

# ARTMAP-FD: Familiarity Discrimination Applied to Radar Target Recognition

Gail A. Carpenter, Mark A. Rubin, William W. Streilein

Department of Cognitive and Neural Systems

677 Beacon Street

Boston University

Boston, MA 02215 USA

Email: gail@cns.bu.edu, rubin@cns.bu.edu, wws@cns.bu.edu

## Abstract

*ARTMAP-FD extends fuzzy ARTMAP to perform familiarity discrimination. That is, the network learns to abstain from meaningless guesses on patterns not belonging to a class represented in the training set. ARTMAP-FD can also be applied in conjunction with sequential evidence accumulation. Its performance is illustrated here on simulated radar range profile data.*

## 1 Introduction

The recognition process involves both identification and familiarity discrimination. Consider, for example, a neural network designed to identify aircraft based on their radar reflections and trained on sample reflections from ten types of aircraft  $A \dots J$ . After training, the network should correctly classify radar reflections belonging to the familiar classes  $A \dots J$ , but it should also abstain from making a meaningless guess when presented with a radar reflection from an object belonging to a different, unfamiliar class. Many neural networks carry out pattern recognition, but most perform identification without first estimating whether a test set input belongs to a class that became familiar during training. Supervised and unsupervised networks that carry out familiarity discrimination [1]-[13] include some that use a positive ARTMAP baseline vigilance value during testing. In the benchmark application developed here, however, this approach was not as successful as the ARTMAP-FD method, that instead uses the ARTMAP choice function to estimate familiarity. This paper describes ARTMAP-FD, an extension of fuzzy ARTMAP that performs familiarity discrimination. ARTMAP-FD

capabilities are demonstrated on data sets consisting of simulated radar range profiles from aircraft targets, with performance evaluated using receiver operating characteristic (ROC) curves. Section 2 summarizes the dynamics of a fuzzy ARTMAP system for classification. Section 3 defines a familiarity function and describes its role in the ARTMAP-FD network, both for individual inputs and for input sequences associated with a given target. Section 4 describes the radar range profiles that are used as simulation inputs in Section 5. In these simulations, multiwavelength input vectors can have as many as 2400 components, so the application uses the ARTMAP properties of scalability and fast learning in an essential way. Finally, Section 6 discusses selection of the familiarity threshold.

## 2 Fuzzy ARTMAP

Fuzzy ARTMAP [14] is a self-organizing neural network for learning, recognition, and prediction. Figure 1 illustrates a fuzzy ARTMAP system for classification problems, where each input  $\mathbf{a}$  learns to predict an output class  $K$ . During training, the network creates internal recognition categories, with the number of categories determined on-line by predictive success. Components of the vector  $\mathbf{a}$  are scaled so that each  $a_i \in [0, 1]$  ( $i = 1 \dots M$ ). Complement coding [15] doubles the number of components in the input vector, which becomes  $\mathbf{A} \equiv (\mathbf{a}, \mathbf{a}^c)$ , where the  $i^{th}$  component of  $\mathbf{a}^c$  is  $a_i^c \equiv (1 - a_i)$ . With fast learning, the weight vector  $\mathbf{w}_j$  records the largest and smallest component values of input vectors placed in the  $j^{th}$  category. The  $2M$ -dimensional vector  $\mathbf{w}_j$  may be visualized as the hyperbox  $R_j$  that just encloses all the vectors  $\mathbf{a}$  that selected

category  $j$  during training.

