Neural Systems for Automatic Target Learning and Recognition

Allen M. Waxman, Michael Seibert, Ann Marie Bernardon, and David A. Fay

■ We have designed and implemented several computational neural systems for the automatic learning and recognition of targets in both passive visible and synthetic-aperture radar (SAR) imagery. Motivated by biological vision systems (in particular, that of the macaque monkey), our computational neural systems employ a variety of neural networks. Boundary Contour System (BCS) and Feature Contour System (FCS) networks are used for image conditioning. Shunting center-surround networks, Diffusion-Enhancement Bilayer (DEB) networks, log-polar transforms, and overlapping receptive fields are responsible for feature extraction and coding. Adaptive Resonance Theory (ART-2) networks perform aspect categorization and template learning of the targets. And Aspect networks are used to accumulate evidence/confidence over temporal sequences of imagery.

In this article we present an overview of our research for the past several years, highlighting our earlier work on the unsupervised learning of three-dimensional (3-D) objects as applied to aircraft recognition in the passive visible domain, the recent modification of this system with application to the learning and recognition of tactical targets from SAR imagery, the further application of this system to reentry-vehicle recognition from inverse SAR, or ISAR, imagery, and the incorporation of this recognition system on a mobile robot called the Mobile Adaptive Visual Navigator (MAVIN) at Lincoln Laboratory.

From the study of biological vision systems, we can learn much that applies to the design of computational neural systems for target recognition. These insights are most relevant to passive vision systems, such as visible and multispectral infrared imaging systems, but similar organizing principles are also useful in the radar imaging domain. In the next section, we summarize the primary lessons that have been learned from the anatomical, physiological, and psychophysical study of vision systems in the macaque monkey and man. These insights are then applied throughout the remaining sections of this review. (Note: An introduction to biological vision, learning, and memory can be found in the September 1992 special issue of *Scientific American,* which is entitled "Mind and Brain.")

Design Constraints from Biological Vision

The vision systems of primates contain two primary processing streams: the *parvocellular* stream, which processes shape information, and the *magnocellular* stream, which processes motion information (see References 1 and 2, and the references cited therein). Both streams begin in the retina and culminate in the parietal and temporal lobes of the cerebral cortex. Our automatic target recognition (ATR) systems have focused on the modeling of the parvocellular stream for the learning and recognition of three-dimensional (3-D) objects, although we have utilized image sequences to accumulate evidence over time. The image motion of objects can also be useful for recognizing potential targets, and we have developed

neurocomputational systems [3] to extract such information in real time (30 velocity fields per second) on the Pipelined Image Processing Engine (PIPE), a video-rate parallel-processing computer. The integration of an object's image motion with its shape information can potentially enhance the ATR process, and is a topic we are currently investigating.

The early visual processing that takes place in the retina, lateral geniculate nucleus, geniculo-cortical connections, and visual cortical areas V1, V2, and V4 of the occipital lobe are responsible for

1. conditioning imagery so as to render it invariant to the prevailing illumination (while producing smoothly shaded percepts of objects),
2. localizing features (such as edges, high-curvature points, and high-contrast points) that describe 2-D shapes, and
3. transforming the resulting feature pattern so as to render it invariant to object location, scale, orientation around the line of sight, and small deformation due to any foreshortening resulting from a rotation in depth (i.e., a rotation around an axis perpendicular to the line of sight), while still retaining measurements of these spatial attributes.

These invariant representations of 2-D object shapes make their way to the inferior temporal cortex via connections between the occipital and temporal lobes, whereas the location/scale/orientation information is relayed to the posterior part of the parietal lobe via connections between the occipital and parietal lobes. Object-location information is conveyed to the parietal lobe also via the superior colliculus, which receives direct connections from the geniculate nucleus and is intimately involved in attentional processes. These two cortical pathways—one subserving *object vision* (in the temporal lobe) and the other subserving *spatial vision* (in the parietal lobe)—have come to be known as the *what and where systems* [4]. Fusion of the *what* and *where* information is achieved via reciprocal connections between the temporal and parietal lobes, as well as by indirect connections between other regions of the brain such as the hippocampus, although the details are not yet understood.

Insight into the later stages of visual processing and 3-D object representation can be gained by studying the superior temporal sulcus (STS) in the temporal lobe of the macaque monkey. This area is known to be the site of cells tuned for the recognition of faces and other body parts. Of course, the faces that a monkey recognizes are indicative of the monkey's visual experiences, and reflect the visual learning process itself. We have learned much from the work of D.I. Perrett and his colleagues at the University of St. Andrews in Scotland [5–8].

The notion of cells specifically tuned to the recognition of certain objects (analogous to the orientationally tuned *edge sensitive neurons* in V1 discovered by D. Hubel and T. Wiesel in 1959) was popularized by H. Barlow in 1972, and became known as the *grandmother-cell hypothesis,* as if to emphasize that a single neuron becomes active to signal the recognition of one's grandmother. And, for the past 20 years, a debate has raged over this notion of single-cell versus distributed-network coding of visual objects. In fact, this seemingly absurd notion of single-cell coding seems to have much supporting evidence, as illustrated in Perrett's work below (and confirmed by other investigators). The strict notion of grandmother cells, however, must be reinterpreted in light of the fact that many layers of processing precede the view-specific coding of objects, and a hierarchical pooling of cells is required to influence the object-specific cell. Moreover, many visual objects may activate this cell, although it is maximally active for a specific object, whereas other cells are more active in the case of the other objects. Hence, a recognition decision must follow a neural competition between grandmother cells, and possibly an evidence-accumulation phase among multiple views when such views are available.

Figure 1 (from Reference 5) illustrates the STS area in the macaque monkey brain. The figure shows the locations of neurons detected by Perrett that are highly tuned to the face and profile views of heads, rotations of heads between specific views, and conjunctions of face views with up/down/left/right motions. Perrett's subsequent work [7] indicates the existence of view-specific cells, each one tuned for a particular view around a certain class of heads, and still other cells, called view-general cells, that respond to any view of a

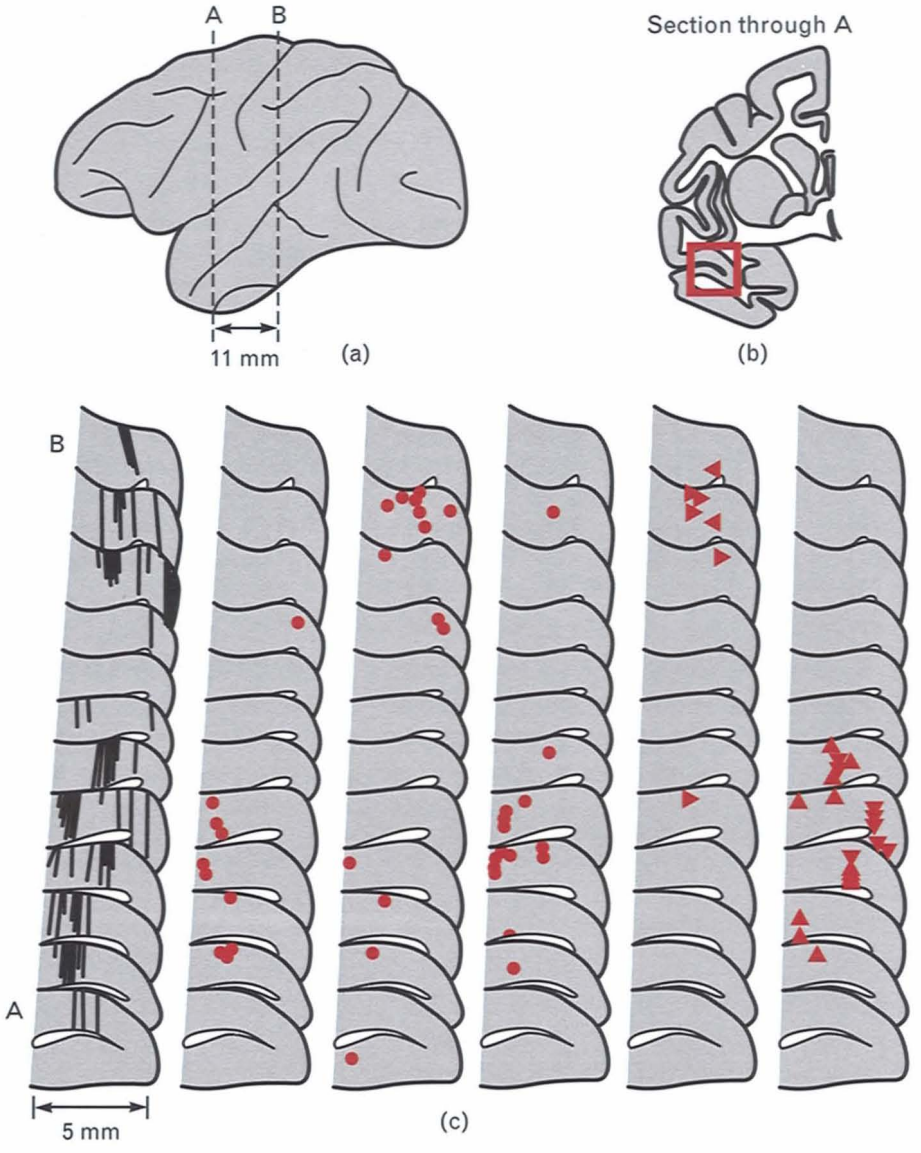


FIGURE 1. View-based coding of faces in the temporal cortex of the macaque monkey: (a) lateral view of the monkey brain, (b) coronal cross section with a red box around the superior temporal sulcus (STS), and (c) serial sections of the STS area investigated. From left to right, the sections illustrate the electrode tracks, cells selective to face views, cells selective to profile views, cells selective to transitions between views during head rotations, cells selective to faces moving left/right, and cells selective to faces moving up/down. (Adapted from D.I. Perrett et al. [5], with permission from *Trends in Neurosciences,* Elsevier Science Publishers B.V.)

specific head (as if the view-general cells were connected to all of the corresponding view-specific cells). View-specific cells respond to the same face views with similar activity levels, regardless of the illumination strength or color, the size or 2-D orientation of the face, and the position of the face in the field of view. Such cells have apparently learned 2-D-invariant shape codes.

Figure 2 (from Reference 6) provides a striking example of view and identity coding in the macaque temporal cortex. In the experiment, a monkey was shown different views of the faces of two familiar

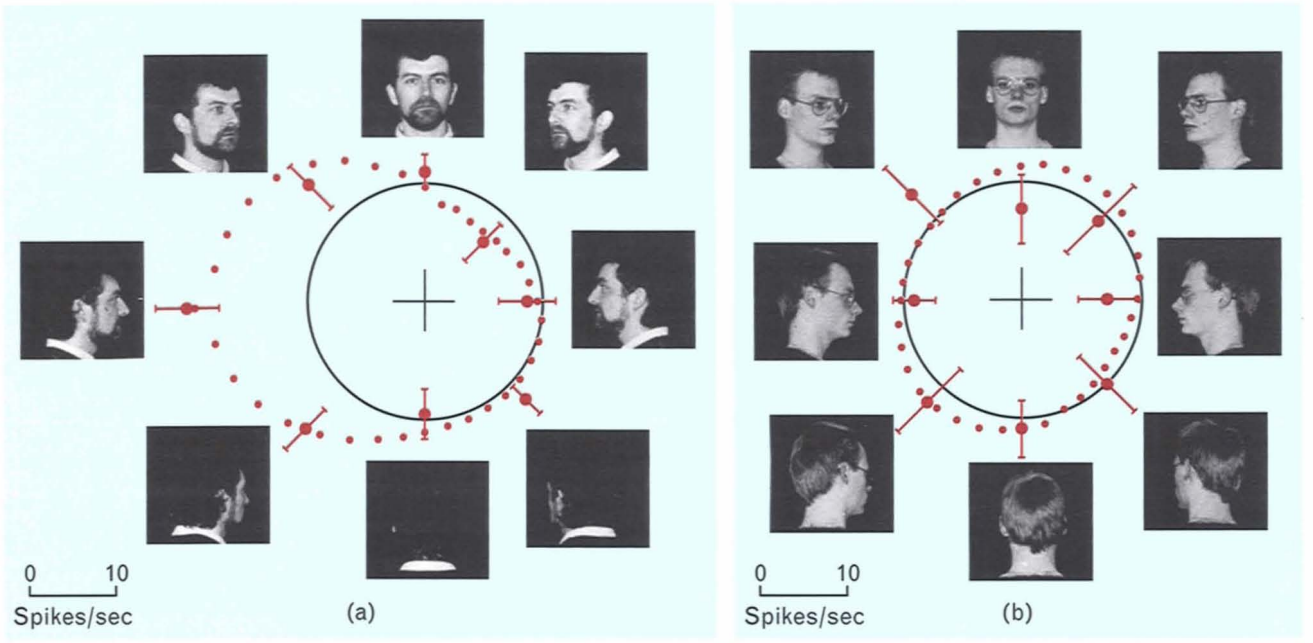


FIGURE 2. View and identity coding in the macaque temporal cortex for (a) subject 1 and (b) subject 2. In the experiment, a monkey was shown different views of the faces of two familiar people (subjects 1 and 2), and the activity of a single STS neuron in the monkey's brain was monitored with an electrical probe. The results are plotted in spikes/sec radially from the "+" symbol; the black solid circle denotes the spontaneous background activity level. The experimental measurements are represented by the large red dots with error bars indicating standard deviations over several repeated trials. Note that the neuron has a clear preference for the right profile view of subject 1, and no significant response to any view of subject 2. (Adapted from Perrett et al. [6], with permission from the *Journal of Experimental Biology*, the Company of Biologists Ltd.)

people, while an electrical probe monitored the activity of a single STS neuron in the monkey's brain. The results are shown in Figure 2, in which cell activity is plotted radially from the "+" symbol and the solid circle denotes the spontaneous background activity level. The experimental measurements are represented by the large dots with error bars indicating standard deviations over several repeated trials. Figure 2(a) shows that the neuron is highly tuned for the right profile view of subject 1. Nearby views (some at a 45° angle from the right profile) still generate cell activity, though at a much reduced rate. All views of subject 2 (a rather different-looking face) generate no significant activity above the background level, as shown in Figure 2(b). Thus this neuron might someday become a *grandfather cell!*

In summary, monkeys learn to recognize faces by employing a view-based strategy. Representations of 2-D shapes are learned and stored in view-specific STS cells. These cells code shape information that is invariant to illumination, position, scale, orientation around the line of sight, and small foreshortening deformation. Other cells code transitions between neighboring views that have been exposed by the rotations of a subject's head. A hierarchical combination of the two types of cells allows the construction of view-general cells that are selectively activated by specific heads regardless of the viewing direction. This same strategy for the learning and recognizing of 3-D heads (and, possibly, other objects) can be applied usefully to the design of artificial neural systems for ATR.

Aircraft Recognition from Visible Image Sequences

We designed our first end-to-end ATR system for the passive visible domain, and applied the system to high-contrast imagery of model F-16, F-18, and HK-1 (Spruce Goose) aircraft moving against textured backgrounds. (Note: Detailed descriptions of this neural system are contained in several papers by M. Seibert and A.M. Waxman [9–11].)

Figure 3 provides a conceptual overview of the system, in which a temporal view sequence of an object leads to the learning of an *aspect graph* [12] representation of that 3-D object. We can divide the system into three main functional stages, the first of which performs 2-D view processing to extract features (invariant to illumination) from the individual images, group these features to locate object position, and transform the features to render the pattern invariant to scale, orientation, and small deformation. The second stage takes these invariant feature patterns and clusters them into categories of similar views, or *aspects*. This 2-D view classification is done in an unsupervised way; i.e., it is strictly data driven without any category definition by a human. Along with the learning of these aspect categories, a prototype feature-pattern template is established for each category. The aspect categories correspond to the nodes of an aspect-graph representation of the target; they also play the role of view-specific cells for aircraft. The third stage detects the transitions over time between aspect categories (while the target is tracked in relative motion), learns these transitions, and accumulates evidence for possible targets. The learned transitions are like the arcs that connect the nodes in the aspect-graph concept, and are reminiscent of the STS neurons that are activated by the rotation of the heads between views in Figure 1. The ability to accumulate evidence over time is significant, for there are often cases in which a single view of a target is not sufficient to identify the target unambiguously; moreover, this fusion of evidence leads to a notion of

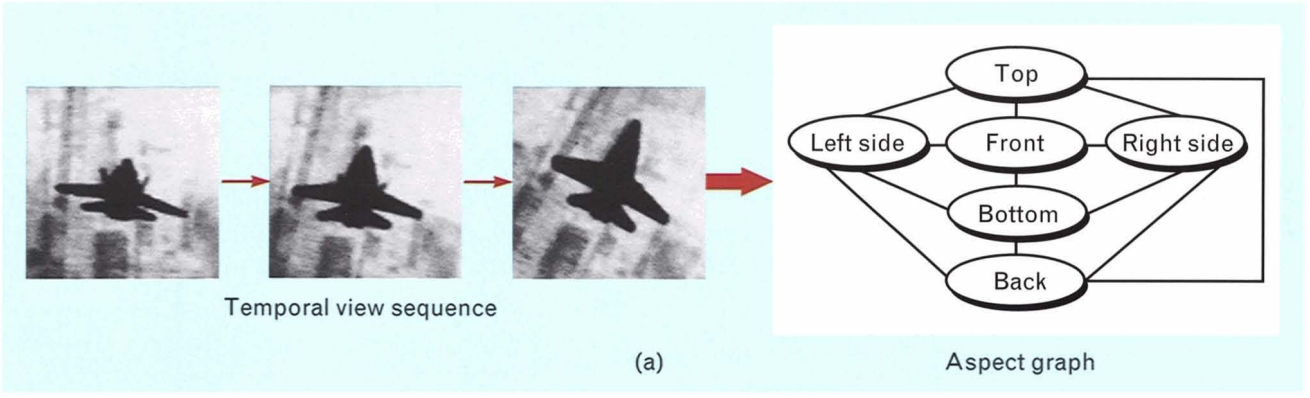


FIGURE 3. Conceptual approach of ATR neural system for passive visible image sequences: (a) temporal view sequence of images and corresponding aspect graph, and (b) functional block diagram of system. As a target moves relative to an observer, qualitatively different views are exposed in a temporal view sequence. The views unfold in an orderly fashion that is represented in the aspect graph. Each image in the sequence is processed by three stages of networks performing feature extraction and invariant mappings, classification of feature maps into aspect categories, and 3-D object evidence accumulation from the recognition of categories and transitions. The learned categories and transitions are analogous to the nodes and arcs, respectively, of an associated aspect graph.

confidence in the recognition decision.

These three processing stages can each be realized with multiple neural networks, and together the networks comprise a neural system architecture, as shown in Figure 4. Here, each module is an individual network that is annotated by the module's functional role in the system. Two processing streams are shown: the gray modules form a *parvocellular stream,* and the red modules form an *attentional stream.*

In the system, images are captured with a conventional CCD camera (which could be replaced by an infrared imaging system), and objects are segmented from the background by using a combination of motion and contrast information. Next, a *shunting center-surround network* enhances the edges of the segmented object, and a *Diffusion-Enhancement Bilayer (DEB)* extracts and dynamically groups the feature points of high edge curvature into a position centroid, as shown in Figure 5. These networks form nonlinear dynamical systems in which individual nodes are governed by (Hodgkin-Huxley–like) cell-membrane equations that resemble the charging dynamics of coupled resistor-capacitor networks. (See Reference 13 by S. Grossberg for a review of his pioneering work on dynamical neural networks, including shunting center-surround networks. Also, see References 14 and 15 for a reformulation of the DEB in terms of coupled dynamical layers of astrocyte glial-like diffusion cells and neural-like contrast-enhancing cells, all inspired by biology and applied to the psychophysical percept of long-range apparent motion.)

The centroid determined by the DEB network is used to track and fixate the object, and serves as the origin of a log-polar transform of the extracted-feature map. This transformation is very closely approximated by the axonal connections between the lateral geniculate nucleus and the primary visual cortex V1 [16]. In our system the transformation serves to convert changes in 2-D scale and 2-D orientation of the visual-feature map into a translation along new orthogonal axes. These processing steps are illustrated for an F-18 silhouette in Figures 6(a), (b), and (c). The log-polar feature map (periodic in orientation angle θ) is then input to a second DEB to determine a new feature centroid in the transformed coordinates. The spatial pattern of features now represents the original view of the F-18 invariant to illumination, position, scale, and orientation.

The next layer of processing indicated in Figure 4 consists of *overlapping receptive fields;* the processing is aligned with the centroid that was detected on the log-polar map, and serves to render the feature pattern somewhat insensitive to nonlinear spatial deformation. In the processing (Figure 7), a small array of Gaussian-weighted overlapping receptors are excited by the underlying features in the log-polar map, and the output of the array provides a much compressed code of the spatial feature pattern. (An individual receptor is activated by the feature within the receptor's field that lies closest to the field's center, and the feature's distance is coded according to a Gaussian falloff.) This compressed code is illustrated for the F-18 in Figure 6(d) for the case of a 5×5 array of overlapping receptors. In the figure, the sizes of the dots correspond to the receptor activation level: the larger the dot, the greater the activation. This coarse coding of spatial feature patterns simultaneously provides for enormous data reduction from the original target image (compared, for example, with a direct template-matching approach), leads to a tolerance for small deformations due to rotations in depth and inaccurate feature extraction, and yields an input vector for the classification network that forms the next system module.

The later stages of vision support the learning and recognition process. In our system, learning and recognition are realized by two modules consisting of an *Adaptive Resonance Theory* network (cf. several papers on various ART networks in Reference 17) and an *Aspect network* [10, 11].

