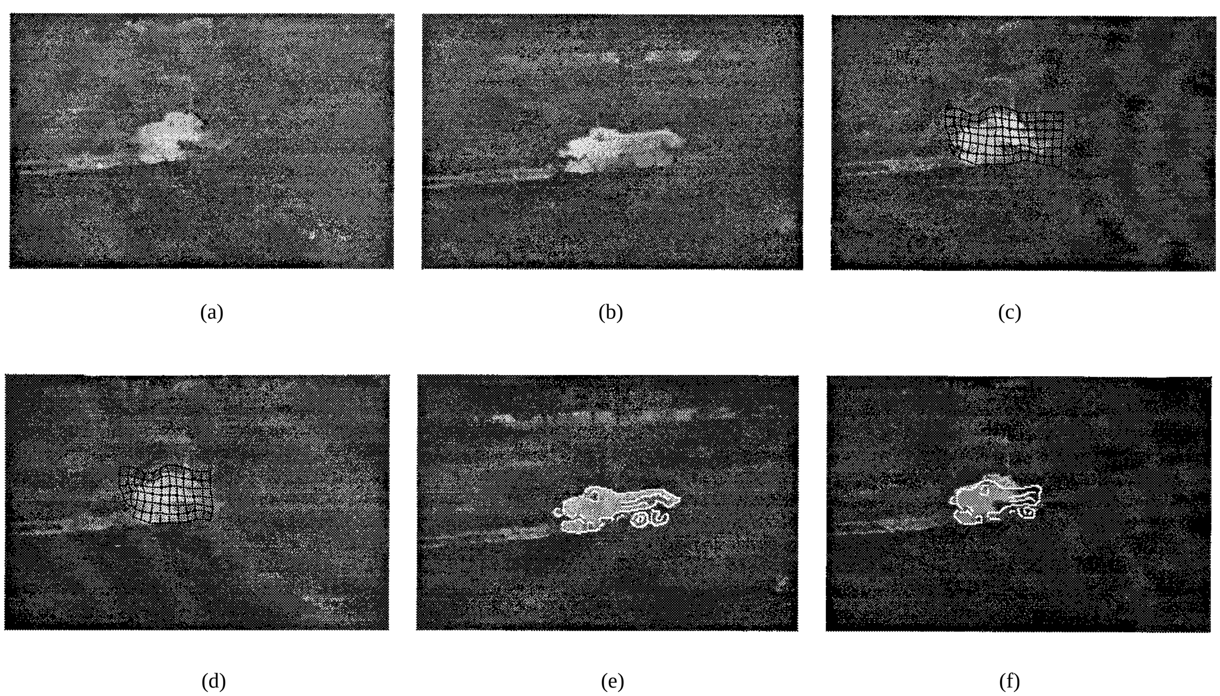


Fig. 15. Example of an occluded truck (ZIL) and the matching results. To illustrate the matching result, the edge boundaries of the model (e) are backprojected onto the occluded object (f) using the repaired distorted Gabor grid (d). (a) Object (ZIL). (b) Object model. (c) Initial matching result. (d) Result after grid erosion. (e) Model edges. (f) Backprojected model.

Gabor magnitude and phase help to improve the recognition performance when object signature varies with changes in environmental parameters, such as air temperature in the range of 12°C to 26°C. For an approach based on modeling context and clutter (time of the day, air temperature, range, depression angle) in infrared images, the reader is referred to [23].

Most of the computation time taken by the object recognition algorithm is for the Gabor decomposition obtained by using 56 filters. For an image of size 300×200, it takes several minutes on a SunSparc10 to finish the decomposition. We are investigating very large-scale integration implementation [5] of Gabor filters, which can dramatically reduce the cost of

Gabor image decomposition. Since the models are generated off-line and stored as Gabor grid, the complexity of matching is significantly reduced by our speed-up grid placement and flexible matching algorithm.

REFERENCES

[1] A. R. Curran, *PRISM 3.1 User's Manual*, Keweenaw Res. Ctr., Michigan Technol. Univ., Houghton, June 1993.

[2] R. Bajcsy and S. Kovacic, "Multiresolution elastic matching," *Comput. Vision, Graphics, Image Processing*, vol. 46, pp. 1–21, 1989.

[3] D. Ballard and L. Wixson, "Object recognition using steerable filters at multiple scales," in *Proc. IEEE Workshop Qualit. Vision*, June 1993, pp. 2–10.