Activation of the coding field  $F_2$  is determined by the Weber law choice function  $T_j$ :

$$T_j(\mathbf{A}) = \frac{|\mathbf{A} \wedge \mathbf{w}_j|}{\alpha + |\mathbf{w}_j|}, \quad (1)$$

where  $(\mathbf{P} \wedge \mathbf{Q})_i \equiv \min(P_i, Q_i)$  and  $|\mathbf{P}| \equiv \sum_{i=1}^{2M} |P_i|$ . With winner-take-all coding, the  $F_2$  node  $J$  that receives the largest  $F_1 \rightarrow F_2$  input  $T_j$  becomes active. Node  $J$  remains active if it satisfies the matching criterion:

$$\frac{|\mathbf{A} \wedge \mathbf{w}_j|}{|\mathbf{A}|} = \frac{|\mathbf{A} \wedge \mathbf{w}_j|}{M} > \rho, \quad (2)$$

where  $\rho \in [0, 1]$  is the dimensionless *vigilance parameter*. Otherwise, the network resets the active  $F_2$  node and searches until  $J$  satisfies (2). If node  $J$  then makes an incorrect class prediction, a *match tracking* signal raises vigilance just enough to induce a search, which continues until either some  $F_2$  node becomes active for the first time, in which case  $J$  learns the correct output class label  $k(J) = K$ ; or a node  $J$  that has previously learned to predict  $K$  becomes active. During testing, a pattern  $\mathbf{a}$  that activates node  $J$  is predicted to belong to the class  $K = k(J)$ .

### 3 Familiarity discrimination with ARTMAP-FD

#### 3.1 Familiarity measure

During testing, an input pattern  $\mathbf{a}$  is defined as *familiar* when a familiarity function  $\phi(\mathbf{A})$  is greater than a decision threshold  $\gamma$ . Section 6 discusses how to choose  $\gamma$  for a given application. Once a category choice has been made by the winner-take-all rule, fuzzy ARTMAP ignores the size of the input  $T_j$ . In contrast, ARTMAP-FD uses  $T_j$  to define familiarity, taking

$$\phi(\mathbf{A}) = \frac{T_j(\mathbf{A})}{T_j^{MAX}} = \frac{|\mathbf{A} \wedge \mathbf{w}_j|}{|\mathbf{w}_j|}, \quad (3)$$

where  $T_j^{MAX} = |\mathbf{w}_j| / (\alpha + |\mathbf{w}_j|)$ . This value is attained by each input  $\mathbf{a}$  that lies in the hyperbox  $R_J$ , since  $|\mathbf{A} \wedge \mathbf{w}_j| = |\mathbf{w}_j|$  for these points. An input that chooses category  $J$  during testing is then assigned the maximum familiarity value 1 if and only if  $\mathbf{a}$  lies within  $R_J$ .

Note that the choice parameter  $\alpha$  in equation (1) is usually taken to be small since the *conservative limit*, where  $\alpha = 0^+$ , minimizes the number of category nodes

formed during training. When  $\alpha \approx 0$ ,  $T_j^{MAX} \approx 1$ , so  $\phi(\mathbf{A}) \approx T_j(\mathbf{A})$ . Simulations below set  $\alpha = 0.0001$ . Then, setting  $\phi(\mathbf{A}) = T_j(\mathbf{A})$  produces essentially the same results as setting  $\phi(\mathbf{A}) = T_j(\mathbf{A})/T_j^{MAX}$ . The former choice of  $\alpha$  is more readily computable in a neural network but the latter has a simpler geometric interpretation.

#### 3.2 Familiarity discrimination algorithm

ARTMAP-FD is identical to fuzzy ARTMAP during training. During testing,  $\phi(\mathbf{A})$  is computed after fuzzy ARTMAP has yielded a winning node  $J$  and a predicted class  $K = k(J)$ . If  $\phi(\mathbf{A}) > \gamma$ , ARTMAP-FD predicts class  $K$  for the input  $\mathbf{a}$ . If  $\phi(\mathbf{A}) \leq \gamma$ ,  $\mathbf{a}$  is regarded as belonging to an unfamiliar class and the network makes no prediction.