Figure 8 illustrates the ART-2 architecture for unsupervised category learning and recognition. (Note: ART-2 is one implementation of Adaptive Resonance Theory for patterns consisting of real numbers.) The ART-2 network takes an N-dimensional input vector (in our case, the overlapping receptive field pattern with dimension of order 10 to 100) and first processes it through circuitry that contrast-enhances and normalizes the input as a short-term memory (STM) pattern. ART-2 then passes this pattern through a bottom-up filter (or template) stored in long-term memory (LTM) to excite a field of STM category

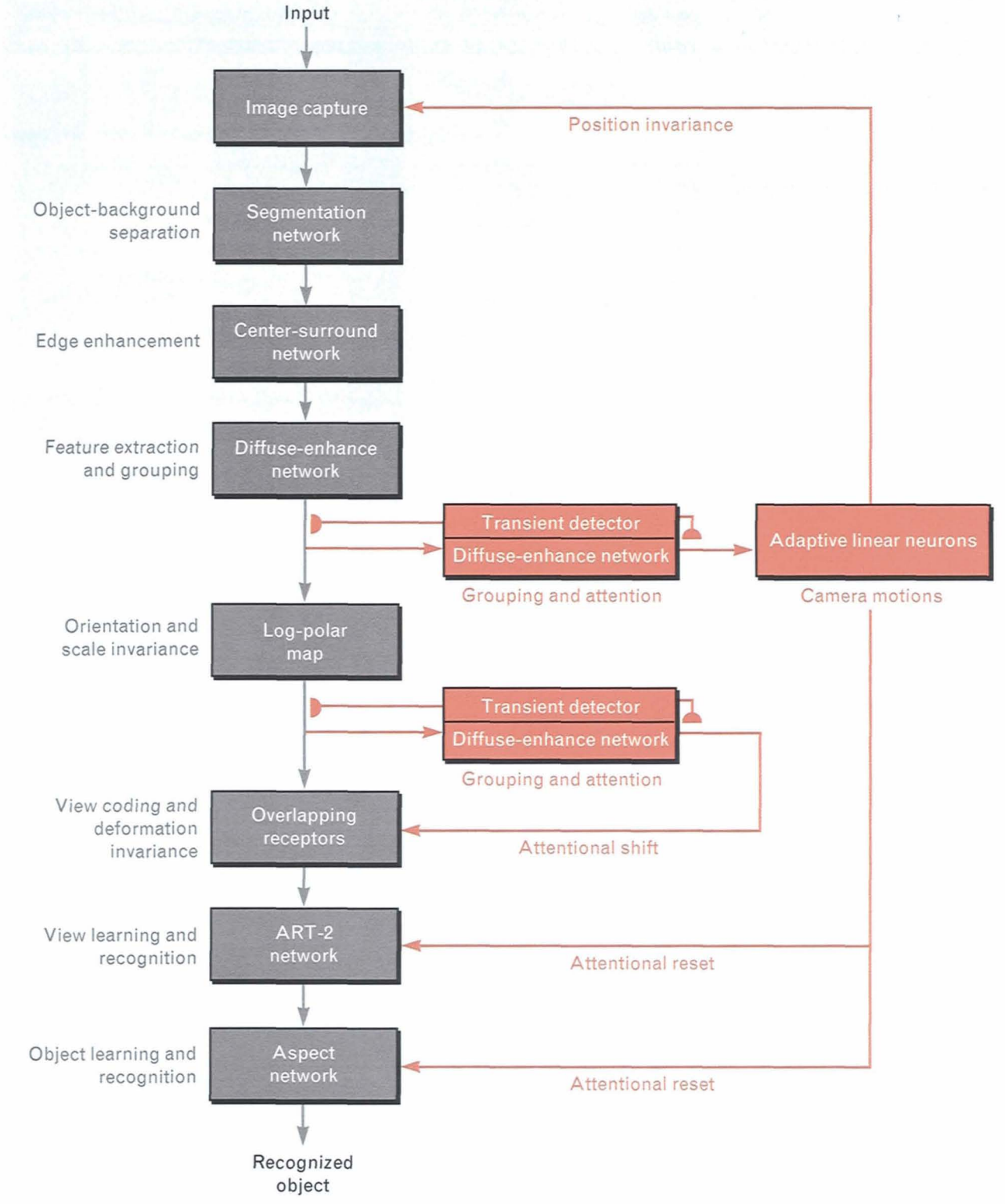
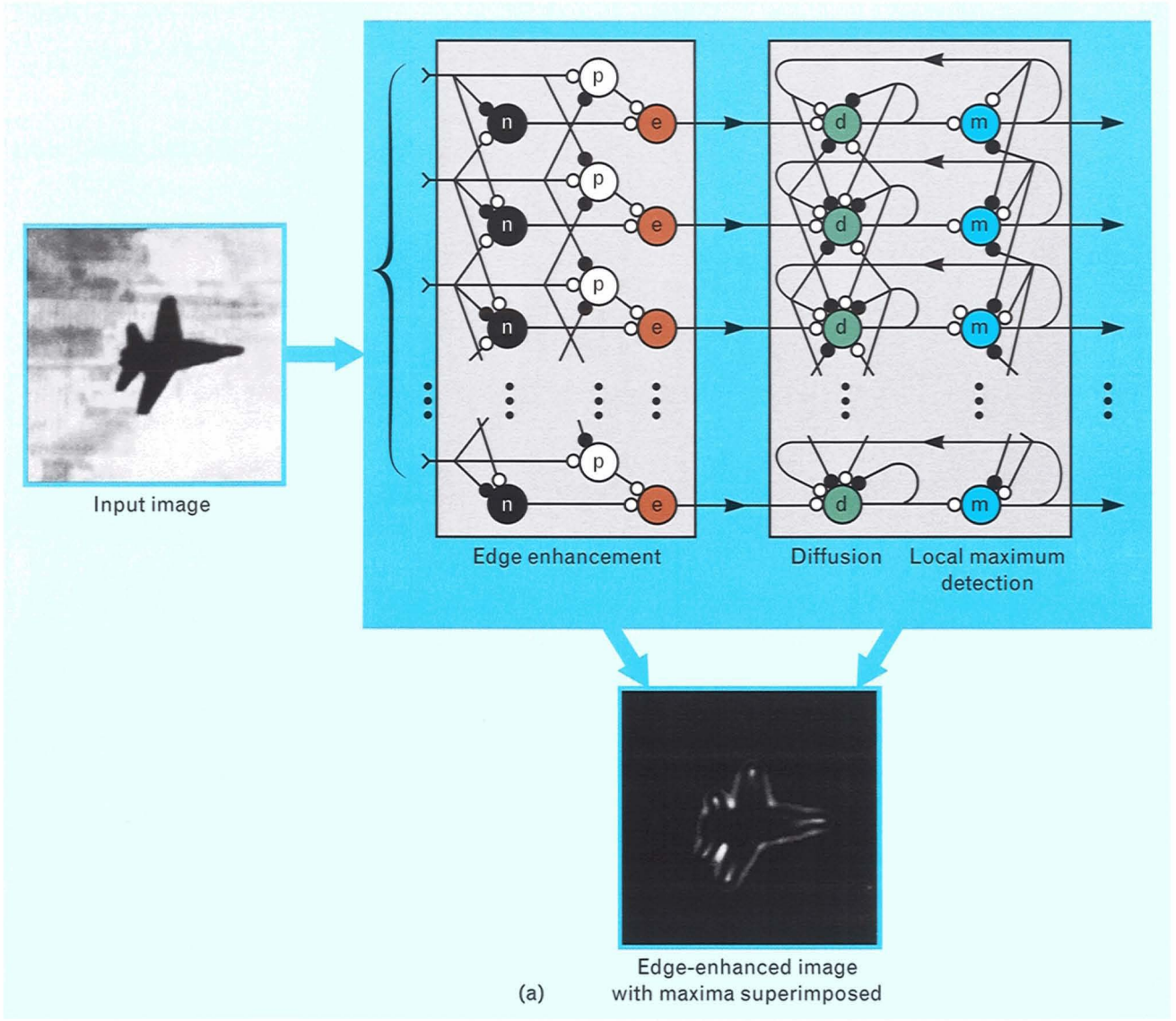


FIGURE 4. Modular system architecture for the learning and recognition of 3-D targets from visible imagery. The system is organized into two streams of neural network modules: the gray *parvocellular* stream for invariant shape learning and recognition, and the red *attentional* stream. The functional role of each module is indicated along with the type of network.



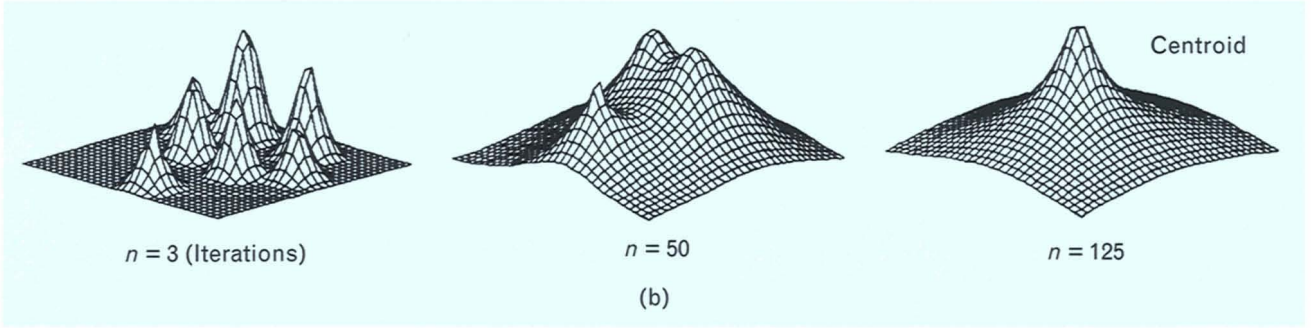


FIGURE 5. Diffusion-Enhancement Bilayer (DEB) for feature extraction and grouping: (a) architecture diagram and (b) evolving map of high-curvature points. The first stages of processing are accomplished by center-surround networks to edge-enhance the segmented object, and a diffusion-enhancement network to isolate points of high curvature along the silhouette. These feature points are dynamically grouped into a centroid (providing a focus of attention) by another DEB, which couples a diffusion layer to a contrast-enhancing layer in a feedforward and feedback configuration. (For a detailed description of DEBs, see References 9, 14, and 15.)

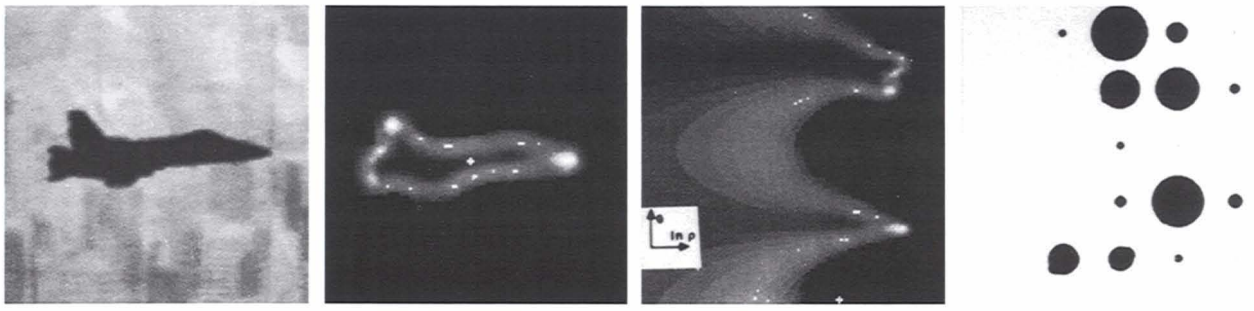


FIGURE 6. Stages in the processing of a 2-D view of a model F-18 aircraft: (a) the original image, (b) the edge-enhanced silhouette with DEB features superimposed and the centroid indicated with a "+," (c) log-polar mapping of the image in part *b*, with the new centroid indicated with a "+," and (d) the resulting output of a 5 × 5 array of overlapping receptive fields (see Figure 7) that forms the pattern fed to the Adaptive Resonance Theory (ART-2) network. In the image in part *d*, larger dots represent greater activity in the corresponding receptive fields.

nodes (our view-specific cells, or aspect nodes). These category nodes compete among themselves to choose a maximally activated winner, which in turn activates top-down feedback of a learned template also stored in LTM. This feedback represents the network's expectation of a specific input pattern. A vigilance parameter ρ (in the interval 0 to 1) that is set in advance by the user mediates the matching of the enhanced input pattern with the top-down template. Thus, simply having a best match among already established categories is not enough; rather, the best match must satisfy the established vigilance. When the match does satisfy the vigilance criterion, the network goes into a state of resonant oscillations between layers, and the bottom-up and top-down filters adapt slightly for better representation of the recent input pattern. When the vigilance criterion has not been met, the network generates a reset signal that flips the category field, thus suppressing the recent winner and reactivating the former losers. In this way, an uncommitted category node can establish a new category and a new template can be learned. ART-2 has several important attributes that make it particularly well suited to ATR applications: it supports on-line, real-time, unsupervised, stable category learning and refinement. We have utilized ART-2 successfully in a number of applications.

To present our results for the ART-2 classification of different aircraft, we introduce the concept of a *viewing sphere,* as illustrated in Figure 9. Note that a

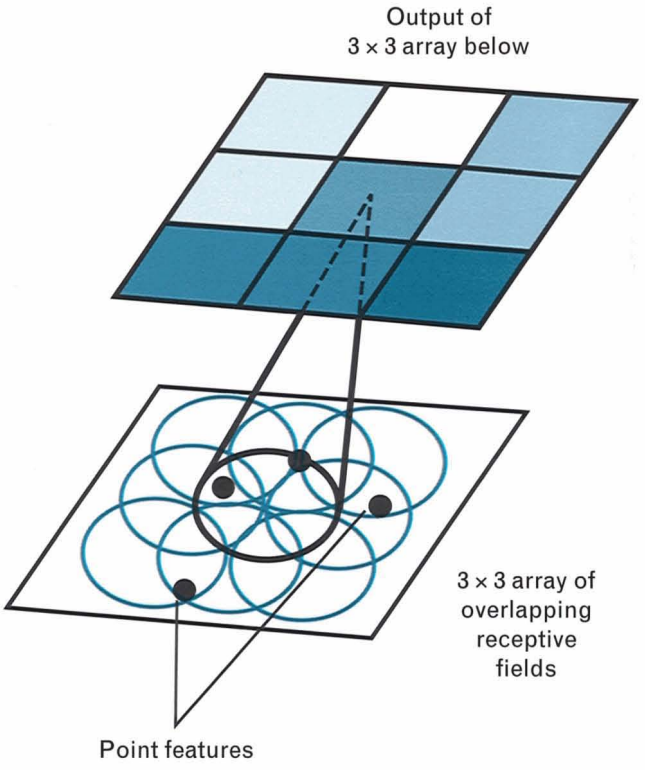
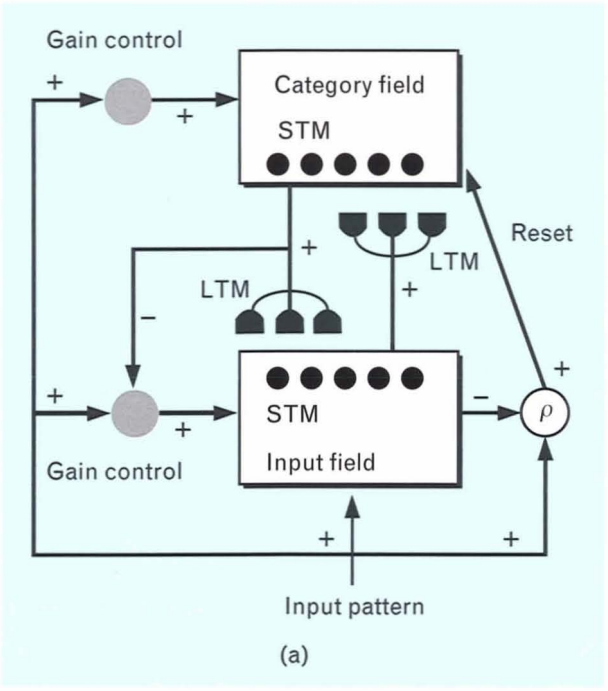
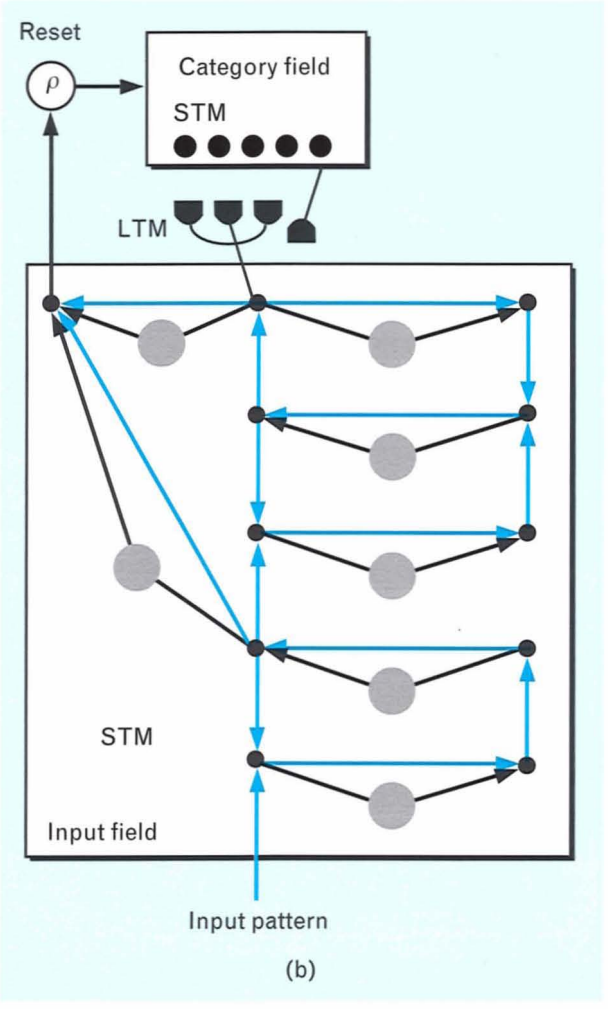


FIGURE 7. Spatial coding of features by overlapping receptive fields. Each circular field is activated according to a Gaussian-weighted distance to the point feature that is closest to the receptor center. (Note: Lighter colors in the figure represent closer distances.) These receptors provide enormous data compression, and they code spatial relations of features robustly with respect to deformations due to foreshortening. The fields convert a binary feature map to an analog pattern that is then suited for ART-2 classification.





location on the viewing sphere for an example aircraft corresponds to the view of that aircraft as seen from that particular direction. Using this viewing-sphere concept, Figure 10 summarizes the results of feature extraction, coding, and ART-2 classification for an F-18 model aircraft. With 535 input views of the F-18 and a vigilance ρ of 0.93, ART-2 generates 12 categories of the aircraft. In Figure 10(a), the categories, or aspects, are shown color coded on an *aspect sphere* with 12 different unrelated colors (i.e., a dark blue has no relation to a light blue). Note that the aspects subtend finite solid angles on the sphere (the target is oriented with its nose to the left). Because of object silhouette symmetry, only one quadrant of the sphere is shown. We can visualize example silhouettes that correspond to the 12 categories by selecting locations on the aspect sphere falling at the centers of each of the established categories, as shown in Figure 10(b). The corresponding silhouettes (numbered 1 through 12 in Figure 10[c]) represent prototype views that the system has created in an unsupervised manner. Notice the variety of silhouettes selected: some prototype views capture the wing shapes, some capture the double tail fins, some capture the dual exhausts, while others emphasize traditional top and side views. Also note the similarity between silhouettes 2 and 5, given the proximity of their corresponding centroids in Figure 10(b). Yet, although similar, silhouettes 2 and 5 do exhibit subtle differences, e.g., the differing slopes of the top portion of the visible tail fin. All of the views in Figure 10(c) were selected automatically. When the vigilance ρ was increased from 0.93 to 0.95, the ART-2 network generated 24 categories.

In addition to the F-18, we have also investigated

FIGURE 8. ART-2 network: (a) architecture and (b) circuit model. ART-2 takes analog input patterns and clusters them into categories by using unsupervised competitive learning. ART-2 can be trained on a dataset, then used to recognize data patterns in the field while continuing to refine its learned category representations (i.e., templates) stored in its adaptive synapses. The vigilance parameter ρ mediates the matching of the enhanced input pattern stored in short-term memory (STM) with a learned template from long-term memory (LTM). (Adapted from G.A. Carpenter et al. [17], with permission. This reference also contains a detailed description of ART.)

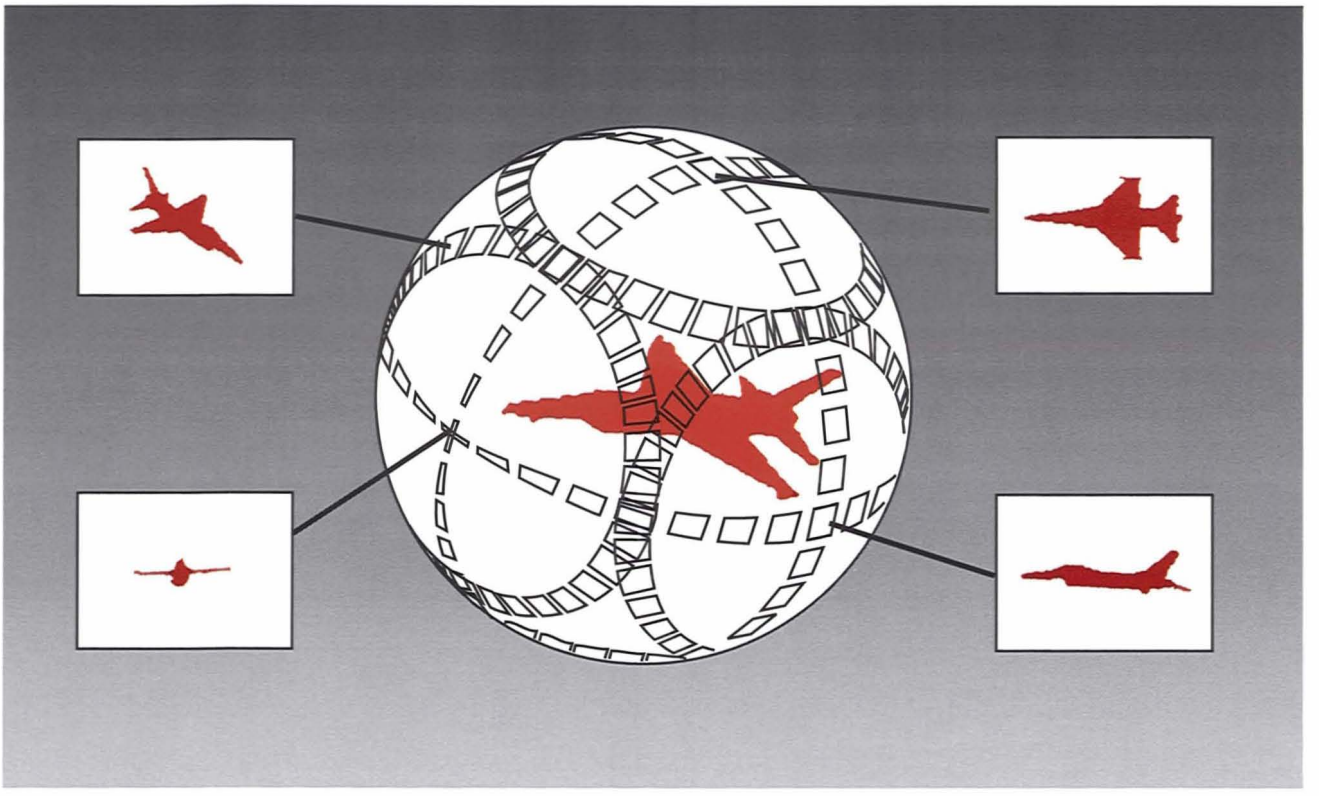


FIGURE 9. Example viewing sphere for a fighter aircraft. Note that a location on the sphere corresponds to the view of the aircraft as seen from that particular direction. Silhouettes of the aircraft are shown from different viewing directions. The silhouettes were obtained by applying thresholds to imagery that was captured with a charge-coupled device (CCD) camera and frame grabber. (The jagged contours reflect the finite pixel sizes of the CCD imager.)