[4] B. Bhanu, "Automatic target recognition: State-of-the-art survey," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, pp. 364–379, July 1986.

[5] B. Bhanu, B. Hutchings, and K. Smith, "VLSI design and implementation of a real-time image segmentation processor," *Mach. Vision Applic.*, vol. 3, no. 1, pp. 21–44, Jan. 1990.

[6] B. Bhanu and T. Jones, "Image understanding research for automatic target recognition," *IEEE Aerosp. Electron. Syst.*, vol. 8, pp. 15–23, Oct. 1993.

[7] I. Biederman, "Human image understanding: Recent research and a theory," *Comput. Vision, Graphics, Image Processing*, vol. 32, pp. 29–73, 1985.

[8] R. Braithwaite, "The use of the Gabor expansion in computer vision systems," Master's thesis, Univ. British Columbia, Vancouver, BC, Canada, Aug. 1989.

[9] R. Braithwaite and B. Bhanu, "Hierarchical Gabor filters for object detection in infrared images," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, Seattle, WA, June 1994, pp. 628–631.

[10] ——, "Error bound for multi-stage synthesis of narrow bandwidth Gabor filters," in *Proc. IEEE Int. Conf. Image Processing*, Washington, DC, Oct. 23–26, 1995, pp. 33–36.

[11] ——, "Target detection using phase-based Gabor element aggregation," in *Proc. ARPA Image Understanding Workshop*, Monterey, CA, Nov. 1994, pp. 511–514.

[12] P. Burt, "Smart sensing within a pyramid vision machine," *Proc. IEEE*, vol. 76, pp. 1006–1015, Aug. 1988.

[13] D. Casasent, J.-S. Smokelin, and A. Ye, "Wavelet and Gabor transforms for detection," *Opt. Eng.*, vol. 31, pp. 1893–1898, Sept. 1992.

[14] J. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1169–1179, 1988.

[15] D. Fleet, A. Jepson, and M. Jenkin, "Phase-based disparity measurement," *Comput. Vision, Graphics, Image Processing: Image Understand.*, vol. 53, pp. 198–210, 1991.

[16] D. Forsyth, J. Mundy, and A. Zisserman, "Transformational invariance: A primer," *Image Vision Comput.*, vol. 10, pp. 39–45, Jan. 1992.

[17] A. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognit.*, vol. 24, no. 12, pp. 1167–1186, 1991.

[18] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simmulated annealing," *Science*, vol. 220, pp. 671–680, 1983.

[19] M. Lades *et. al.*, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Comput.*, vol. 42, pp. 300–311, Mar. 1993.

[20] R. Mehrotra, K. Namuduri, and N. Ranganathan, "Gabor filter based edge detection," *Pattern Recognit.*, vol. 25, no. 12, pp. 1479–1494, 1992.

[21] M. Porat and Y. Zeevi, "The generalized Gabor scheme of image representation in biological and machine vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 452–468, July 1988.

[22] A. Resenfeld and A. Kak, *Digital Picture Processing*, 2nd ed. New York: Academic, 1982.

[23] S. Rong and B. Bhanu, "Modeling clutter and context for target detection in infrared images," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, San Francisco, CA, June 1996, pp. 106–113.

[24] T. Sanger, "Stereo disparity using Gabor filters," *Biolog. Cybern.*, vol. 59, pp. 405-418, 1988.

[25] S. Ullman and R. Basri, "Recognition by linear combinations of models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 992–1006, Oct. 1991.

[26] Z. Wang and M. Jenkin, "Using complex Gabor filters to detect and localize edges and bars," in *Advances in Machine Vision: Strategies and Applications*, vol. 32, C. Archibald and E. Petriu, Eds. River Edge, NJ: World Scientific, 1992, pp. 151–170.

Xing Wu received the B.S. degree in electrical engineering from Tsinghua University, China, in 1985, the M.S. degree in remote sensing applications from Academia Sinica, China, in 1988, and the M.S. degree in computer science from the University of California, Riverside, in 1994. His main research interests are in image understanding, object recognition, genetic learning, and remote sensing.

Since 1992, Mr. Wu has been actively involved in several major research projects supported by Honeywell, DARPA, and AFOSR.

Bir Bhanu (S'72–M'82–SM'87–F'96) For photograph and biography, see this issue, p. 4.