Note that fuzzy ARTMAP can also abstain from classification, when the baseline vigilance parameter  $\bar{\rho}$  is greater than zero during testing. Typically  $\bar{\rho} = 0$  during training, to maximize code compression. In radar range profile simulations such as those described below, fuzzy ARTMAP can perform familiarity discrimination when  $\bar{\rho} > 0$  during both training and testing. However, accurate discrimination requires that  $\bar{\rho}$  be close to 1, which causes category proliferation during training.

Range profile simulations have also set  $\bar{\rho} = 0$  during both training and testing, but with the familiarity measure set equal to the fuzzy ARTMAP match function:

$$\phi(\mathbf{A}) = \frac{|\mathbf{A} \wedge \mathbf{w}_J|}{M}. \quad (4)$$

This approach is essentially equivalent to taking  $\bar{\rho} = 0$  during training and  $\bar{\rho} > 0$  during testing, with  $\bar{\rho} = \gamma$ . However, when a test set input  $\mathbf{a} \in R_J$ , the function defined by (4) sets  $\phi(\mathbf{A}) = |\mathbf{w}_J| / M$ , which may be large or small. Thus this function does not provide as good familiarity discrimination as the one defined by (3), which always sets  $\phi(\mathbf{A}) = 1$  when  $\mathbf{a} \in R_J$ . All the simulations below employ the function (3) with  $\bar{\rho} = 0$ .

#### 3.3 Familiarity discrimination with sequential evidence accumulation

ART-EMAP (Stage 3) [16] identifies a test set object's class after exposure to a sequence of input patterns, such as differing views, all identified with that one object. Training is identical to that of fuzzy ARTMAP, with winner-take-all coding at  $F_2$ . ART-EMAP generally employs distributed  $F_2$  coding during testing. With winner-take-all coding during testing as well as training,

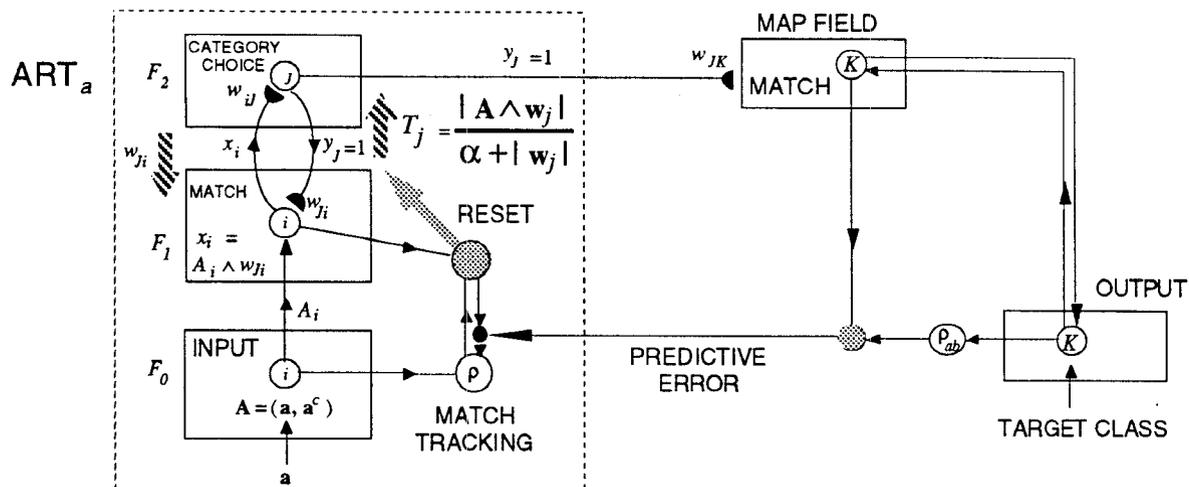


Figure 1. A fuzzy ARTMAP network for classification.

ART-EMAP predicts the object's class to be the one selected by the largest number of inputs in the sequence. Extending this approach, ARTMAP-FD accumulates familiarity measures at each predicted class  $K$  as the test set sequence is presented. Once the winning class is determined, the objects familiarity is defined as the average accumulated familiarity measure of the predicted class.