ART-2 classification for an F-16 and HK-1 (Spruce Goose) model aircraft, as shown in Figure 11. Approximately 500 views of each aircraft were collected and processed with a single ART-2 module. (The next section, "Tactical Target Recognition in the Synthetic-Aperture Radar [SAR] Spotlight Mode," presents an alternative strategy of using one ART-2 module per target for the SAR application.) All the views together resulted in only 41 independent categories at a vigilance ρ of 0.93. (Note: Figure 10 investigated the categorization of a single target by using views of just that target. Thus, at the same vigilance setting of 0.93, ART-2 generated only 12 categories, in contrast to the 26 categories of Figure 11[a].) The individual aspect spheres show the similarity in category layout between the two fighter aircraft, and the obvious differences between fighter and transport-like aircraft. The spheres also indicate how certain views of the two fighters are ambiguous, at least in terms of the features extracted by the system. The individual 3-D targets are represented by roughly 25 categories each. Note in Figure 11 that some of the categories are common to two or more of the targets; i.e., the light yellow in Figure 11(a) corresponds to the same category that is represented by the same light yellow in Figure 11(b).

The aspect spheres in Figure 11 also illustrate the neighbor relations among categories as one rotates or explores a target in 3-D. These neighbor relations correspond to permitted transitions among categories, and are learned and exploited by our *Aspect network.* Much like the STS cells that code view transitions, and the hierarchical pooling of view-specific cells to form view-general object-specific cells, our Aspect networks self-organize into connections among aspect category nodes that preferentially channel activity into corresponding 3-D object nodes when successive aspects occur in a permitted sequence. The Aspect networks learn these aspect transitions incrementally during controlled training sessions, or ini-

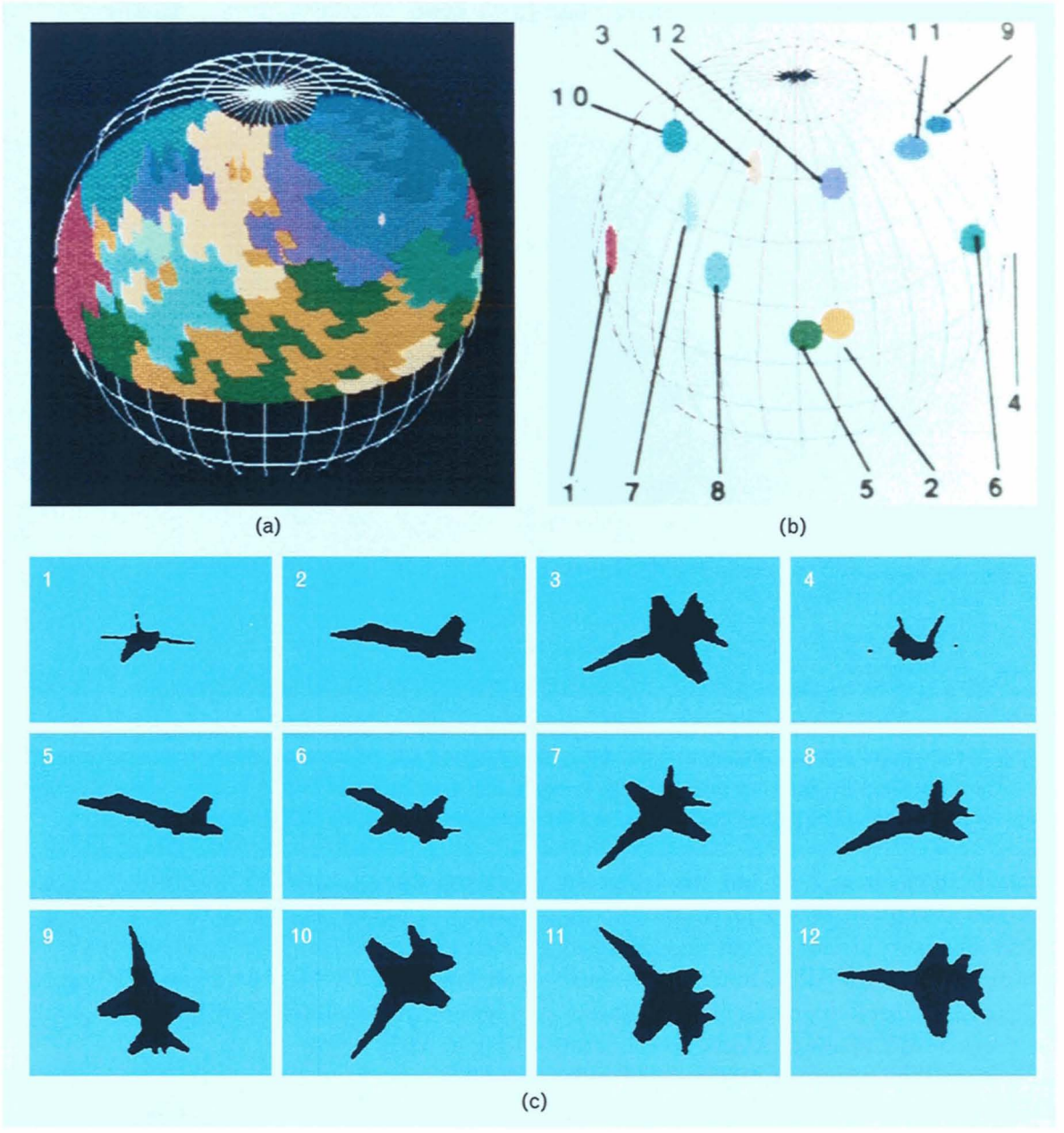


FIGURE 10. Results of feature extraction, coding, and ART-2 classification of an F-18 model aircraft alone at a vigilance ρ of 0.93: (a) aspect sphere showing the 12 aspects (color coded) generated by ART-2, (b) centroids of the largest regions of the 12 aspects, or categories, and (c) corresponding example silhouettes of the regions in part *b*. These views have been selected automatically by the system. (Note: The 12 colors used for the aspect sphere have been selected arbitrarily; i.e., a dark blue has no relation to a light blue.)

tially in the field after the aspect categorization has stabilized (i.e., after repeated exposures to the training data yield the same categorization). Then, during the imaging of a target in motion, multiple viewpoints are experienced, leading to recognition of multiple aspects by the ART-2 network, followed by evidence accumulation by the object nodes in the Aspect network. Target trajectories are realized as a set of aspect categories linked together by aspect transitions. Much more information becomes available when we consider the aspect transitions among ambiguous views. For example, even if both views of a two-aspect sequence are each ambiguous among potential targets, the additional aspect-transition information is often

sufficient for the preferential activation of the correct target node in the Aspect network.

Figure 12 illustrates an Aspect network for a single object, along with an enlarged view of the network's *adaptive axo-axo-dendritic synapse.* This synapse brings together in close physical proximity projections from pairs of aspect nodes onto a branch of the dendritic tree leading to an object node. When ART-2 categories are excited in temporal succession, the aspect nodes shown charge or discharge exponentially like capacitors, and their temporal overlap of activity supports a Hebbian form of correlational learning on the connecting synapse (cf. Reference 13 for a discussion of modified Hebbian learning with gated decay). The synaptic weights lie in the interval [0,1], and, as category transitions are experienced, the weights asymptotically approach the extreme values of 0 (implying no allowed transition between corresponding categories) and 1 (indicating a permitted transition). These values correspond to the absence or presence of an arc in the associated aspect-graph representation. The dendritic tree with its synaptic connections resembles the symmetric state-transition matrices that are commonly used in system-modeling techniques.

Extending the Aspect-network concept to multiple targets leads to the network architecture shown in Figure 13. In this design we consider all aspect categories of all targets as belonging to the same ART-2

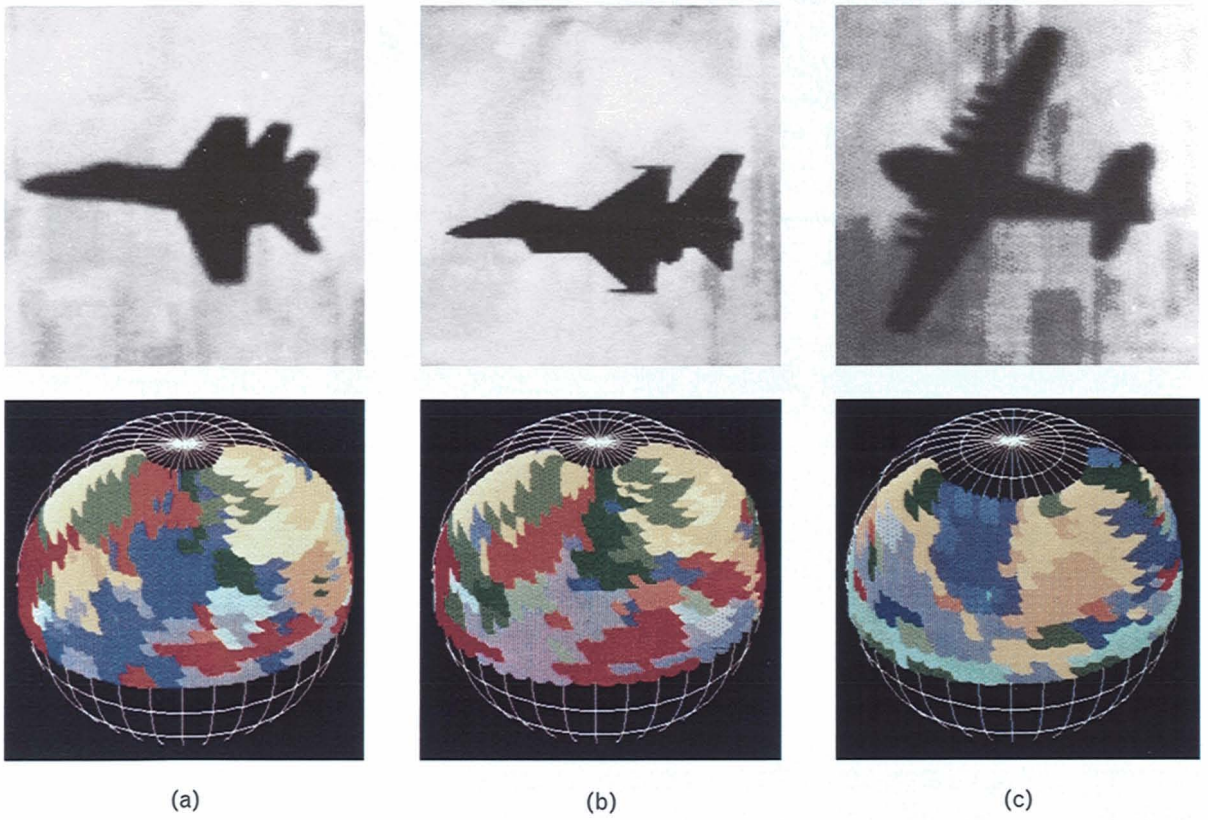


FIGURE 11. Aspect spheres for the (a) F-18, (b) F-16, and (c) HK-1 (Spruce Goose) have been generated from 535, 530, and 423 views, respectively, of each aircraft. Feature extraction, invariant mappings, and ART-2 categorization of all 1488 views generate a total of 41 aspects, or categories, at a vigilance ρ of 0.93. The number of categories generated for the individual aircraft is 26 for the F-18, 24 for the F-16, and 28 for the HK-1. Note that many categories are common to more than one target; i.e., the light yellow in part *a* corresponds to the same category that is represented by the same light yellow in part *b.* Also note the resemblance of the aspect spheres for the two fighter aircraft, in contrast to the HK-1 aspect sphere.

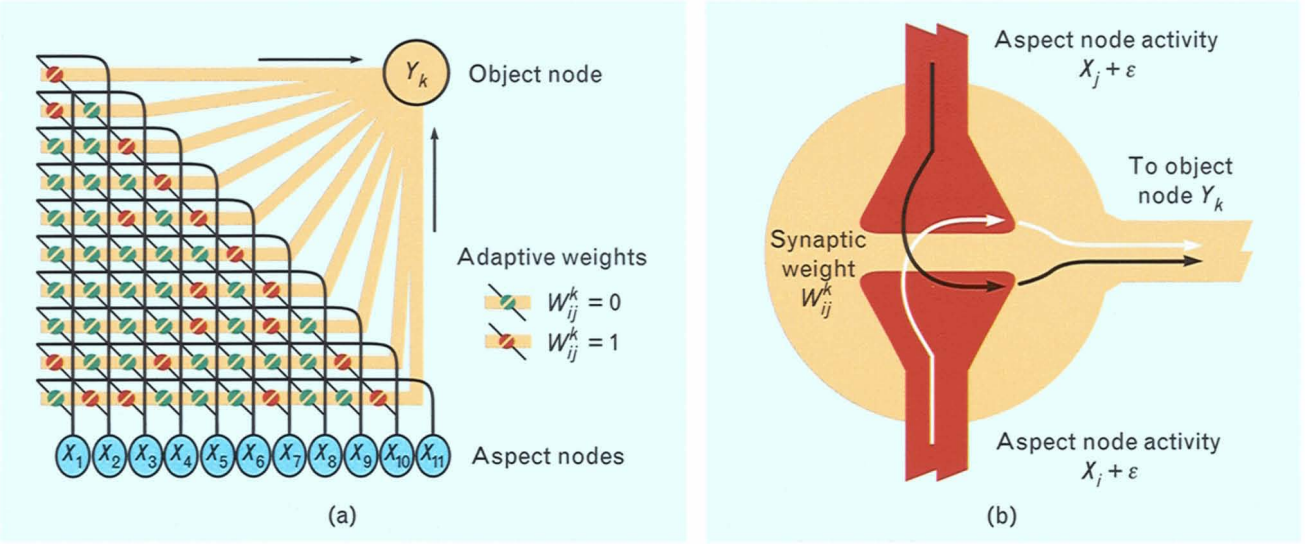


FIGURE 12. Aspect network for the single-object case: (a) network and (b) enlarged view of one synapse of the network. The aspect nodes (blue) are each coupled to corresponding categories allocated by the ART-2 network; the nodes charge and decay like capacitors. Axons (wires) emanating from each aspect node cross each other to form a transition matrix, and each crossing has an associated axo-axo-dendritic synapse (red) onto the dendritic tree (orange) of the object node. When two aspect nodes are simultaneously active (during view transitions), they strengthen the synapse (red) via modified Hebbian learning, and conduct activity onto the dendrite toward the object node. Object nodes thus pool activity from aspect nodes, exploiting transition information to amplify this activity, thereby accumulating evidence over time. In the enlarged view, the synapse brings together activity from aspect nodes X_i and X_j (as well as a background-noise level ε) and channels it onto the dendritic tree. (Note: The box "Aspect Network Learning Dynamics" contains a description of the equations that govern the aspect nodes, object nodes, and synaptic weights. For further details of Hebbian learning and Aspect networks, see References 10, 11, and 13.)

network. The aspect categories of the ART-2 network drive a single set of aspect nodes that fan out to all the synaptic arrays of possible targets. Activity (i.e., evidence) is then channeled into the object nodes, which compete to select the target with the maximum evidence at that moment. The winning object is then able to modify its own transition array. Sudden saccadic eye/camera motions to other locations in a scene initiate a reset of object-node activities to zero; smooth tracking motions do not cause such resetting.

Figure 14 contains an example of aircraft recognition by the Aspect network. In the training sequence, each of the three model aircraft experiences an identical trajectory of 2000 views covering one quadrant of the viewing sphere. Then a test sequence of 50 F-16 images is generated, and evidence is accumulated for each of the three targets as well as for an unlearned other target representing a none-of-the-above category. The graphs shown in the figure illustrate the corresponding category (and transition) sequence, the evidence accumulation and decay for each possible target, and the winning object with the instantaneous maximum evidence. Note that initially the system begins selecting the "other" target until sufficient evidence accumulates to declare the F-16 the winner, and it remains so. Reference 11 contains further details of this experiment.

At this point we have the basic design of a neural ATR system. The system has a number of definite strengths, but it also suffers from a few shortcomings. For example, a difficulty exists in adding new targets once the system has stabilized, because new data may modify the existing ART-2 category templates and lead to the need to retrain the Aspect network. A more efficient design is to assign a separate ART-2 network and (much compressed) Aspect network to each potential target, but allow the unsupervised assignment of aspect categories during the controlled exposure in a training session. By doing so, we can add new targets at a later time by simply adding new

ASPECT NETWORK LEARNING DYNAMICS

WE DEVELOPED the Aspect network (Figures 12 and 13) as a means to fuse recognition events over time. The network embodies a hierarchical pooling of view-specific aspect categories so as to exploit the additional information associated with permitted category transitions. These transitions are learned by exploring the object.

The dynamics of Aspect networks is in the form of differential equations (shown below) governing the short-term memory activity of the aspect nodes X_i and object nodes Y_k, and long-term memory of the adaptive axo-axodendritic synapses W_{ij}^k. Aspect nodes are excited by their corresponding ART-2 category nodes I_i (with rate constant K_X) and passively decay back to their resting state (with rate constant λ_X).

Object nodes accumulate evidence for each object by summing the activity (with rate constant K_Y) entering from the aspect nodes on the dendritic tree. Activity riding atop background noise ε enters via the learned synapses corresponding to permitted transitions, and activity is channeled most effectively by paired aspect nodes in a permitted sequence. The function $\Phi_B(A)$ is a threshold linear function that passes activity levels when $A > O(B)$. Similar to the aspect nodes, the object nodes also decay passively to their resting state (with rate constant λ_Y).

The synaptic weights learn aspect transitions by experiencing correlated activity from two aspect nodes, as long as the object node activity is changing (i.e., $\dot{Y}_k \neq 0$) for the winning object Z_k. The function $\Theta_\varepsilon(C)$ is a binary threshold gate that equals unity when $C > O(\varepsilon)$. The weights approach asymptotes toward the fixed points of 0 and 1 because of the quadratic shunting terms that modulate the rate constant K_W. For further details of Aspect networks, see References 1 and 2.

References

1. M. Seibert and A.M. Waxman, "Learning and Recognizing 3D Objects from Multiple Views in a Neural System," chap. II.12 in *Neural Networks for Perception, Vol. 1,* ed. H. Wechsler (Academic Press, New York, 1991), pp. 426–444.
2. M. Seibert and A.M. Waxman, "Adaptive 3-D Object Recognition from Multiple Views," *IEEE Trans. Pattern Anal. Mach. Intell.* **14,** 107 (1992).

$$\text{Aspect nodes:} \quad \frac{dX_i}{dt} = K_X I_i - \lambda_X X_i$$

$$\text{Object nodes:} \quad \frac{dY_k}{dt} = K_Y \left\{ \sum_i \sum_{j>i} \Phi_{\varepsilon^2}\left[\left(X_i + \varepsilon\right) W_{ij}^k \left(X_j + \varepsilon\right) \right] - \lambda_Y Y_k \right\}$$

$$\text{Synaptic weights:} \quad \frac{dW_{ij}^k}{dt} = K_W W_{ij}^k \left(1 - W_{ij}^k\right) \left\{ \Phi_\varepsilon \left[\left(X_i + \varepsilon\right)\left(X_j + \varepsilon\right) \right] - \lambda_W \right\} \Theta_\varepsilon(\dot{Y}_k) \Theta_\varepsilon(Z_k)$$

ART-2 and Aspect networks, without any modification to the existing networks. Moreover, separate ART-2 networks for each target better support the ATR task given only a single view (as opposed to a sequence of views), because each target will have generated its own set of learned templates within its ART-2 module. This design has been adopted for the next application—target recognition from SAR spotlight sequences. For this application we also introduce a measure of *recognition confidence* derived from the accumulated evidence.

Tactical Target Recognition in the Synthetic-Aperture Radar (SAR) Spotlight Mode

High-resolution radar imaging of a scene can be accomplished by flying a radar that is transmitting chirp pulses from many closely spaced look angles (Figure 15). The moving radar thus synthesizes a long aper-

ture, and the return pulses determine a reflectivity image of the scene as projected into the range and cross-range coordinates of the plane formed by the synthetic aperture and the radar line of sight. (This plane is referred to as the *synthetic-aperture radar [SAR] slant plane.*) The range resolution is proportional to the bandwidth of the chirp pulse; the cross-range resolution is proportional to the angle subtended by the synthetic aperture. As the radar moves along the flight path, it can be "squinted" so as to track a fixed location on the ground. Hence, the radar beam *spotlights* a particular scene, and a sequence of SAR images is obtained of that scene from multiple views.

The reader may look ahead to Figure 26(a) to view a typical clutter scene—an overpass that crosses the New York State Thruway—obtained from the Lincoln Laboratory Advanced Detection Technology Sensor (ADTS), a millimeter-wave radar, operating in the SAR mode. (In our work, only single-channel vertical-vertical [VV] polarization imagery is used.) Note that the image is quite speckled, a consequence of the coherent imaging method. Nonetheless, at first glance this scene has a rather natural appearance.