## 4 Radar range profiles

A radar range profile is a one-dimensional representation of a target, produced from a recording of a radar pulse reflection at high temporal resolution [17]-[19]. A continuous radar return is quantized into range bins that sample information about the target over a downrange extent  $\Delta x$ . The signal, integrated over each bin, produces a discrete radar range profile vector (Figure 2a, top). Several range profiles, constructed from the same view of the target but using pulses of different center frequencies, can also be concatenated to form a multiwavelength radar range profile [20,21] (Figure 2a, bottom). The simulations presented here have multiwavelength range profiles with center frequencies evenly spaced between 18GHz and 22GHz. In the simulations,  $\Delta x = 2/3$  m and the range profile covers 40m, so the number of components in a range profile equals the number of center frequencies times 60. Simulations below use 2, 10, or 40 center frequencies, yielding input vectors  $\mathbf{a}$  of size  $M = 120, 600, \text{ or } 2400$ .

Range profiles are here simulated by computing, in the

far-field approximation, reflections from 100 scattering centers placed randomly on each image. Images represent a set of two-dimensional "aircraft" that differ from one another by wing position and wing length (Figure 2b). Each set of targets represents airplanes with wing positions that range from the middle of the fuselage to near the tail and with wing lengths whose range is independent of the set size. Larger sets thus contain more targets and targets that are more similar to one another. Figure 2b depicts the scattering centers of a set of targets with 6 wing positions and 6 wing lengths. The network is trained on range profiles generated by 18 targets (in boxes), which define the set of familiar classes.

The ARTMAP-FD network is trained on range profiles obtained from 21 viewing angles in the plane of the targets, evenly spaced  $0.5^\circ$  apart over a range of  $10^\circ$ , centered on the front of the target. At each viewing aspect and for each familiar target, training set range profiles are computed with 15 downrange shifts evenly spaced from  $-1/2$  to  $+1/2$  bin widths. The trained network is tested on at least 2000 range profiles of all the targets, familiar and unfamiliar, taken at random angles within the  $10^\circ$  range and with random shifts of the distance to the target spanning one half the downrange extent of the range profile. A temporal sequence of range profiles is generated using a model of small-amplitude stochastic fluctuations of the heading of an aircraft attempting to fly in a fixed direction. Twenty sequential views correspond to one second of observation time. (For a more extensive discussion of the procedures

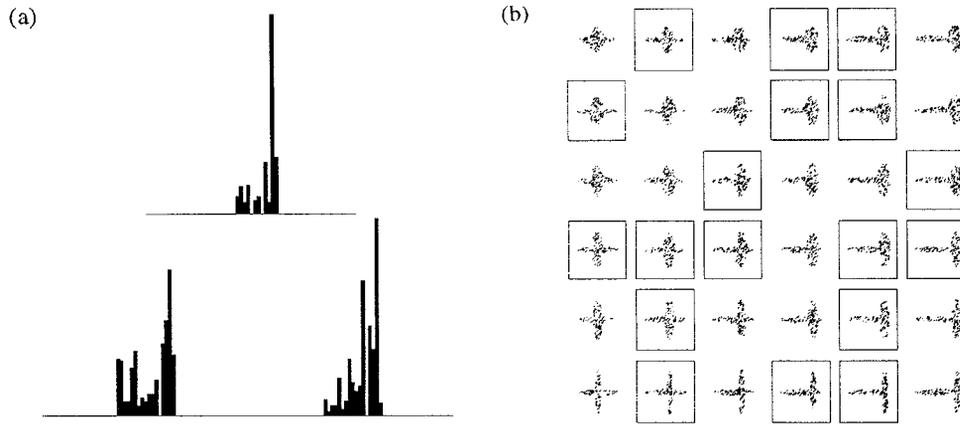


Figure 2. (a) Simulated range profiles. Top: Single wavelength. Bottom: Multiwavelength, with two center frequencies. (b) 36 simulation targets with 6 wing positions and 6 wing lengths and 100 scattering centers per target. Boxes indicate randomly selected familiar targets.

for generating simulated radar range profiles, see [21].)