To illustrate here the appearance of objects such as ground vehicles, we refer to the inverse SAR, or ISAR, images shown in Figure 16. Three tactical targets are shown at a radar depression angle (or slant-plane

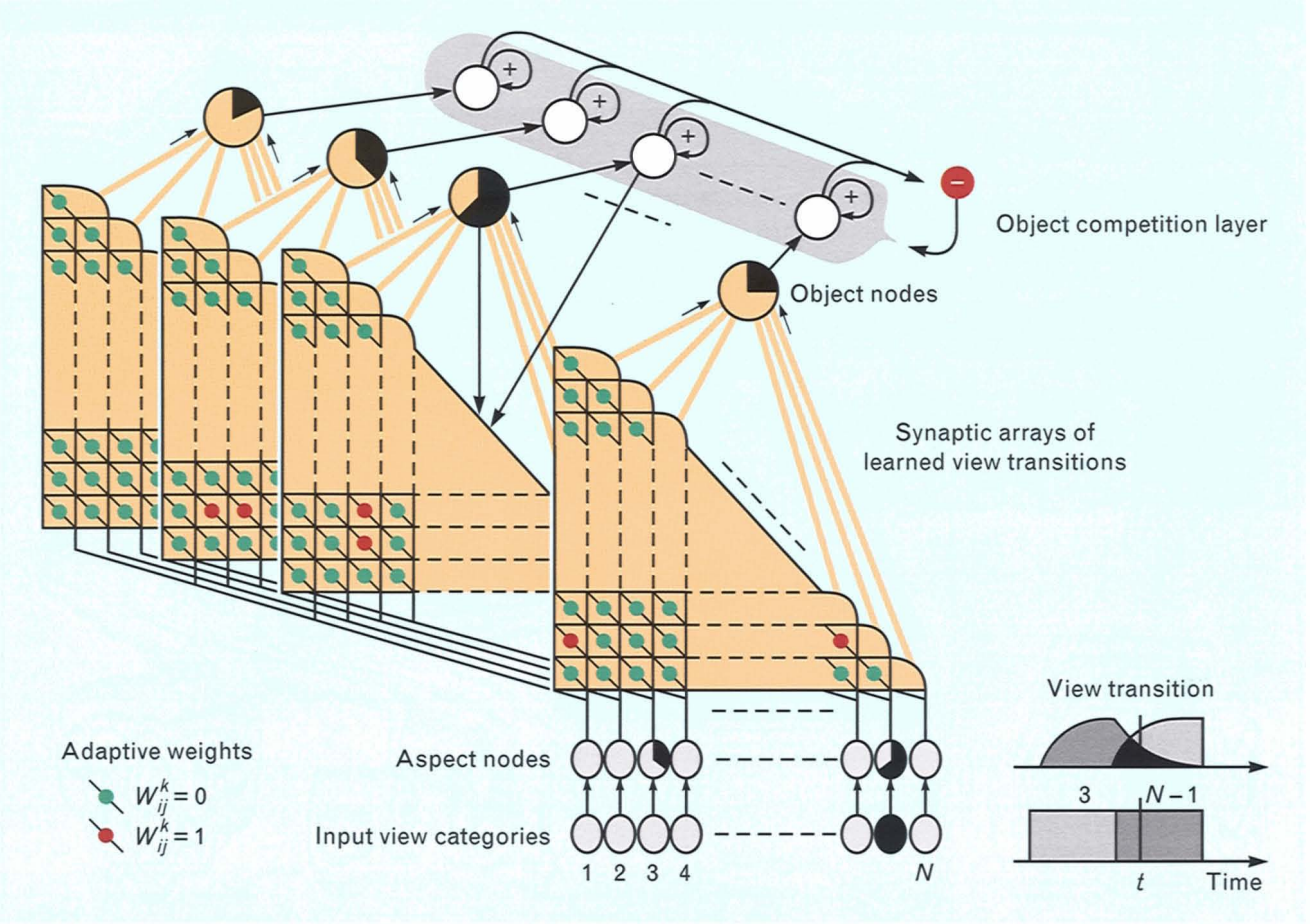


FIGURE 13. Aspect network for the multi-object case. Input aspect categories from a single ART-2 network (coding all aspects of all targets) excite aspect nodes that fan out to all synaptic arrays of learned view transitions, each of which conducts activity (i.e., evidence) to its corresponding object node. A competition layer (created from self-excitation and collective inhibition) determines the target of maximum evidence at any moment, and allows the corresponding synaptic array to be refined. Sudden eye/camera motions can cause the object nodes to reset their evidence to zero. (For a detailed description of Aspect networks, see References 10 and 11.)

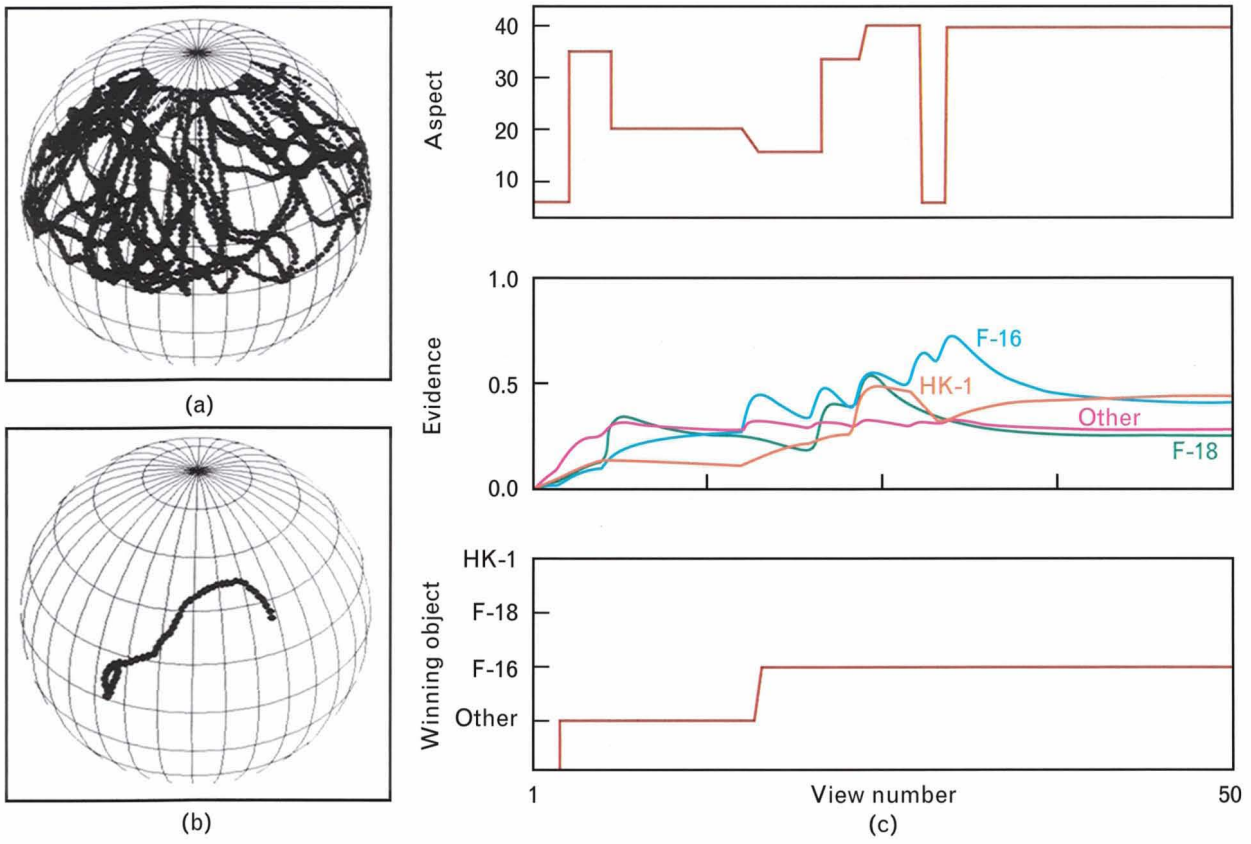


FIGURE 14. Example of training and recognition by evidence accumulation: (a) view sphere showing the trajectory from which 2000 views of each aircraft were used for training the system, (b) view sphere showing the trajectory from which 50 views of an F-16 were selected for testing the system, and (c) graphs showing the recognition test results. In part *c*, the first graph plots the sequence of aspects that were recognized by the system (note the transitions). The second graph shows the activity (i.e., evidence) of the aspect node for each aircraft target, including an unlearned target (referred to as "other"). And the final graph shows the "winning object," or target of maximum evidence at each moment. Note that the system first declares the target as "other," but then generates sufficient evidence to declare it correctly as an F-16, and that correct recognition response is maintained.

slope) of 15° and three azimuthal angles corresponding to front-on, intermediate, and broadside views. The images were obtained by rotating each target on a turntable in front of a stationary radar. Unlike with Figure 26(a), the man-made metallic objects in Figure 16 do not yield radar images that resemble their visible counterparts. The ISAR images are dominated by strong returns from select scattering centers on the target, sidelobe responses, and speckle noise. Both Figures 16 and 26(a) possess 1-ft resolution in range (oriented vertically) and cross-range (oriented horizontally), with the near-range (closest to the radar) located at the top of the image.

To build an ATR system that exploits spotlight-mode SAR sequences, we can utilize many of the ideas and neural modules developed for the visible imaging domain, as presented in the preceding section. The different sensing modality of radar, however, provides us with direct range and cross-range information, and hence object size, which can be exploited in the grouping process that is used to detect potential targets. On the other hand, our earlier methods of invariant processing must be altered. In particular, the log-polar transform must be discarded because the slant-plane image is not an angle-angle image (as is obtained in passive visible or infrared imaging).

Borrowing heavily from our work in the passive

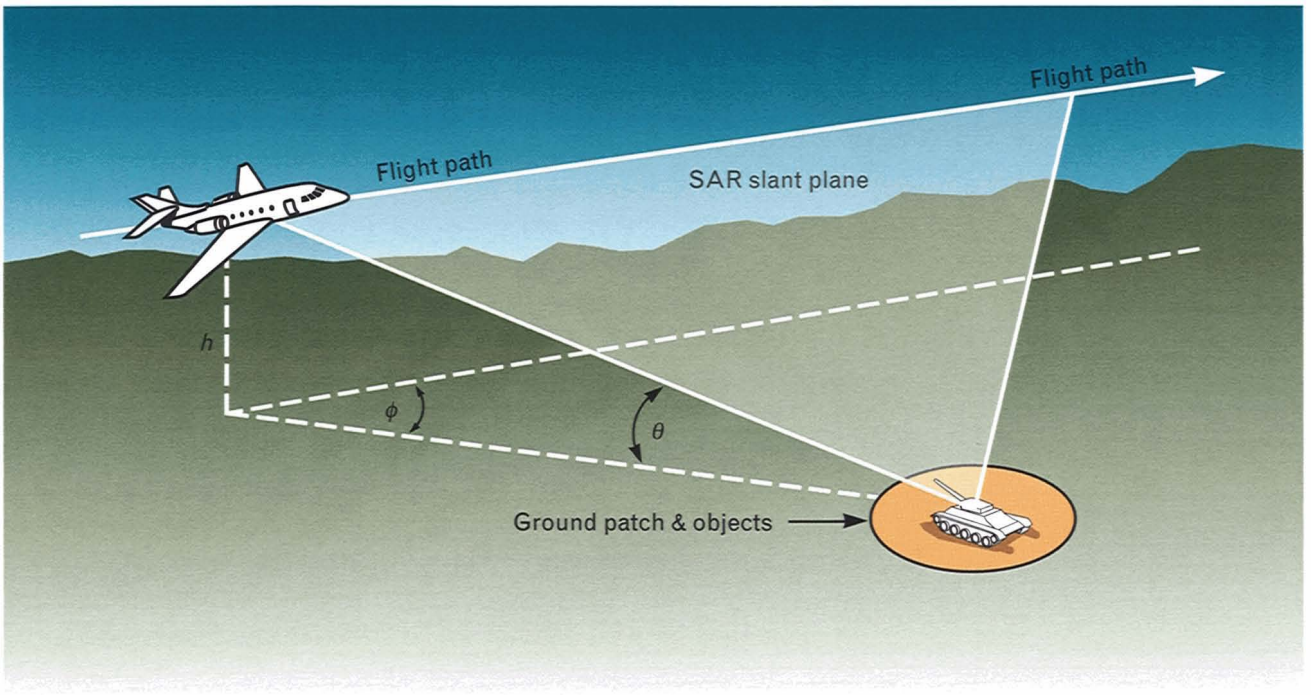


FIGURE 15. Imaging geometry for spotlight-mode synthetic-aperture radar (SAR). A radar on board an aircraft illuminates an area of interest on the ground by pointing at a depression angle θ and squint angle ϕ. As the aircraft flies along a straight path at altitude h, the radar transmits chirp pulses from many closely spaced look angles, and the return pulses determine a reflectivity image of the ground patch and objects of interest. Progressing along the flight path, the radar is steered to illuminate the same area of interest, and thus obtains a sequence of SAR images from multiple look angles.

visible domain, the conceptual approach to SAR target learning and recognition is summarized in Figure 17 (compare to Figure 3). Again, each image of the spotlight sequence is processed through three stages. The first stage extracts features, detects potential targets by grouping the features, and estimates the orientation of each potential target. The second stage is again responsible for the adaptive categorization of feature patterns into aspects, or categories (leading to an aspect-sphere representation of the targets). And the third stage detects aspect transitions (analogous to the arcs of a corresponding aspect graph), accumulates evidence over time, and generates a recognition decision as well as a dynamic confidence measure.

Figure 18 shows the end-to-end neural system that we have developed. A quick inspection of the modules, which are organized into three rows representing the three stages of processing of Figure 17, reveals many of the same neural networks that were used for the passive visible domain. It is also evident that we have learned from our work in that area: we now utilize a separate ART-2 network for each target, and a separate Aspect network connected to each of these ART-2 networks. The following figures illustrate the various processing modules shown in the system diagram.

Each SAR image of the spotlight sequence is processed by the entire chain of neural modules. There are, however, several opportunities to exploit the temporal flow of information inherent in the processed data. The very first module uses shunting center-surround networks, either in isolation or as part of the Boundary Contour System (BCS) and Feature Contour System (FCS) networks for image conditioning. (BCS/FCS networks are discussed in the following section, "SAR Image Conditioning Using BCS/FCS Networks.") Figure 19 provides some details on shunting center-surround networks, as applied to SAR imagery for feature extraction. In this application, the shunting center-surround network converts the slant-plane reflectance image into a locally normalized con-

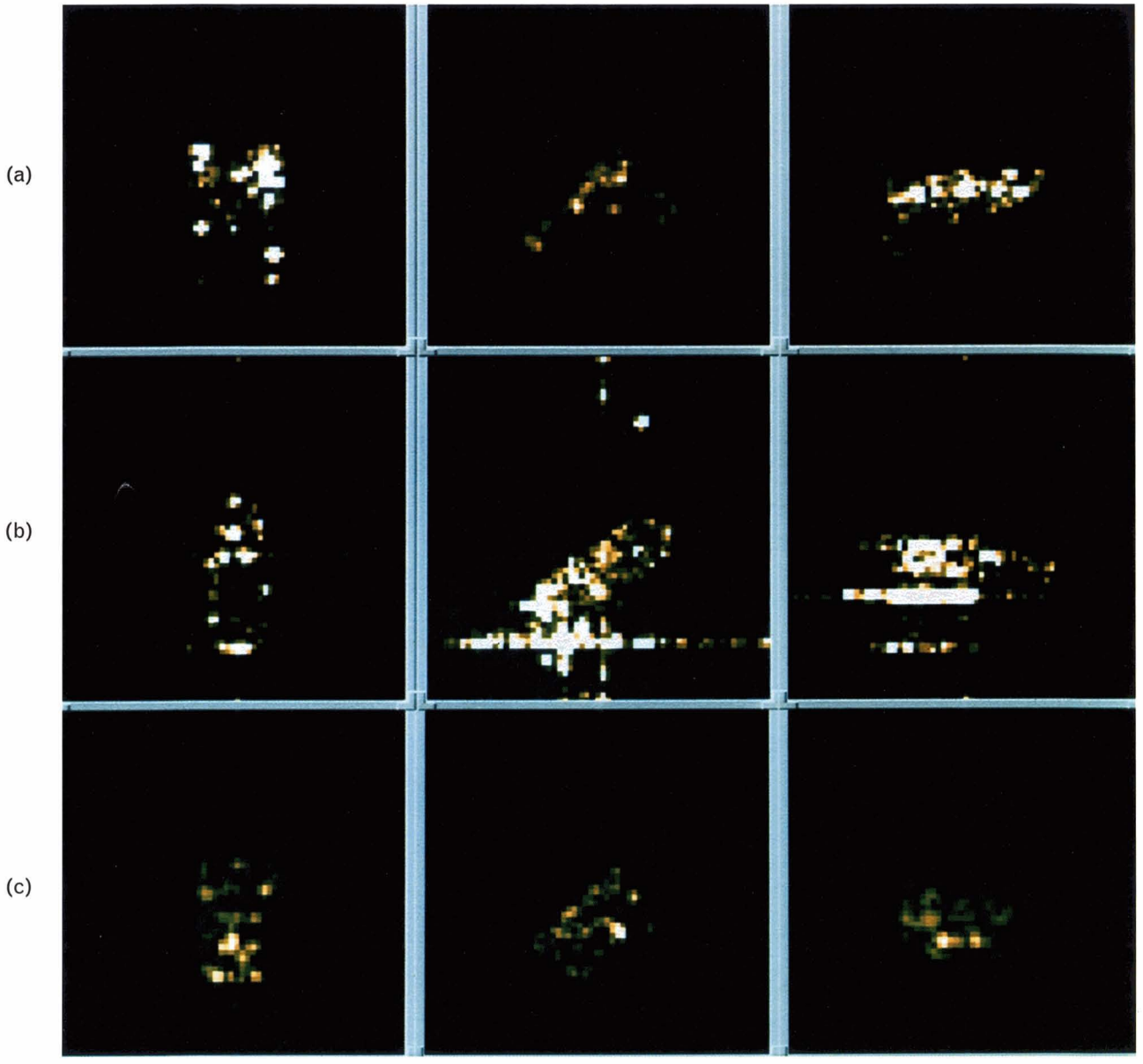


FIGURE 16. Examples of inverse SAR, or ISAR, imagery of three tactical ground vehicles: (a) target 1, (b) target 2, and (c) target 3. The three targets are shown at three different orientations: the left, middle, and right columns of images are for azimuth angles of 0° (front-on view), 45° (intermediate view), and 90° (broadside view), respectively. The images are for a radar depression angle of 15° and vertical-vertical (VV) polarization.

trast image. Thresholds are then applied to the value at each pixel of the locally normalized contrast image, and an AND operation is used to combine the resulting image with a low-threshold version of the log reflectance input image to obtain a set of high-contrast feature blobs that can then be projected from the slant plane to the ground plane by using the known radar imaging geometry.

In the locally normalized contrast image, the local contrast is dependent on the choice of spatial scales for the excitatory-center and inhibitory-surround areas of the receptive field, as shown in Figure 19. These scales are chosen so as to capture the texture of scattering centers on a vehicle as compared to the vehicle as a whole. (Note: The network does *not* try to detect bright pixels on the target as compared with the sur-

rounding clutter, as is typical of constant false-alarm rate [CFAR] filtering methods.) Another advantage of using shunting networks here is that they perform an automatic gain-control operation, and, as a result, the large dynamic range of radar reflectances collapses into a predefined range in a locally adaptive fashion. These networks are modeled as dynamic membranes [13] and resemble bipolar and ganglion receptive fields in the retina.

Figure 20 illustrates the four steps involved in processing an ISAR image that contains four targets. The input image is shown in the upper left quadrant, and the result of feature-blob extraction is shown in the upper right. The spatial patterns of the extracted feature blobs show strong resemblance to the scattering patterns obtained from SARTOOL simulations of radar imagery. (SARTOOL decomposes a target object into its principal scatterers and then combines the radar signatures of those scatterers.) Note that we have discarded the original reflectance values of the feature blobs because, in practice, they can vary considerably from one instance of a target to another. In the more realistic case of targets in clutter, feature blobs generated by nontargets will also be extracted from the clutter. Thus, to simulate clutter, we added 2% random noise to the feature-blob image before proceeding with the processing. (In Figure 20 the feature-blob image is shown without the superimposition of any noise so that we could illustrate clearly the target feature blobs that emerge from the extraction process.)

Because the image axes are measured in units of physical size, we can use the images directly to detect potential targets and discriminate them from clutter and nontarget objects by grouping the feature blobs into clusters of approximately the same image size as the targets of interest. This grouping is performed in the ground-plane coordinates, first by using an iso-

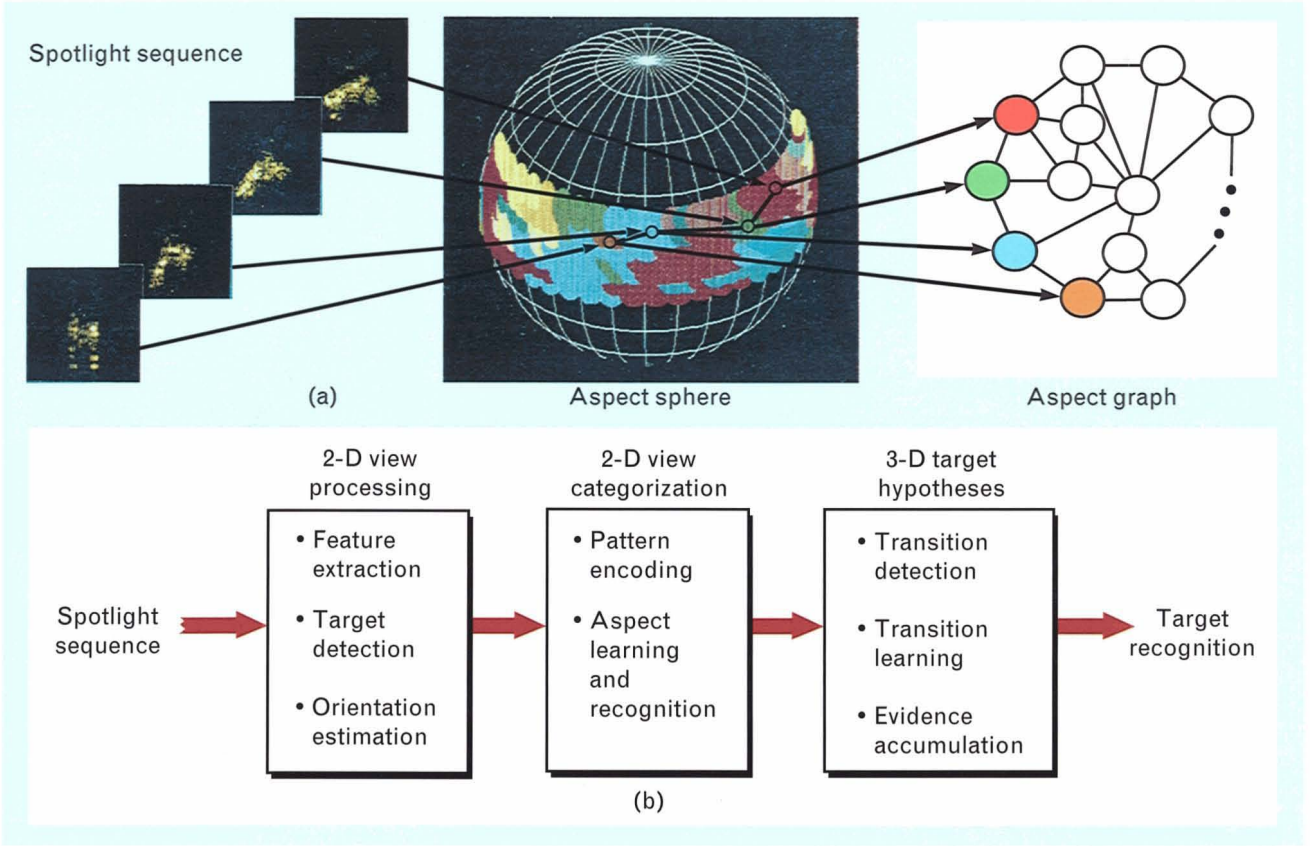


FIGURE 17. Conceptual approach of ATR neural system for SAR image sequences: (a) spotlight sequence of SAR images and corresponding aspect sphere and aspect graph, and (b) functional block diagram of system. Note that the approach is analogous to the approach for passive visual imagery shown in Figure 3.

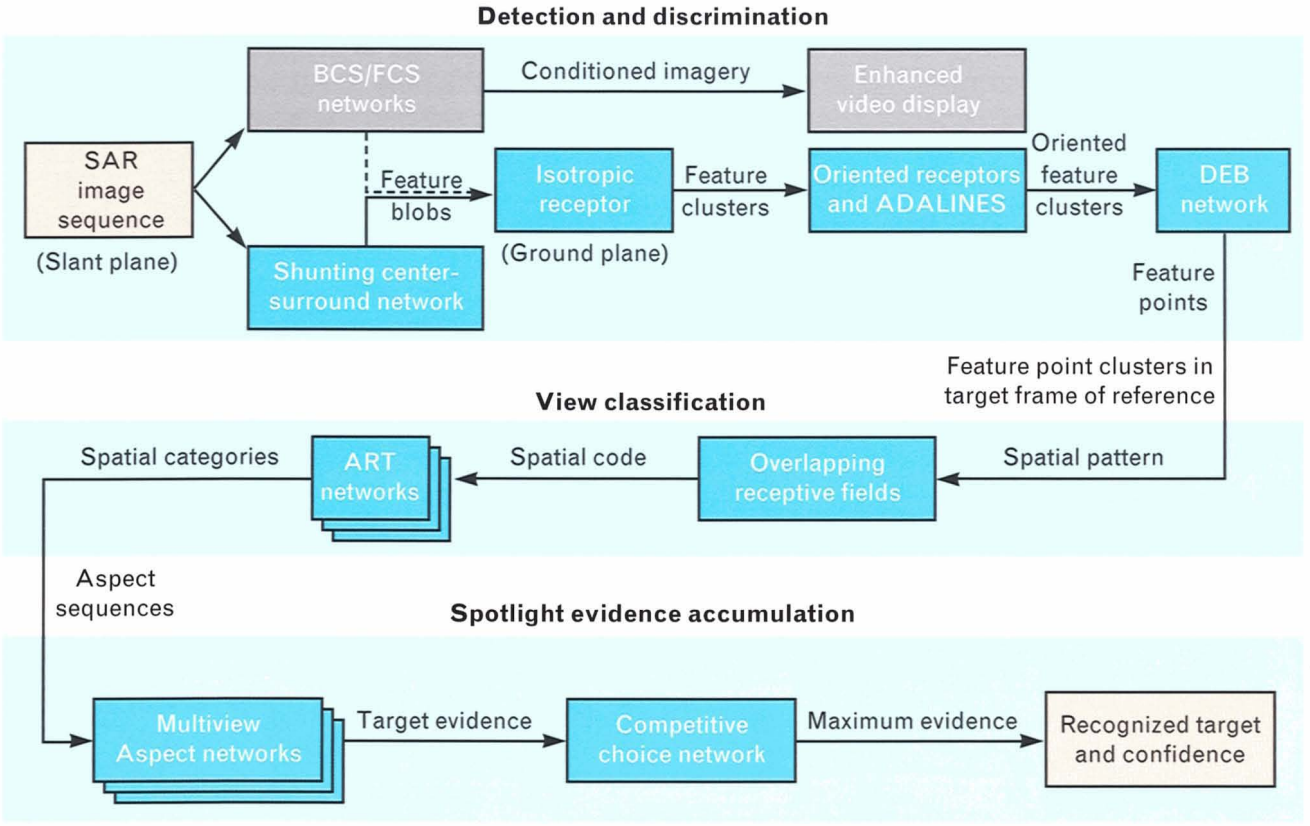


FIGURE 18. Modular system architecture for the learning and recognition of 3-D targets from SAR imagery. The three rows of modules represent the three stages of processing shown in Figure 17. Each individual module is a neural network that transforms the imagery as indicated. From a sequence of SAR images the recognized targets generate a dynamic measure of confidence.

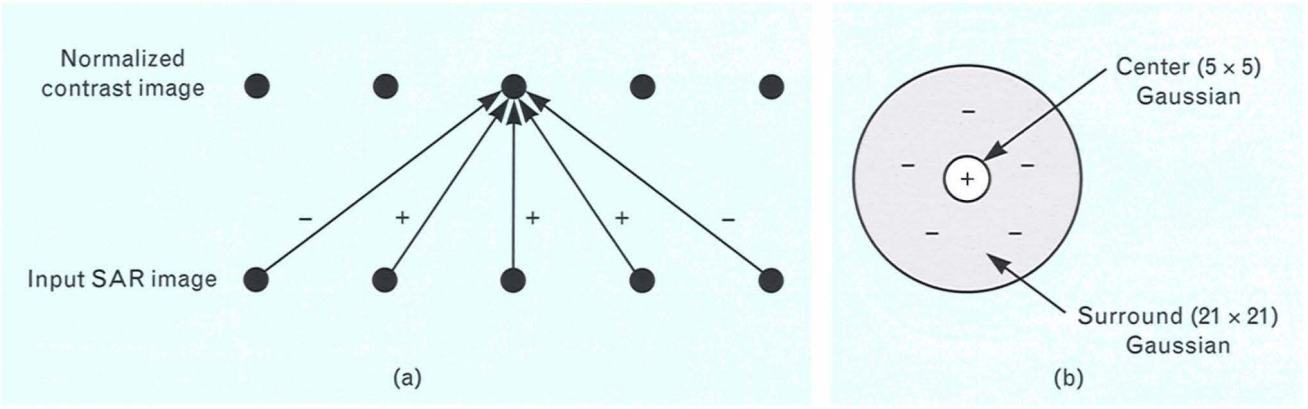


FIGURE 19. Shunting short-term memory model for feature extraction from SAR imagery: (a) center-surround feedforward architecture and (b) center-surround receptive field. The model is implemented as a feedforward dynamical system with an excitatory-center/inhibitory-surround receptive field. In equilibrium the resulting image represents locally normalized contrast. The scales of the receptive field—5 × 5 for the center region and 21 × 21 for the surround region—are chosen to capture the contrast between scatterers and target objects. (Note: A description of the equations that govern shunting short-term memory and the equilibrium condition are given in the box "Shunting Short-Term Memory" on page 98. For further details, see Reference 13.)

SHUNTING SHORT-TERM MEMORY

THE FULL DYNAMIC range radar image I_j serves as input to a shunting short-term memory (STM) network (Figure 19), as governed by the dynamics of a charging membrane (essentially Ohm's law). Excitatory input from a Gaussian center C, shunted inhibitory input from a Gaussian surround S, and passive decay (with rate constant λ_A) yield an equilibrium contrast measure A_i that is normalized with respect to the local mean amplitude. The Gaussian center is weighted by $G_{ij}^{\sigma_c}$, and the Gaussian surround is weighted by $G_{ik}^{\sigma_s}$. More general shunting networks are described in Reference 1.

Reference

1. S. Grossberg, "Nonlinear Neural Networks: Principles, Mechanisms, and Architectures," *Neural Networks* **1**, 17 (1988).

Activity dynamics:

$$\frac{dA_i}{dt} = -\lambda_A A_i + \sum_j G_{ij}^{\sigma_c} I_j - (1 + A_i) \sum_k G_{ik}^{\sigma_s} I_k$$

Equilibrium contrast:

$$A_i = \frac{\sum_j G_{ij}^{\sigma_c} I_j - \sum_k G_{ik}^{\sigma_s} I_k}{\lambda_A + \sum_k G_{ik}^{\sigma_s} I_k} \equiv \frac{C - S}{\lambda_A + S}$$

tropic receptive field (shown as circular areas in the lower left quadrant of Figure 20), and then by using oriented rectangles in the vicinity of the isotropic groupings. The rectangles are constructed from inhibitory-center/excitatory-surround receptive fields, motivated by the scatterer distributions that are typical of the targets of interest. Given a view sequence in which targets may be considered stationary as compared to the moving and squinting radar, Adaptive Linear Neurons (ADALINES) [18] performing a recursive least-squares estimation from the measurements can be used to estimate and refine the target locations and orientations. After a target has been detected and localized, it can then be segmented from the scene, rotated into a reference frame aligned with the target (with an ambiguity between whether the target is facing forward or backward), and processed by the remaining modules. The oriented feature blobs for each detected target are shown in the lower right quadrant of Figure 20. Note that sidelobe responses outside the targets have been discarded.

Figure 21 illustrates the processing of an individual target. The input slant-plane imagery is shown in the upper left quadrant of the figure, and the localized target feature blobs are shown in the upper right. After the features are reoriented in a frame of reference with respect to the target, a DEB network is used to reduce the features to points, as shown in the lower left quadrant. This oriented spatial pattern of feature points covers an extent of approximately 20 × 30 pixels at 1-ft resolution. Moreover, as the target orientation and radar depression angle change, this pattern must change quickly, too. Of equal importance is the deformation of this feature pattern with varying radar squint angle (given an identical depression angle and target orientation with respect to the radar). For these reasons, a template constructed from this feature pattern (no less a template that incorporates the original reflectance values) is both memory intensive and fraught with difficulties. Thus we can again utilize the large overlapping receptive fields (cf. Figure 7) to reduce this binary feature pattern to a 9 × 9 array of analog numbers that code the spatial distribution of features in a compressed manner that is robust to

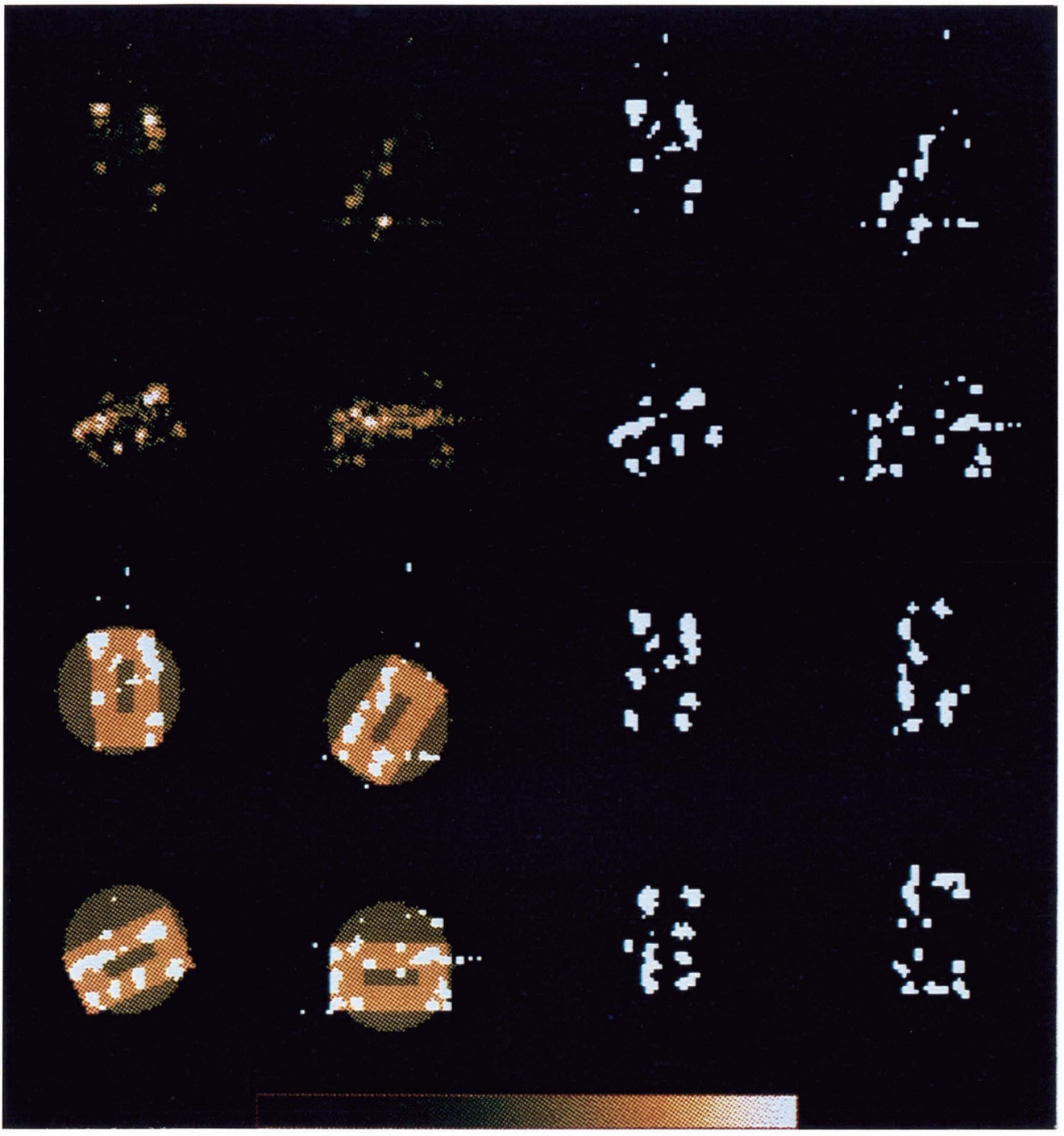


FIGURE 20. Multitarget localization and orientation estimation is illustrated for the case of four tactical targets. The upper left quadrant shows the input slant-plane imagery for the composite of four targets at a depression angle of 15°. (Note: In each of the four quadrants, the four targets are arranged with target 1 in the upper left, target 2 in the upper right, target 3 in the lower left, and a modified version of target 1 in the lower right.) The upper right quadrant contains the feature blobs that have been extracted and projected in the ground plane. Each target is then grouped with a circular mask, and the target's orientation is estimated with an inhibitory-center/excitatory-surround oriented rectangular mask, as shown in the lower left quadrant. This processing allows the detected targets to be reoriented in a target frame of reference, as has been done in the lower right quadrant. The color bar at the bottom of the figure denotes increasing (from black to white) reflectivity for the imagery in the upper left quadrant.

spatial deformation. Such a coding is illustrated in the lower right quadrant of Figure 21.

The 9 × 9 array of 81 numbers forms the input to an ART-2 network that is dedicated to the learning of a particular target. The target is learned during a training session in which the target exposure is controlled. In a testing session, the system is run in recognition mode, and the 81-member spatial code vector is fed to all ART-2 networks representing all targets of interest. Figure 22 illustrates the results of training independent ART-2 networks on each of the three ISAR targets of Figure 16. The resulting aspect categories that were established are shown color coded on aspect spheres seen both from the side with the targets facing left, and from above (compare to Figure 11). The data consisted of ISAR images created at all even azimuths 360° around each target, for radar depression angles between 15° and 32°, comprising approximately 3000 views of each target. (Note the missing data at a few intermediate depression angles for targets 2 and 3.) The resulting unsupervised classification generated islands of common category that extended over large azimuthal extents and many depression angles. We purposely changed the vigilance parameter setting between targets to emphasize the user's control over the fineness of categorization. With a finer categorization, more details survive the learning process. The roughly 3000 views of each target have been compressed into only 34 categories for targets 1 and 2, and 75 categories for target 3, for which we used the highest vigilance setting.

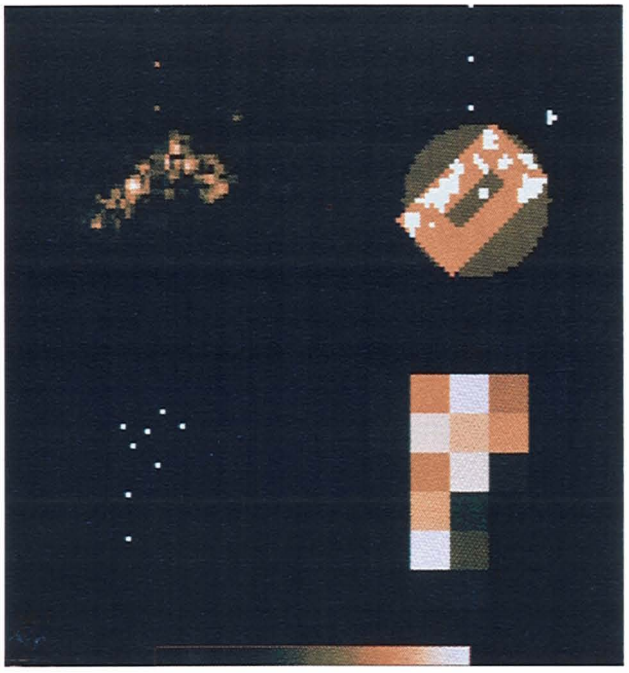


FIGURE 21. Target feature extraction and spatial coding for a single target. The upper left quadrant shows the input slant-plane image of a target at a depression angle of 15°. In the upper right quadrant, feature blobs have been extracted from the input image and projected in the ground plane. After the feature blobs are reoriented in the target frame of reference, a DEB network is used to reduce the blobs to points, as shown in the lower left quadrant. The feature points are then coarsely coded by a 9 × 9 array of overlapping receptive fields (cf. Figure 7), as shown in the lower right quadrant. The color bar at the bottom of the figure denotes increasing (from black to white) reflectivity for the upper left image, and increasing receptive-field activity for the lower right image.

Associated with each category allocated by each ART-2 network is a template of the prototype 9 × 9 array (contrast enhanced and normalized) that was learned by the synapses (the adaptive LTM sites) in the network, as shown in Figure 8. Eight of the principal templates for target 1 are illustrated in Figure 23, along with sample slant-plane images from the corresponding viewing directions. Note that the learned templates include two broadside views, frontal and end-on views, and the characteristic L-shapes near the four corner views. These color patterns code prototype spatial feature patterns (*not* reflectance patterns).

To complete the learned description of each target, the permitted transitions among aspect categories must be detected and imposed on the synapses of each Aspect network. The result of this process is contained in the last row of photographs in Figure 22. The photographs show the transition matrices for each target (cf. Figure 13). In the matrices, a red pixel corresponds to a permitted transition while a green pixel codes the absence of such a transition.

Figure 24 contains an example of our ATR system running in recognition mode. We used an ISAR image sequence that consisted of 45 views of target 1 (only the odd azimuths in the interval 67° to 157°) at a depression angle of 21°. (Note: Although this dataset was not part of the original training set, it was admittedly not very different from the training data. This lack of adequate training and test datasets is a prob-

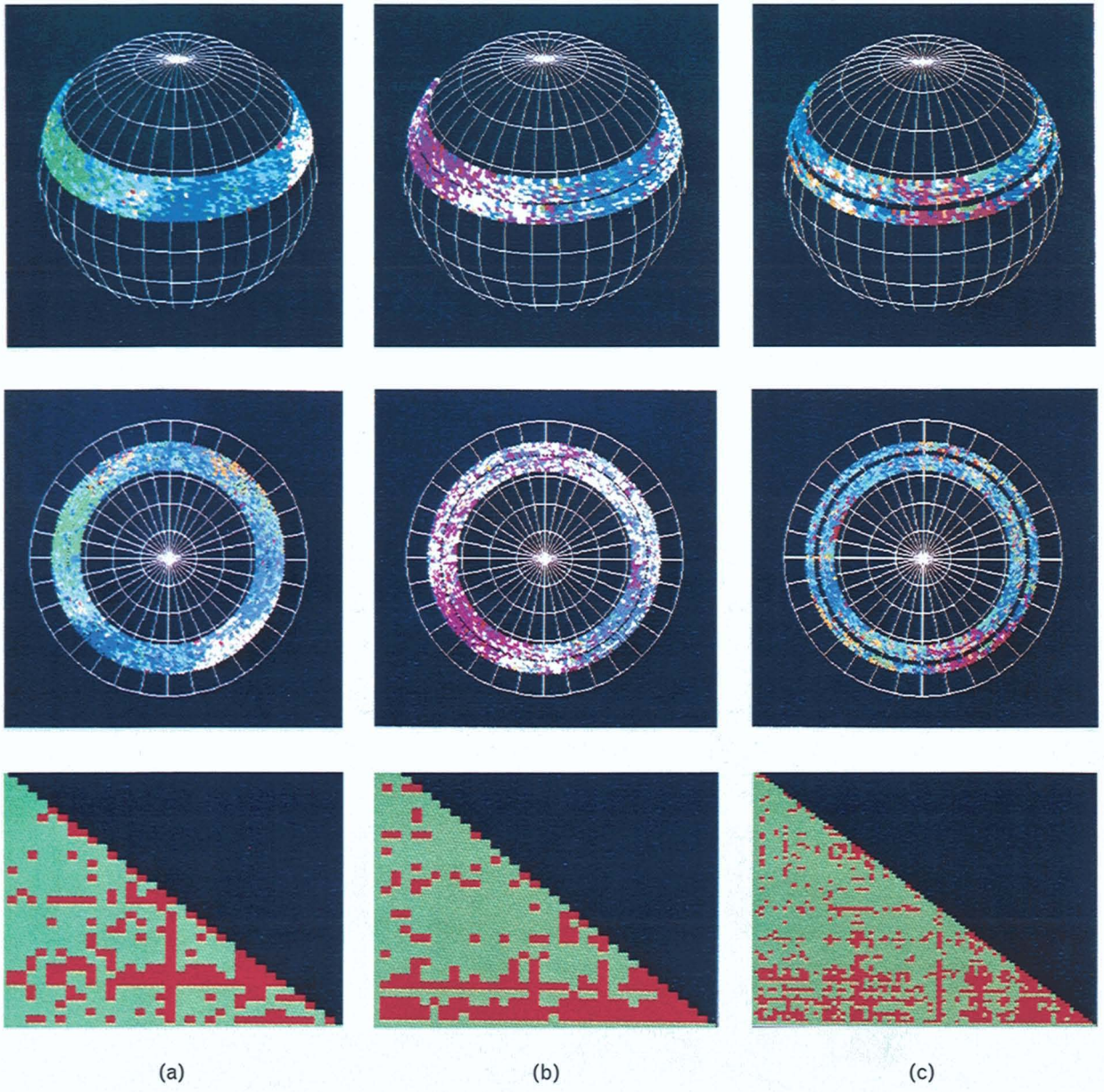


FIGURE 22. Aspect spheres and transition matrices for (a) target 1, (b) target 2, and (c) target 3. The spheres were generated with independent ART-2 networks. For each target, approximately 3000 views, collected for even azimuths 360° around and depression angles from 15° to 32°, have been compressed into 34, 34, and 75 categories for targets 1, 2, and 3, respectively. The categories, or aspects, have been assigned colors and are shown on a viewing sphere from the side with the target facing left (top row of photographs), and from above (center row of photographs). Note the category islands that emerge over large viewing extents, particularly for target 1. The fineness of categorization is controlled by the vigilance parameter of the ART-2 network, and can be chosen to be more or less sensitive to variations in the feature patterns. The vigilance ρ is 0.97, 0.98, and 0.99 for targets 1, 2, and 3, respectively. The last row of photographs shows the category transitions that were learned for each target by independent Aspect networks coupled to each ART-2 network. Each transition matrix codes possible category transitions in red, while green denotes the absence of such a transition (cf. Figure 13). Note that the transition patterns are quite different among the targets. Detected transitions between categories contribute to the evidence accumulation during the recognition process.

lem with many ATR studies.) The test imagery was passed through the early modules of our system, and then to four ART-2 networks (with learning turned off) coupled to four Aspect networks corresponding to the training targets 1, 2, and 3, in addition to a fourth unlearned target (referred to as "other") that was represented by random synaptic weights. Each ART-2 network determined the best matching aspect category for the test target, and each ART-2 network activated its corresponding Aspect network to accumulate evidence over the test view sequence. A competition between the object nodes in the different Aspect networks then selected the target with the instantaneous maximum evidence.