## 5 Familiarity discrimination simulations

Since familiarity discrimination involves placing an input into one of two sets, familiar and unfamiliar, the receiver operating characteristic (ROC) formalism [22,23] can be used to evaluate the effectiveness of ARTMAP-FD on this task. The *hit rate*  $R_H$  is the fraction of familiar targets the network correctly identifies as familiar and the *false alarm rate*  $R_F$  is the fraction of unfamiliar targets the network incorrectly identifies as familiar. Each of these quantities depends upon a decision threshold familiarity parameter  $\gamma$ . An ROC curve is a plot of  $R_H$  vs.  $R_F$ , parameterized by  $\gamma$ . With  $\gamma = 0$ , all inputs meet the familiarity criterion, so the curve begins in the upper right-hand corner. There, the hit rate  $R_H$  equals 1 but the false-alarm rate  $R_F$  also equals 1. As  $\gamma$  increases, the ROC curve moves toward the lower left-hand corner, where  $\gamma = 1$ . Then, all inputs are regarded as unfamiliar and  $R_H = R_F = 0$ . Good discrimination potential is characterized by an ROC curve that approaches the upper left-hand corner of the square, the point where all true positives are identified ( $R_H = 1$ ) without any false positives ( $R_F = 0$ ). The area under the ROC curve is the *c-index*, a measure of predictive accuracy that is independent of both the fraction of positive cases in the test set and the positive-case decision threshold  $\gamma$ .

Figure 3a shows ROC curves for a network trained on 2 targets from a 4-target set (the upper-right and lower right corner targets out of the 4 corner targets in Figure 2b). Successive curves show simulation results for range

profiles having 2, 10, and 40 center frequencies. Just as an increase in the number of center frequencies in a range profile increases the accuracy of classification on test sets with purely familiar targets [21], increasing the number of frequencies also increases the network's ability to distinguish between familiar and unfamiliar targets.

A larger target set makes classification more difficult, even with many center frequencies in each range profile. Familiarity discrimination is more difficult as well, but is again improved by sequential evidence accumulation. This can be seen from the ROC curves in Figure 3b, obtained from 18 familiar targets and 18 unfamiliar targets selected at random from a set of 36 targets (Figure 2b). Sequential evidence accumulation was performed for 1, 3, and 100 observations, corresponding to 0.05, 0.15, and 5.0 seconds of observation time.

## 6 Familiarity threshold selection

The c-index and the shape of the ROC curve measure the network's potential ability to discriminate between familiar and unfamiliar targets. However, when the network is placed in operation, one particular decision threshold  $\gamma = \Gamma$  must be chosen. The optimal  $\Gamma$  corresponds to a point on the parameterized ROC curve that is typically close to the upper left-hand corner of the unit square [22,23], to maximize correct selection of familiar targets while minimizing incorrect selection of unfamiliar targets. In a given application, selection of  $\Gamma$  depends upon the relative cost of errors due to missed targets and false alarms. The value of  $\Gamma$  can be determined by a validation procedure [23].

Because of noise and varying target patterns encountered during operation, the robustness of the choice of