In Figure 24, the category sequence recognized by the ART-2 network for target 1 is illustrated both on an aspect sphere seen from above, and as a graph of category versus view number. The second graph in Figure 24(b) plots the accumulation of evidence for each target, while the third graph indicates the selected target with the maximum evidence accumulated at each view. Although the selected target is target 1, we can see that target 3 also accumulates a significant amount of evidence. Thus selecting target 1 solely on the basis of maximum evidence can be risky because the evidence for target 1 may exceed that for target 3 by only a slight amount. This possibility suggests looking at the *differential evidence* between the two targets of highest accumulated evidence, as illustrated in the fourth graph. The differential evidence may be small for some views, but it too can be integrated along the temporal view sequence, giving rise to a *dynamic confidence measure.* As shown in the bottom graph of Figure 24, the confidence measure increases monotonically along the view sequence in this example. It is a matter of preference to select the threshold level of confidence that the system should use in declaring a target as recognized. Clearly, the number of views required to reach this confidence threshold will depend on the target itself, as well as the starting view in a sequence.

SAR Image Conditioning Using BCS/FCS Networks

We have already noted that single-channel SAR imagery is characterized by a very large dynamic range and

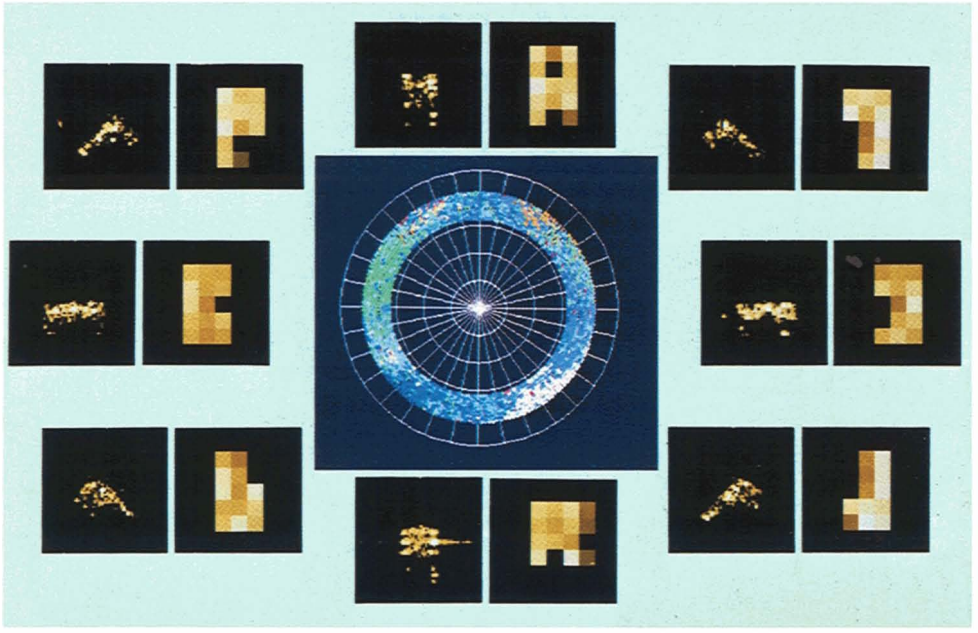


FIGURE 23. Aspect sphere, example typical views, and corresponding learned templates for target 1 of Figure 22(a). The learned templates include two broadside views, frontal and end-on views, and the characteristic L-shapes near the four corner views. Note the ability of the ART-2 network to quantize the viewing space around a target in an unsupervised fashion. The learned templates are then used for the recognition process.

excessive speckle noise, and man-made objects possess rather broken signatures that vary rapidly with small changes in viewing angle. To a great extent, we can alleviate these problems by first conditioning the imagery with the *Boundary Contour System* and *Feature Contour System* (BCS/FCS) network paradigm developed by S. Grossberg, E. Mingolla, and D. Todorović, (see chapters 1 to 4 in Reference 19). This neural processing architecture is strongly motivated by the known anatomy and physiology of the early visual processing stages, including that of the retina, LGN, V1, V2, and V4. The architecture, which essentially incorporates a general theory of preattentive vision, has also been quite successful in explaining a very large body of psychophysical perceptual data.

The BCS/FCS networks are shown as an alternative first module in the ATR system of Figure 18. Our preliminary work indicates that the initial processing of SAR imagery with BCS/FCS networks, in lieu of a shunting center-surround network, improves the target detection (and false-alarm rejection) process.

Preattentive vision, in its simplest form, is a computational process in which contours are contextually established and the perceived brightness (and color) is generated primarily from local-contrast information. In BCS/FCS theory, the role of the BCS network is to establish such contours in the context of local fields of

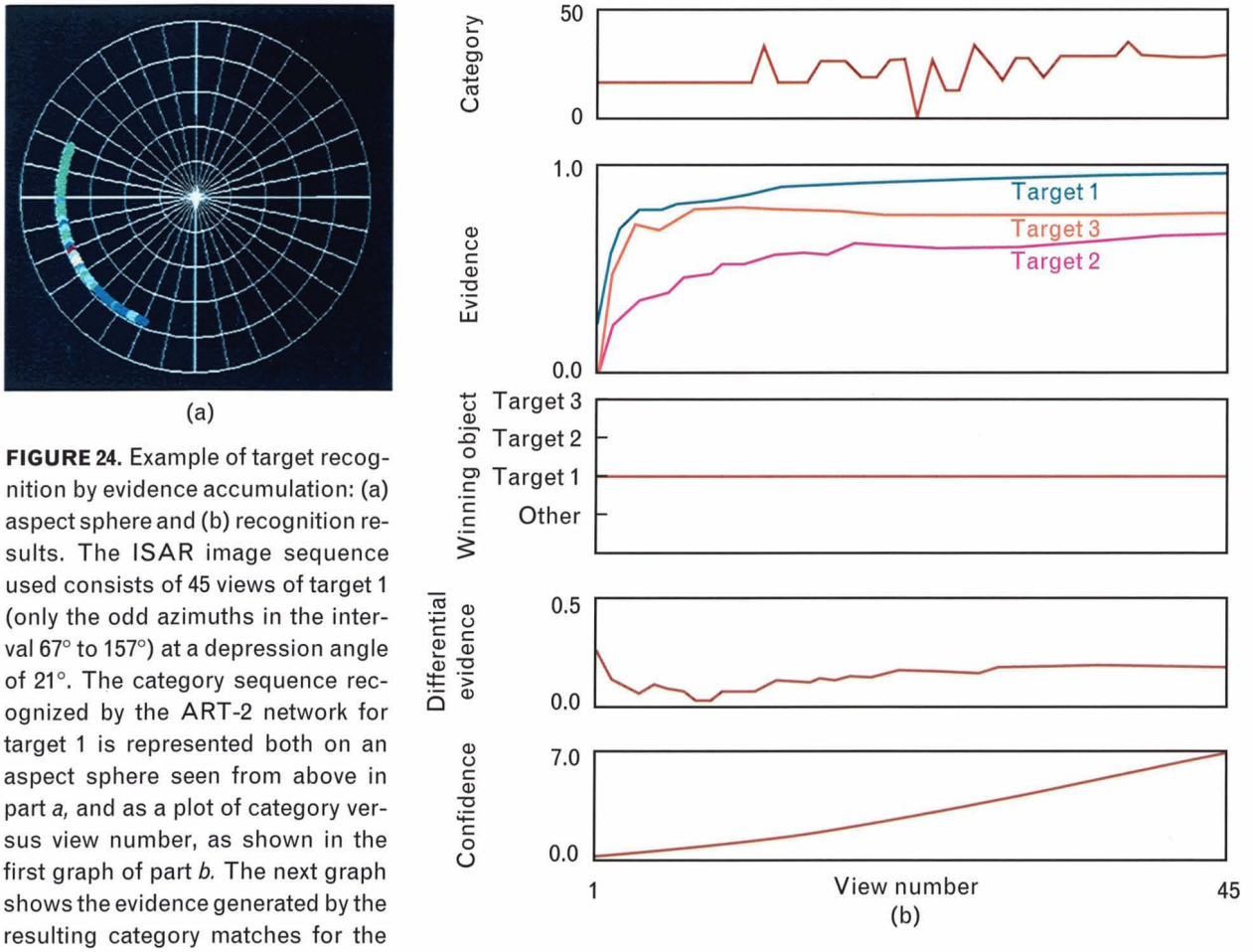
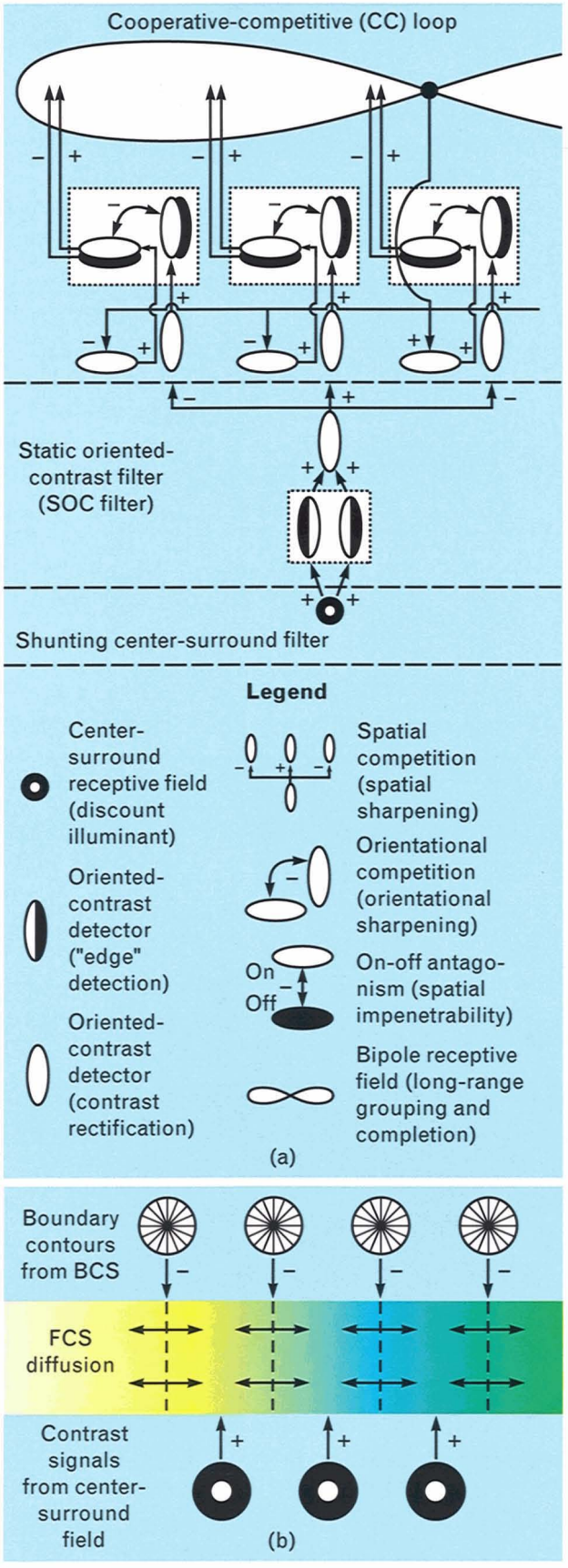


FIGURE 24. Example of target recognition by evidence accumulation: (a) aspect sphere and (b) recognition results. The ISAR image sequence used consists of 45 views of target 1 (only the odd azimuths in the interval 67° to 157°) at a depression angle of 21°. The category sequence recognized by the ART-2 network for target 1 is represented both on an aspect sphere seen from above in part *a*, and as a plot of category versus view number, as shown in the first graph of part *b*. The next graph shows the evidence generated by the resulting category matches for the training targets 1, 2, and 3. The third graph indicates the "winning object," i.e., the selected target with the maximum evidence accumulated at each view. The target with the instantaneous maximum evidence is consistently target 1, although target 3 also has a strong response. The differential evidence between those two targets is plotted in the fourth graph, and integrated across view numbers in the fifth graph to generate a monotonically increasing confidence measure.



edge fragments, or oriented contrast. Although such *boundary contours* are themselves invisible, they modulate the dynamics of a diffusive filling-in process in the FCS network whereby local contrast and brightness information mix and spread within such boundaries to create a smoothly shaded figure.

The architecture of the BCS network is illustrated in Figure 25(a). Beginning with monocular preprocessing in the form of shunting center-surround receptive fields, local measures of normalized isotropic contrast are made. An oriented-contrast filter then derives evidence for local *edge fragments,* which are then used as input to a *cooperative-competitive (CC)* feedback loop. The CC loop performs long-range completion of contours in the context of the local edge statistics. We have found that one pass through the CC loop is typically sufficient for our purposes. The boundary contours obtained from the BCS network provide input to the FCS filling-in network, along with the center-surround contrast signals, as shown in Figure 25(b). Essentially, the contrast signals try to spread diffusively to neighboring nodes, but the BCS signals modulate the local diffusivity

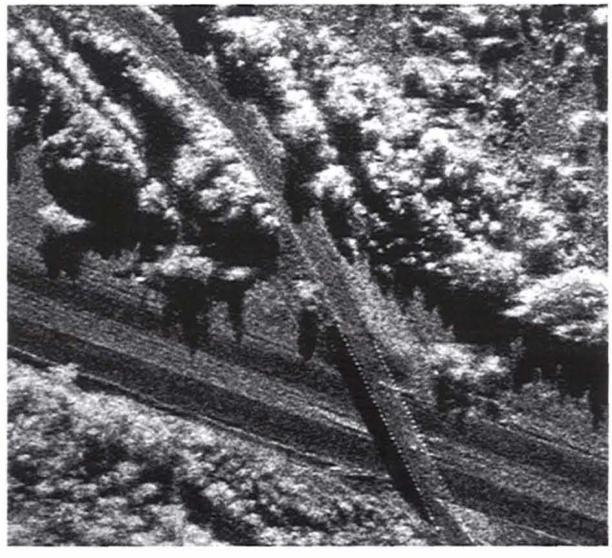
FIGURE 25. Architecture of the (a) Boundary Contour System (BCS) of S. Grossberg and E. Mingolla and the (b) Feature Contour System (FCS) of Grossberg and D. Todorović. The BCS architecture in part *a* models the neurodynamics of preattentive visual processing in the LGN, V1, and V2 visual areas of the brain. Shunting center-surround receptive fields provide input to oriented-contrast, or edge, detectors that compete across position and orientation. The resulting local edge fragments are grouped over large distances by oriented bipole receptive fields that feed back to the oriented-contrast detectors to complete broken boundaries comprising the edge fragments. The boundary contours obtained from the BCS network provide input to the FCS network, which uses the information to modulate the local diffusivity between compartments, as shown in part *b.* The local diffusivity between compartments affects the FCS diffusion layer, where the contrast signals from the shunting center-surround network spread laterally in two dimensions. In the diffusion layer, strong boundary contours inhibit diffusion across the boundaries. The FCS architecture models hypothesized filling-in interactions in the V4 visual area of the brain. (Adapted from Figures 15 and 17 of chapter 1 in Reference 19, with permission. This reference also contains a detailed description of BCS/FCS networks.)

such that strong boundary contours inhibit diffusion across the boundaries. Thus the boundary contours impede the spreading contrast signals. In the presence of a dense web of boundary contours of varying strength, the FCS diffusion process results in smoothly shaded images while retaining sharp transitions in brightness.

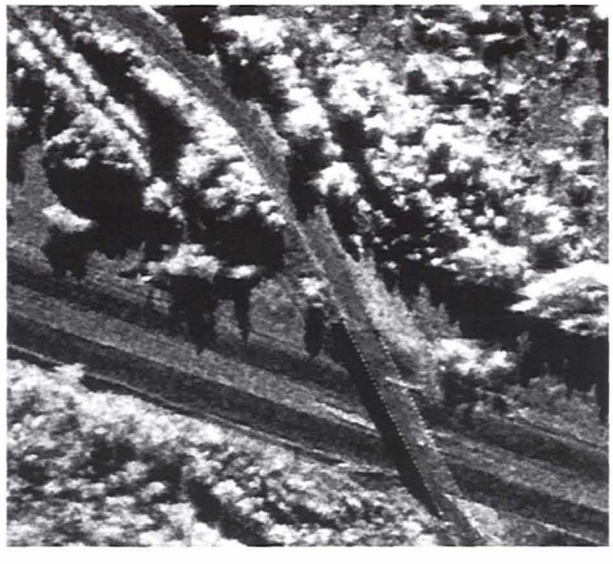
In applying BCS/FCS processing to SAR imagery, various parameters in the governing dynamical system need to be selected so that the pixel values are not discounted completely (because the original SAR image brightnesses are actually reflectance measures). We can then preserve the ordering of the resulting brightnesses in fairly uniform areas so as to mimic the ordering of the initial reflectance values. In nonuniform areas, however, the resulting signals indicate a mixture of reflectance and local contrast. The overall effect is SAR imagery with significantly less speckle noise, darkened and sharpened shadows, and more smoothly shaded signatures. Figure 26(a), obtained with the Lincoln Laboratory ADTS SAR, illustrates a clutter scene of trees, roads, and an overpass that crosses the New York State Thruway. The image was obtained with single-channel VV polarization at 1-ft resolution. Because of the large dynamic range, the scene is displayed as a log-amplitude image. Figure 26(b) shows the same scene after BCS/FCS processing of the full-dynamic-range SAR image. Note the dramatic reduction in speckle, the darkening of shadows, the sharpening of shadow contours, and the smooth shading of the treetops, roads, and grass.

Figure 27 illustrates the various stages of processing for the three ISAR targets oriented at a 45° azimuth and a 15° radar depression angle. The log-amplitude ISAR imagery is shown in the first column, the contrast-enhanced output of the shunting center-surround network is contained in the second column, the boundary contours derived from the BCS network are given in the third column, and the smoothly shaded signatures obtained from the FCS network are in the fourth column. An important attribute of the FCS filled-in signatures is that they are quite stable with respect to small changes in target orientation.

For ATR applications involving SAR imagery, BCS/FCS processing is a useful image-conditioning proce-



(a)



(b)

FIGURE 26. SAR image conditioning with BCS/FCS networks: (a) original SAR image of an overpass that crosses the New York State Thruway and (b) image after BCS/FCS processing. The original single-channel VV-polarization image (shown as log amplitude of the reflectance) is corrupted by speckle noise, which results in many false alarms making target detection difficult. In the BCS/FCS-processed image, note the reduction of speckle noise, the darkening of shadow areas, and the crispness of the shadow contours. Such image conditioning improves target detection while reducing false alarms. The SAR image was obtained with the Advanced Detection Technology Sensor (ADTS), a Lincoln Laboratory millimeter-wave radar.

dure. We expect it to improve both the target detection and recognition stages of our ATR system.

Reentry-Vehicle Recognition from ISAR Sequences

We have also applied our SAR target-recognition system to the identification of reentry vehicles imaged by a ground radar while the vehicles were spinning and traveling along a trajectory. The resulting reentry-vehicle images are thus ISAR imagery, although they are in general simpler than that obtained with tactical targets in clutter. Radar processing is typically done in the range-Doppler domain to extract peaks corresponding to isolated scattering centers on the vehicle's shroud. We can then apply to these data the same three stages of processing that we applied to the ISAR tactical-target imagery: the range-Doppler peaks can be used as point feature patterns, the patterns can be encoded by overlapping receptive fields followed by classification with ART-2, and evidence and confidence can be accumulated with the Aspect network. (Note: An alternative approach to the learning and recognition of reentry vehicles has recently been reported by A.M. Aull et al. [20].)

With this approach in mind, we have constructed ISAR imagery of point scatterers for three reentry vehicles over several rotations at a single angle of attack (i.e., a single depression angle). The vehicles are designated as RV-1, RV-2, and RV-3. Figure 28 illustrates the result of coding and ART-2 category learning for vehicle RV-2 at a vigilance setting of 0.95. Aspect categorization over multiple rotations are shown on an aspect sphere, along with the learned templates and typical feature patterns for the six categories that ART-2 established. Figure 29 shows the results of separate ART-2 categorizations for all three reentry vehicles over multiple rotations, as well as the learned transitions among aspect categories used by the Aspect network. The three vehicles differ in their complexity, which is reflected by the number of categories required by ART-2 to cluster the data: 3, 6, and 16 categories for RV-1, RV-2, and RV-3, respectively.

The results of a recognition experiment are shown in Figure 30 (compare to Figure 24). In the experiment, a sequence of views of vehicle RV-3 was input to the system. (Note: This sequence was not part of the data used for training the system.) Again, evidence was accumulated for all three targets in addition to an unlearned target, the target of maximum evidence was chosen, differential evidence was computed from the two targets of highest evidence, and the difference was integrated along the view sequence to generate a confidence measure. In Figure 30 we see an example of the system changing its selection. The system first (correctly) chooses RV-3 although the confidence is still relatively low, then the system gets confused and switches between the other two vehicles. The switching resets the confidence to zero, and it remains very small due to the small differential evidence generated. Finally, the system locks back onto the correct decision, and confidence builds monotonically.

Table 1 (page 109) summarizes the results of preliminary recognition experiments on these three reentry vehicles. In each case the test sequence consisted of 90 images starting at randomly selected azimuths. In all cases the correct vehicle was recognized, and fewer than 25 images were required in each sequence to converge to a high-confidence correct decision. By converting this result to the fraction of each vehicle's rotation cycle that is required to achieve such recognition, we find that fewer than two revolutions were required in each case.

Learning and Recognition Using Salient Object Parts

The 3-D object learning and recognition system described thus far processes the views of objects as a whole. But this approach can lead to a decline in recognition ability when an object is partially occluded or disguised, or when a part of the object is articulated or variable (removed or replaced). To deal with these situations, we return to biology for guidance.

The brain processes information by using a *principle of contrast.* Many operations seem to be cast in terms of differences, or in terms of the detection of novelties or transitions in space, time, or patterns. Mechanisms exist that detect novel changes, as reflected by the peak in EEG measurements that occurs 300 msec after the introduction of an unexpected

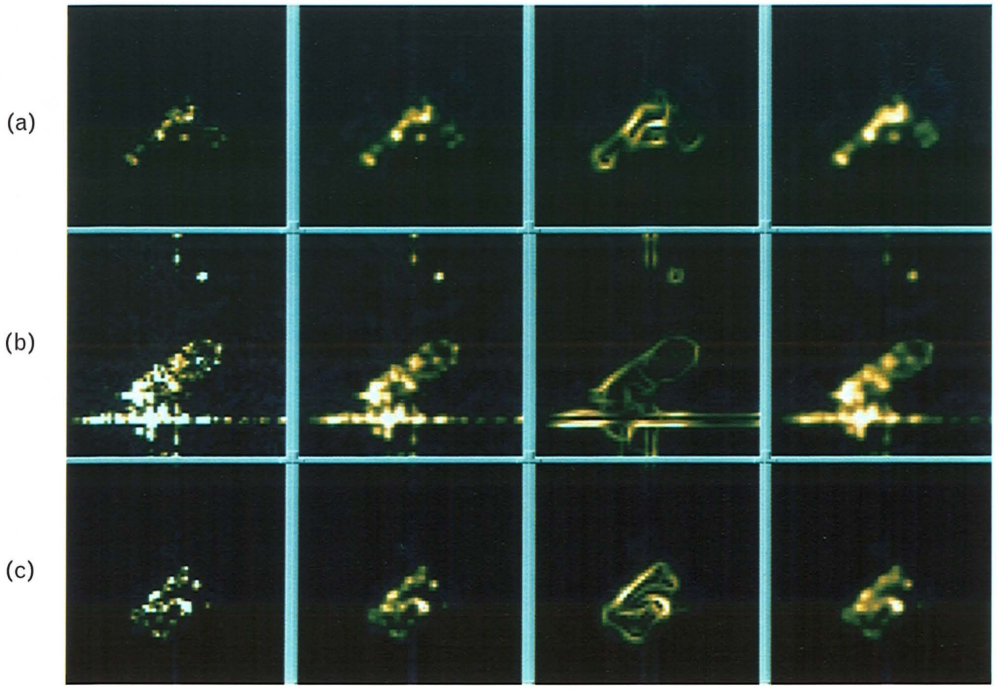


FIGURE 27. BCS/FCS processing applied to the ISAR images of (a) target 1, (b) target 2, and (c) target 3. From left to right, the columns show different stages of the BCS/FCS processing. The ISAR imagery (first column) is contrast enhanced by a shunting center-surround network (second column) and boundary contours are extracted (third column). The contrast-enhanced imagery diffuses within the boundary contours to produce filled-in target signatures (fourth column). All three targets are oriented at a 45° azimuth and a 15° radar depression angle.

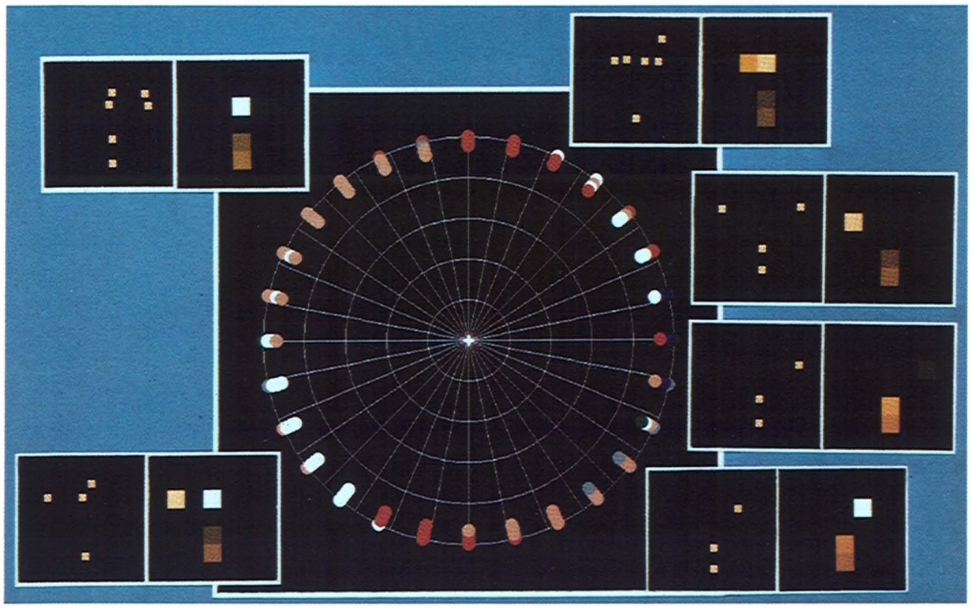
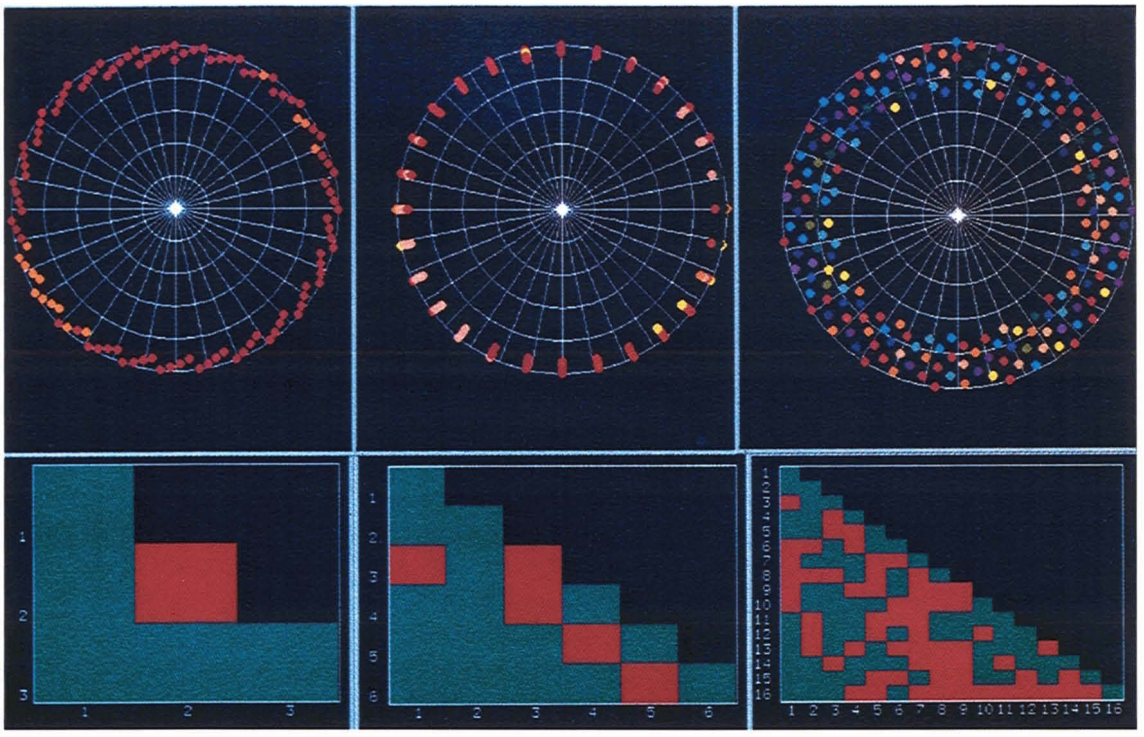


FIGURE 28. Learned categories for reentry vehicle RV-2 are plotted on an aspect sphere over four rotations of the target. (For convenience, the data for each of the four rotations have been plotted on the aspect sphere at different shifted depression angles. Note the four rounds of colored dots on the sphere.) During the learning process, ART-2 generated only six categories (at a vigilance setting of 0.95). The learned templates along with the representative scatterer patterns for the six categories are shown. From the upper right corner of the overall figure, the corresponding colors for the learned templates are dark brown, dark blue, green, light blue, white, and light brown.



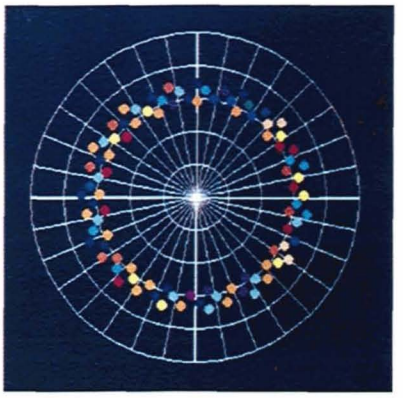
(a) (b) (c)

FIGURE 29. Learned categories (aspect spheres) and transition matrices for the (a) RV-1, (b) RV-2, and (c) RV-3 reentry vehicles over multiple rotation cycles. The vigilance setting is 0.95, 0.95, and 0.96 for the three reentry vehicles, respectively, and the resulting number of categories established is 3, 6, and 16, respectively. (The difference in the number of categories reflects differences in the complexities of the three vehicles.) In the transition matrices, the possible category transitions are coded in blue, while red denotes the absence of such a transition.

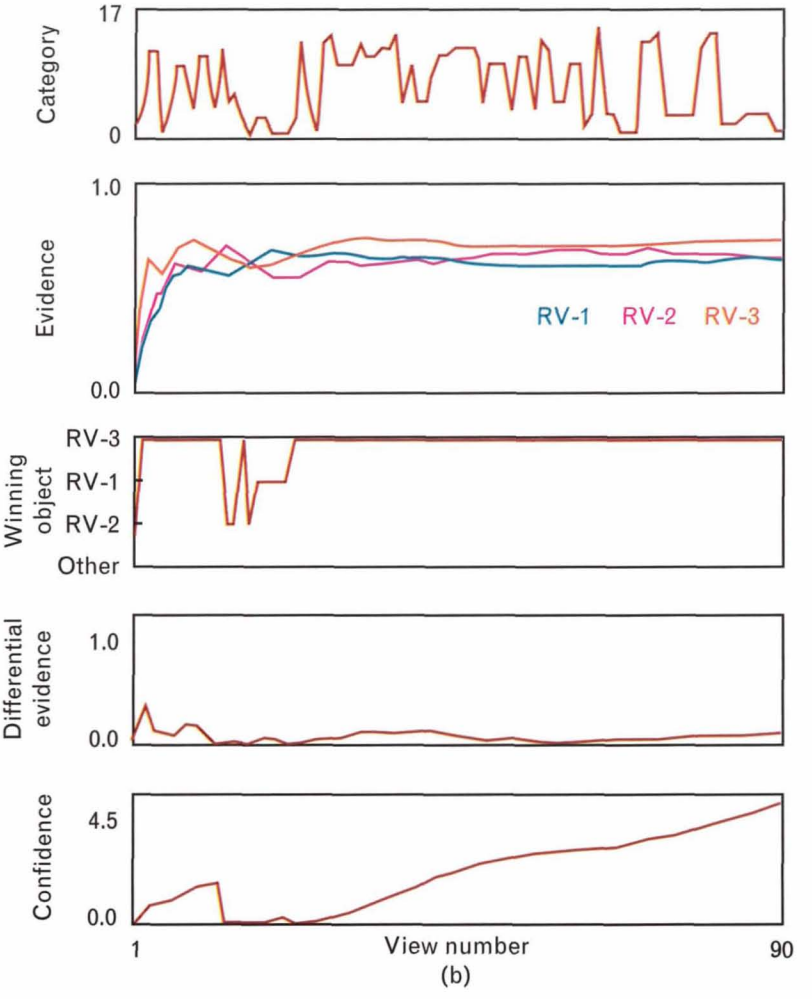
new stimulus. Other mechanisms are responsible for suppressing information that is not changing. For instance, stabilized retinal images fade away in about one second. In fact, all sensory systems become habituated to constant or repetitive input patterns. Indeed, human vigilance decreases after long periods of waiting, and we become bored. This principle of contrast has been exploited earlier in our system in the form of center-surround receptive fields, edge detectors, competitive learning, view-transition detection, and confidence estimation via differential evidence. We now use the principle again, this time as a foundation for *Saliency Maps,* hierarchical object-part representations, and caricature-based recognition [21].

Visual attention not only focuses processing power on an object in a scene, it often isolates only a part of the object for closer inspection. (Note: Evidence for such a finely tuned attentional mechanism has been found in psychological studies in which subjects are demonstrably unaware of stimuli external to the attended visual area.) A serial examination of the object takes place in which the examination is focused on the different *parts* of the stimuli, which may or may not correspond to different parts of the object. But what actually constitutes an *object part?*

For a specific recognition task, some parts may carry more information than other parts, and determining those key parts and the amount of information they carry depends on the specific task. For example, human faces typically have two eyes, so that particular piece of information is not very useful in discriminating between different people, although it would be useful in differentiating human faces from clock faces. In a tactical military application, we need



(a)



(b)

FIGURE 30. Example of recognition by evidence accumulation: (a) aspect sphere and (b) recognition results (cf. Figure 24). The category sequence recognized by the ART-2 network for reentry vehicle RV-3 of Figure 29(c) is represented both on an aspect sphere seen from above in part *a,* and as a plot of category versus view number, as shown in the first graph of part *b.* The next graph shows the evidence generated for the three different reentry vehicles, the following graph shows the selected target with the maximum evidence, and the last two graphs show the differential evidence between the two highest scoring targets and this differential evidence integrated along the view sequence to give a measure of confidence. In this case, the system initially identifies the target correctly and confidence grows, although the differential evidence remains small. But the system then changes its decision, causing the confidence to be reset to zero. Finally, the system reverts to the correct identification and locks in on that decision, and confidence grows monotonically.

Table 1. Reentry-Vehicle Recognition Results

Test Vehicle	*Number of Images in Test Sequence*	*Correct Recognition*	*Number of Images Required for Convergence*	*Fraction of Rotation Cycle Required for Convergence*
RV-1	90	Yes	4	0.15
RV-2	90	Yes	9	0.33
RV-3	90	Yes	23	1.58

to determine what information is useful in recognizing the differences between the various types of vehicles that are being sought.

In our research, we use *differencing* to generate expectation-driven part segmentation cues. As with the 3-D object learning and recognition system described earlier, in Figure 31 the best-match aspect category for tank 1 can be located on the tank-1 aspect sphere. The category carries with it a learned template of the invariant appearance of the object. For the extension to part-based representations, the category must also carry a more complete description to include characteristic attributes of the object such as scale, orientation, context, and other information. (Recall the *what and where* visual pathways, and their interaction, mentioned earlier.) In addition to a description of specific views of specific objects, the system also requires a description of the views of generic (i.e., average) objects of a class. The generic-object description is necessary to represent efficiently the hierarchical descriptions that have been learned, as well as to navigate quickly through the representation during the recognition phase, as described below. A generic-object description subsumes the descriptions of all the specific objects that are associated with it. For instance, a generic cannon-tank side view is the average of the side views of all tanks that have cannons. Thus the generic view is a generalized composite representation.

After an ART-2 category is activated, the next step in the object-part process is to compare the description associated with the activated category node of tank 1 with the corresponding previously learned description for the generic tank. The differences between the two descriptions are reported in the form of a visual map called the *Saliency Map.* If all tanks have exactly the same treads, turret, and cannon, then these parts are not salient to the recognition or discrimination tasks, and they will not appear in the Saliency Map. On the other hand, if the gun is longer for tank 1 than for other tanks, then this difference will be evident in the Saliency Map, and the degree to which it is highlighted is used to prioritize the serial attentional examination strategy.

Of course, an input image can activate (to various degrees) the category nodes in many different tank ART-2 networks. Each of these category nodes has a corresponding view-description template with other associated information, including its own Saliency Map. Each tank's Saliency Map indicates which parts are most salient to discriminating that particular tank, and the saliencies predict which parts should be in the image from that vantage point, if the object in question is indeed that particular tank. The predictions become expectation-driven attentional cues for segmenting the most salient parts of the image, as shown in Figure 32. With Saliency Maps, we not only know *what* parts to look for, but we also know *where* to look for those parts relative to other parts and to the object as a whole. As each expectation is investigated, it either confirms or contradicts the hypothesized description, and evidence is accumulated or dissipated for each potential model target.

A Saliency Map can be obtained for a particular object by computing the difference between the description of the characteristic view of that object and the corresponding description of the characteristic view of the class of objects to which that particular object belongs, i.e., the generic object. Figure 31 illustrates this process. (Note: For simplicity, the Saliency Map shown in Figure 31 was derived from the original gray-scale imagery. In a complete implementation, however, the Map should be obtained from an invariant description of a view, such as a log-polar mapped image with the illuminant discounted in the case of passive visible sensors.) As an example, we might have a generic class of objects that are tanks with cannons, turrets, and treads, and included within that class we might have M48 and M60 tanks. Then the Saliency Map for the front view of an M60 tank represents the differences between the front view of the M60 and the front view of a similar class of tanks in general. Such differences are referred to as "activity"—the greater the difference in a particular area of the Saliency Map, then the greater the activity in that part. Areas in which there are no differences (i.e., no activity) are ignored in the scheduling of attentional shifts.

Using Saliency Maps, we can organize a hierarchical representation of the learned objects. Figure 33 illustrates an example hierarchy of tanks. Beginning at the upper left are the descriptions of a generic tank

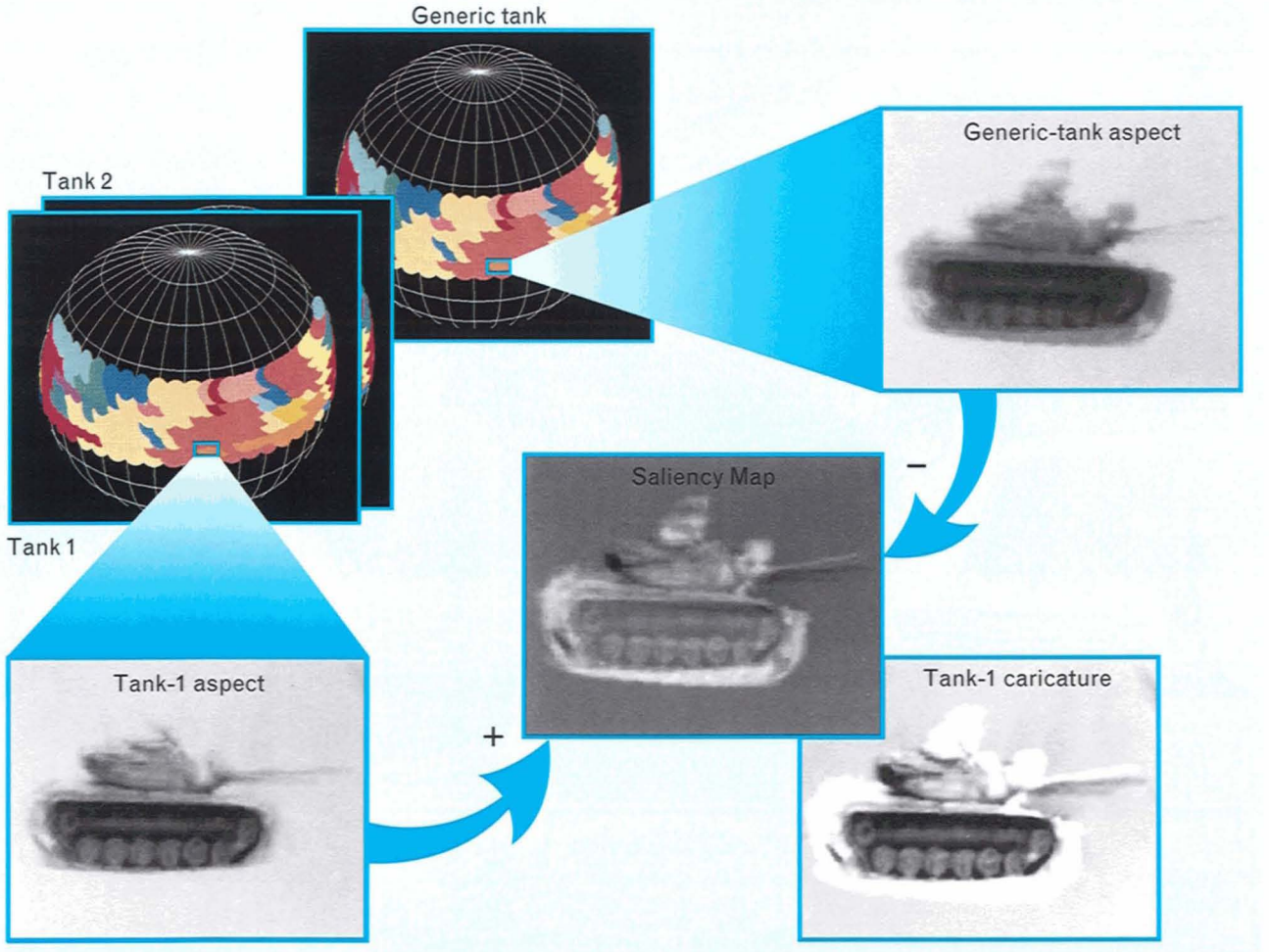


FIGURE 31. Construction of a Saliency Map and corresponding caricature image for the side view of tank 1, an M60 tank. The Saliency Map is created by taking differences between the side view of tank 1 and the side view of a similar class of tanks in general. (This class of tanks is collectively called a *generic tank*). The caricature image emphasizes the salient parts of tank 1 with respect to the generic tank. The salient parts in a Saliency Map are used to generate attentional cues for the recognition and discrimination of a particular object among similar objects, and the use of caricatures increases the efficiency of this process.

from various aspects. If all we desire is to discriminate between tanks and aircraft, then this level of description may be adequate. If, instead, we desire to discriminate between a flamethrower tank and a cannon tank, then more detailed descriptions indicating the information-carrying attributes of both types of tanks are needed. The Saliency Maps described earlier naturally contain this information, so that if an object has been determined to be a tank, the Saliency Maps indicate exactly what must be investigated to make a more refined decision about which specific tank that object is. Once the tank has been recognized as a cannon tank, either an M48 or an M60 in this example, additional Saliency Maps indicate the differences between these two types of tanks and the generic cannon tank. Although Figure 33 shows only 2-way branching, the branching often is *N*-way.

Caricatures of the object descriptions can be used to increase the efficiency of the recognition process. For the recognition of human faces, there are many different possible facial caricatures that can be used, depending on what qualities are emphasized. A caricaturist might emphasize age, sex, beauty, or simply the differences evidenced between a particular face and a corresponding generic age-matched, sex-matched face. P.J. Benson and D.I. Perrett [8] have demon-

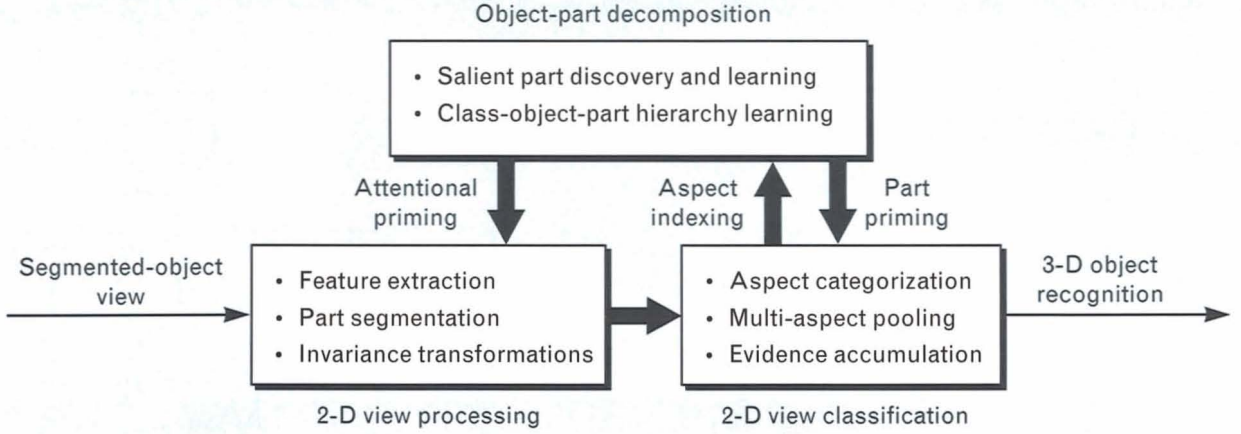


FIGURE 32. Conceptual approach to the learning and recognition of class-object-part hierarchies. Again, views are quantized into aspects through the use of unsupervised learning, but objects of a class are averaged together to form generic-object representations. Differences between specific objects and the generic object of that class are highlighted on a *Saliency Map* (Figure 31), which is then used to focus attention on salient parts during the recognition process. Recognized aspect categories for salient parts generate evidence for targets. In addition, the categories prime the system with expectations for other parts at certain locations.

strated a reduction in reaction time in the recognition task for subjects who are shown a caricatured face versus a non-caricatured face.

Caricaturing occurs naturally in the class-object-part hierarchical representations of Figure 33. Computing a difference map between a description of an input target image and a previously learned generic description leads to the detection of differences between the two descriptions. With that information, the differences can then be emphasized, resulting in a caricature of the input description. Because certain parts in the caricatured map have been exaggerated, they stand out even more strongly, and, because the non-differences have been suppressed, attention can be focused more quickly on the parts of the input image description that are most unusual and therefore most likely to carry discrimination information. Figure 31 contains an example caricature image of a tank.

Visual Navigation by MAVIN

The ATR system design described in the section "Aircraft Recognition from Visible Image Sequences" has been implemented at Lincoln Laboratory on a mobile robot called the Mobile Adaptive Visual Navigator (MAVIN). Shown in Figure 34, MAVIN can be programmed to travel a reconnaissance path, detect and track objects as it moves, and recognize objects it has learned. Arrays of light bulbs, such as the ones shown in Figure 34, have been used for the target objects. Currently, MAVIN is also able to recognize silhouettes of objects that can be segmented easily from the background. Equipped with binocular cameras, MAVIN operates in real time, with feature extraction running on a PIPE video-rate parallel-processing computer, and all other neural network computations running on SUN computers. (Capable of 1-billion 8-bit integer operations per second, PIPE was developed for robotic vision at the National Institute of Standards and Technology [NIST] and manufactured by Aspex Corp. of New York.)

Our past investigations have incorporated the visual learning and recognition system into a neural architecture that is capable of supporting various Pavlovian behavioral-conditioning paradigms based on learned associations and expectations, including excitatory conditioning, inhibitory conditioning, secondary conditioning, and the extinction of conditioned excitors [22, 23]. We have recently extended the MAVIN system to incorporate the learning and recognition of environments that are defined by the layout of visual landmarks observed during explora-

tion [24, 25]. Associative learning methods similar to those used for learning 2-D feature patterns have been applied to spatial patterns of recognized visible landmarks to establish *place cells,* which qualitatively map an environment based on its visual surroundings. We are currently incorporating *displace cells* into the architecture to code *place field* transitions that are induced by robot motions. (A place field corresponds to an area in the environment where recognized target landmarks possess a similar spatial layout.) These concepts for the qualitative mapping and navigation of space are based on behavioral experiments with rats, and on the physiological measurements of neurons in the rat hippocampus.

An important motivation for developing MAVIN has been to demonstrate in the laboratory the system's ability to recognize in real time both fixed landmarks and mobile targets from a sensor platform that can navigate through, explore, and map an environment, viewing the scene from a variety of vantage points. Indeed, MAVIN has proven to be an excellent experimental domain to test the ATR systems that we have developed.

Conclusion

Our strategy of using the unsupervised learning of view-based, invariant representations in conjunction with evidence accumulation that exploits view transitions has proven effective in several sensory domains, and is relevant to both automatic target recognition

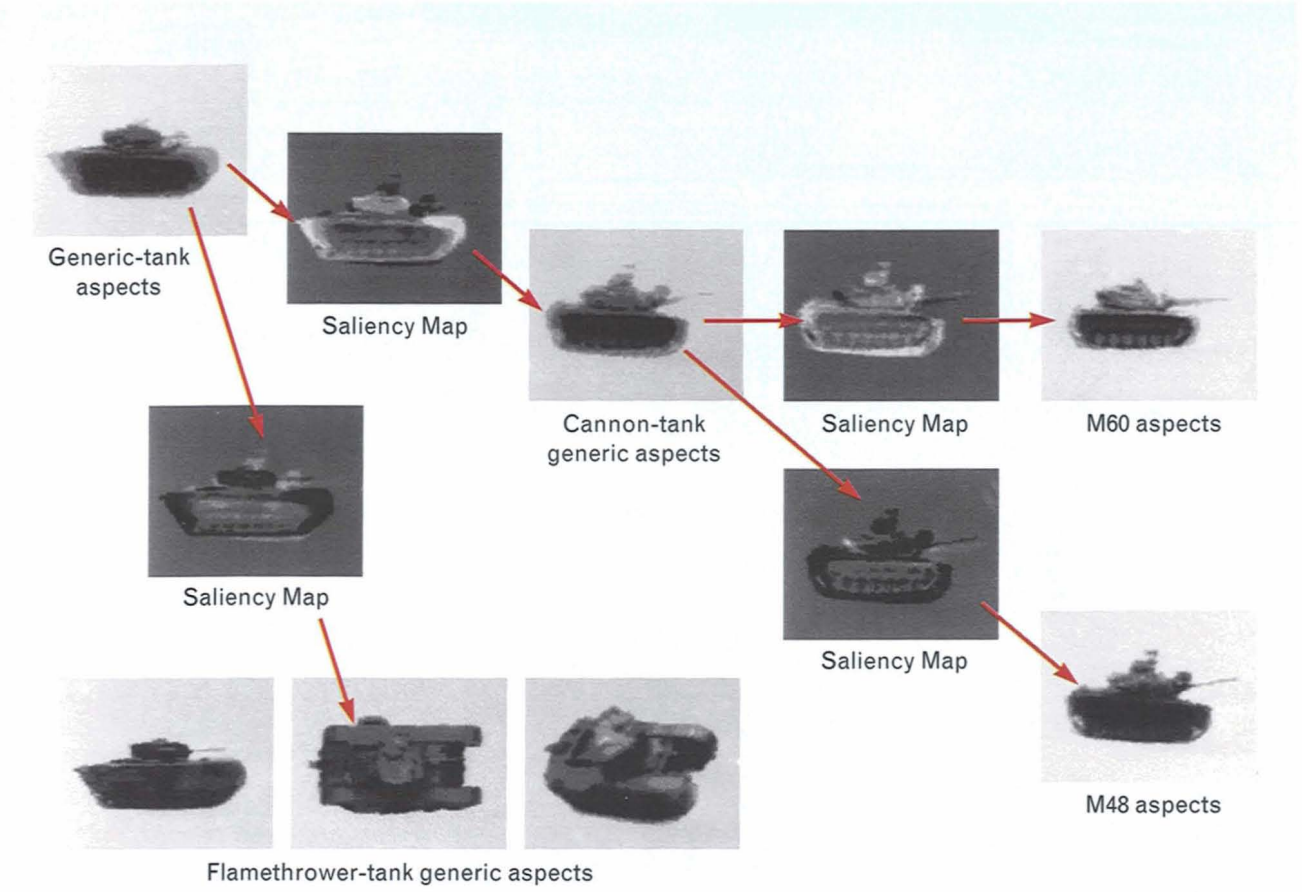


FIGURE 33. Hierarchical object representations are a natural consequence of the Saliency Map approach. The Saliency Maps direct a branching down from generic object to specific target, which may be unique because of some specific part. Because of this hierarchy, Saliency Maps can be used in the recognition process to guide a rapid search among learned categories.

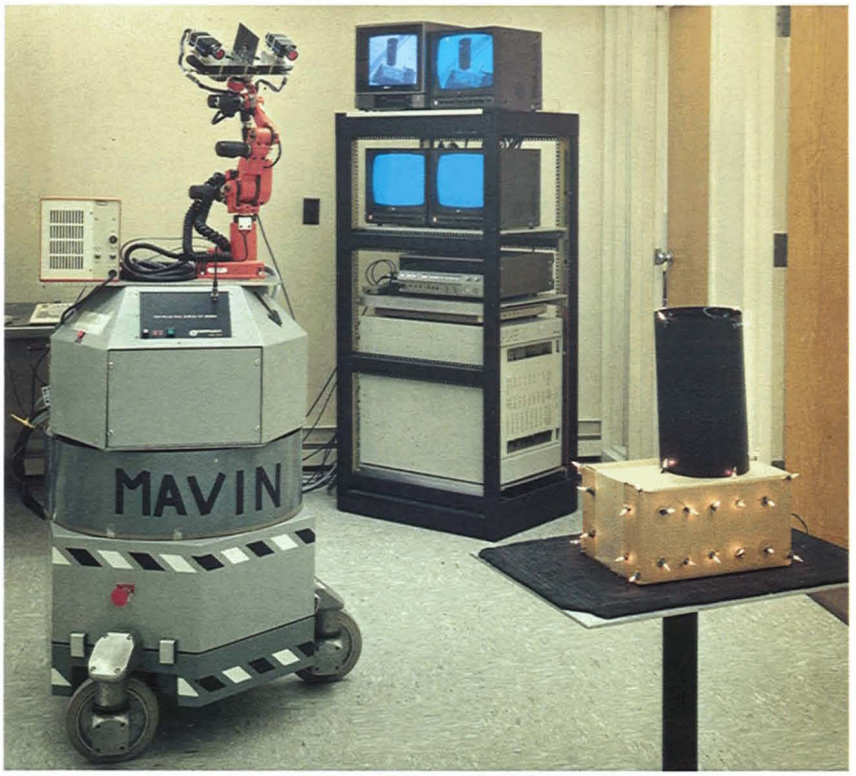


FIGURE 34. The Mobile Adaptive Visual Navigator (MAVIN) developed at Lincoln Laboratory. MAVIN, a mobile robot with binocular cameras, provides a testbed for a passive-vision ATR system in which the concepts that underlie 3-D object learning and recognition have been extended to the learning of representations for environments that are defined by distributions of visual landmarks. This extension supports the ability for an autonomous sensor platform to explore, map (in a qualitative fashion), and navigate through environments consisting of fixed landmarks and moving targets. The neural architecture being developed is based on studies of the rat hippocampus.

(ATR) and environment navigation. But perhaps the most important lesson we have learned is that many valuable insights can be gained from serious study of the brain and behavior. Anatomical, physiological, and psychophysical studies have all helped shape the computational theories and system architectures used in our work. We believe that such studies will continue to enable rapid progress in the ATR field.

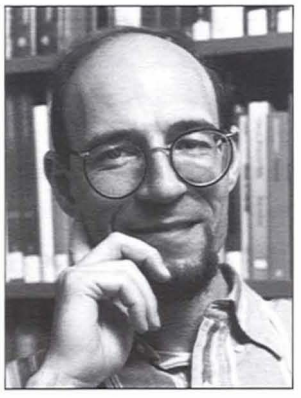
Acknowledgments

We wish to thank the members of the Surveillance Systems Group at Lincoln Laboratory for providing us with the ADTS imagery of clutter and tactical targets, as well as the radar phase history data for the ISAR targets. We are grateful to Jacques Verly and Carol Lazott of the Machine Intelligence Technology Group for constructing the ISAR target imagery from the radar phase histories. We are also indebted to the Signature Studies and Analysis Group for providing us with the range-Doppler peak data. This work on reentry-vehicle recognition was done in conjunction with Bob Gabel of the Machine Intelligence Technology Group. We also wish to acknowledge our ongoing collaboration with Professors Stephen Grossberg and Ennio Mingolla of Boston University's Department of Cognitive and Neural Systems.

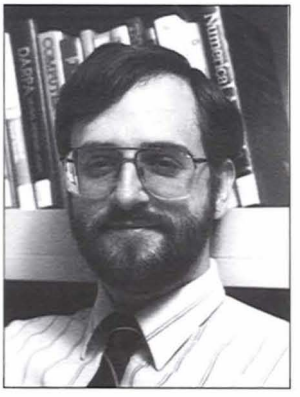
This work has been supported by the U.S. Department of the Air Force and the Office of Naval Research.

REFERENCES

1. E.A. DeYoe and D.C. Van Essen, "Concurrent Processing Streams in Monkey Visual Cortex," *Trends in Neuroscience* **TINS-11,** 219 (1988).
2. S. Zeki, "The Visual Image in Mind and Brain," *Scientific American* **267,** 68 (Sept. 1992).
3. D.A. Fay and A.M. Waxman, "Neurodynamics of Real-Time Image Velocity Extraction," chap. 9 in *Neural Networks for Vision and Image Processing,* eds. G.A. Carpenter and S. Grossberg (MIT Press, Cambridge, MA, 1992), pp. 221–246.
4. M. Mishkin, L.G. Ungerleider, and K.A. Macko, "Object Vision and Spatial Vision: Two Cortical Pathways," *Trends in Neuroscience* **TINS-6,** 414 (1983).
5. D.I. Perrett, A.J. Mistlin, and A.J. Chitty, "Visual Neurones Responsive to Faces," *Trends in Neurosciences* **TINS-10,** 358 (1987).
6. D.I. Perrett, M.H. Harries, R. Bevan, S. Thomas, P.J. Benson, A.J. Mistlin, A.J. Chitty, J.K. Hietanen, and J.E. Ortega, "Frameworks of Analysis for the Neural Representation of Animate Objects and Actions," *Journal of Experimental Biology* **146,** 87 (1989).
7. D.I. Perrett, M.W. Oram, M.H. Harries, R. Bevan, J.K. Hietanen, P.J. Benson, and S. Thomas, "Viewer-Centred and Object-Centred Coding of Heads in the Macaque Temporal Cortex," *Experimental Brain Research* **86,** 159 (1991).
8. P.J. Benson and D.I. Perrett, "Perception and Recognition of Photographic Quality Facial Caricatures: Implications for the Recognition of Natural Images," *European Journal of Cognitive Psychology* **3,** 105 (1991).
9. M. Seibert and A.M. Waxman, "Spreading Activation Layers, Visual Saccades, and Invariant Representations for Neural Pattern Recognition Systems," *Neural Networks* **2,** 9 (1989).
10. M. Seibert and A.M. Waxman, "Learning and Recognizing 3D Objects from Multiple Views in a Neural System," chap. II.12 in *Neural Networks for Perception,* vol. 1, ed. H. Wechsler (Academic Press, New York, 1991), pp. 426–444.
11. M. Seibert and A.M. Waxman, "Adaptive 3-D Object Recognition from Multiple Views," *IEEE Trans. Pattern Anal. Mach. Intell.* **14,** 107 (1992).
12. J.J. Koenderink and A.J. van Doorn, "The Internal Representation of Solid Shape with Respect to Vision," *Biological Cybernetics* **32,** 211 (1979).
13. S. Grossberg, "Nonlinear Neural Networks: Principles, Mechanisms, and Architectures," *Neural Networks* **1,** 17 (1988).
14. R.K. Cunningham and A.M. Waxman, "Astroglial-Neural Networks, Diffusion-Enhancement Bilayers, and Spatio-Temporal Grouping Dynamics," *SPIE* **1611,** 411 (1991).
15. R.K. Cunningham and A.M. Waxman, "Parametric Study of Diffusion-Enhancement Networks for Spatiotemporal Grouping in Real-Time Artificial Vision," *Technical Report No. ESC-TR-92-207,* MIT Lincoln Laboratory (6 Apr. 1993).
16. E.L. Schwartz, "Computational Anatomy and Functional Architecture of Striate Cortex: A Spatial Mapping Approach to Perceptual Coding," *Vision Research* **20,** 645 (1980).
17. G.A Carpenter and S. Grossberg, *Pattern Recognition by Self-Organizing Neural Networks* (MIT Press, Cambridge, MA, 1991), chaps. 9–15.
18. B. Widrow and S. Stearns, *Adaptive Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1985).
19. S. Grossberg, *Neural Networks and Natural Intelligence* (MIT Press, Cambridge, MA, 1988), chaps. 1–4.
20. A.M. Aull, R.A. Gabel, and T.J. Goblick, "Real-Time Radar Image Understanding: A Machine-Intelligence Approach," *Linc. Lab. J.* **5,** 195 (1992).
21. M. Seibert and A.M. Waxman, "An Approach to Face Recognition Using Saliency Maps and Caricatures," *Proc. World Congress on Neural Networks,* Portland, OR (to be published in July 1993).
22. A.A. Baloch and A.M. Waxman, "Visual Learning, Adaptive Expectations, and Behavioral Conditioning of the Mobile Robot MAVIN," *Neural Networks* **4,** 271 (1991).
23. A.A. Baloch and A.M. Waxman, "Behavioral Conditioning of the Mobile Robot MAVIN," chap. 6 in *Neural Networks: Concepts, Applications, and Implementations,* vol. IV, eds. P. Antognetti and V. Milutinovic (Prentice-Hall, Englewood Cliffs, NJ, 1991), pp. 162–200.
24. I.A. Bachelder and A.M. Waxman, "Neural Networks for Mobile Robot Visual Exploration," *SPIE* **1831,** 107 (1993).
25. I.A. Bachelder, A.M. Waxman, and M. Seibert, "A Neural System for Mobile Robot Visual Place Learning and Recognition," *Proc. World Congress on Neural Networks,* Portland, OR (to be published in July 1993).

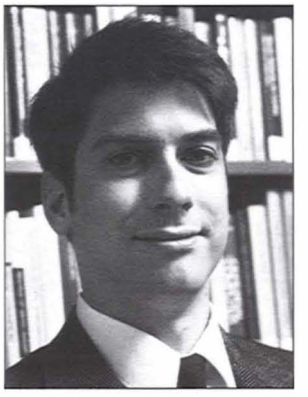


ALLEN M. WAXMAN
is a senior staff member in the Machine Intelligence Technology Group, where his focus on research has been in vision processing, neural networks, mobile robots, and electronic aids for the visually impaired. Allen also currently holds a joint appointment with the Center for Adaptive Systems at Boston University. Before joining Lincoln Laboratory four years ago, he was with the Department of Electrical, Computer, and Systems Engineering at B.U. He received a B.S. degree in physics from the City College of New York, and a Ph.D. degree in astrophysics from the University of Chicago. In 1992, he was the corecipient (with Michael Seibert) of the Outstanding Research Award from the International Neural Network Society.



MICHAEL SEIBERT
received a B.S. and an M.S. degree in computer and systems engineering from the Rensselaer Polytechnic Institute, and a Ph.D. degree in computer engineering from Boston University. He has been with Lincoln Laboratory for six years; he is currently a staff member in the Machine Intelligence Technology Group. Michael's focus on research has been in vision and neural networks, and in 1992 he was the corecipient (with Allen M. Waxman) of the Outstanding Research Award from the International Neural Network Society.

ANN MARIE BERNARDON
is a staff member in the Machine Intelligence Technology Group, where her research has been on machine intelligence, neural networks, and signal processing. Before joining Lincoln Laboratory seven years ago, she worked for Voice Processing Inc. She received the following degrees in electrical engineering: a B.S. from Purdue University and an S.M. from MIT.



DAVID A. FAY
received a B.S. degree in computer engineering and an M.A. degree in cognitive and neural systems from Boston University. While pursuing his graduate studies at B.U., Dave joined Lincoln Laboratory in 1989, and is currently a staff member in the Machine Intelligence Technology Group. His research has been on the development of neural network systems for enhancing radar imagery.