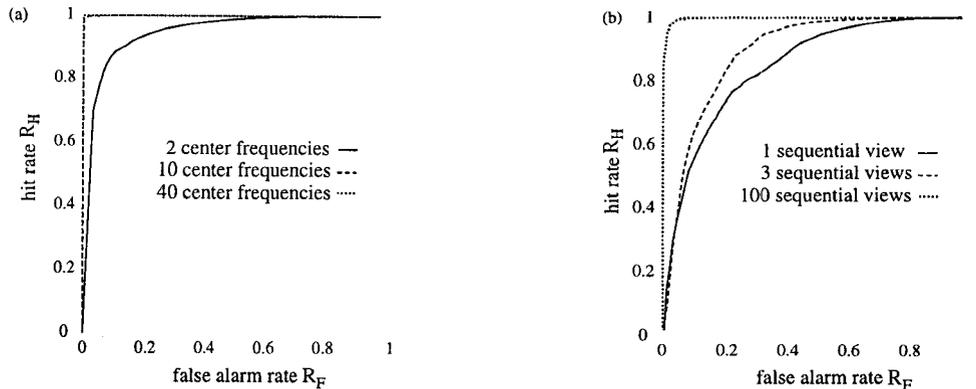


Figure 3. ROC curves from ARTMAP-FD simulations. (a) 4 targets, with multiwavelength range profiles having 2, 10, and 40 center frequencies, with ROC curves for 10 and 40 center frequencies lying almost on the edges of the unit square. Classification accuracy for familiar targets is 77.4%, 95.9%, and 99.2% for 2, 10, and 40 center frequencies, respectively, among 8000 test patterns. The network created 7, 3, and 3 category nodes. (b) 36 targets, with multiwavelength range profiles having 40 center frequencies, with sequential evidence accumulation for 1, 3 and 100 views. Classification accuracy for familiar targets is: 89.5%, 97.0%, and 100.0% for 1, 3, and 100 sequential inputs, among 2016 test pattern sequences. The network created 44 category nodes.

the optimal  $\gamma = \Gamma$  is an important factor in the success of applications. To see the effect of noise in the current simulations, consider the ROC curve from an ARTMAP-FD network trained on 2 familiar targets out of the 4-target set with 40 center frequencies, and tested on range profiles from all 4 targets (Figure 3a). At the point where the curve almost reaches the upper left-hand corner of the box,  $\Gamma = 0.9989$ , which gives a hit rate  $R_H = 0.9997$  and a false-alarm rate  $R_F = 0.0003$ . When 1% Gaussian noise is added to the test range profiles, the ROC curve looks exactly like the noise-free curve, but the corner threshold is now  $\Gamma = 0.9986$ . If this ideal value were known and used during testing, the hit rate would be  $R_H = 0.9998$  and the false alarm rate would be  $R_F = 0.0$ . If, however, the larger parameter value  $\Gamma = 0.9989$ , obtained in noise-free simulations, were being used in a noisy fielded application where a new  $\Gamma$  could not be computed, the hit rate would have been slightly lower ( $R_H = 0.9973$ ), with the false-alarm rate remaining  $R_F = 0.0$ .

Similarly, unfamiliar targets in the field may differ from unfamiliar test-set targets that provided an ideal threshold value. Consider again the 40-center-frequency simulation but with only one of the two unfamiliar targets present during testing. This situation would predict that  $\Gamma = 0.9985$  using one of the targets and  $\Gamma = 0.9989$  using the other target. Both values are close to the ideal threshold  $\Gamma = 0.9989$  obtained when both targets are present in the test set. Similar considerations apply to discrimination between the presence or absence of a target in a noisy environment. In that task, the constant-false-alarm-rate (CFAR) technique [22] estimates the

noise level and adjusts  $\gamma$  accordingly. An approach of this type may be useful in improving the robustness of familiarity discrimination in the presence of noise. Modified familiarity measures that improve robustness while retaining effective familiarity discrimination are currently being investigated.

## Acknowledgements

This research was supported in part by grants from the Advanced Research Projects Agency, the National Science Foundation, and the Office of Naval Research: NSF IRI-94-01659 (G. A. C.), ONR N00014-95-1-0657 (G. A. C.), ONR N00014-95-0409 (G. A. C., M. A. R.), and ONR N00014-96-0659 (M. A. R., W. W. S.). M. A. R. would like to thank Brett Borden and Michael Cohen for helpful discussions.

## References

- [1] Carpenter, G. A., Gjaja, M. N., Gopal, S., & Woodcock, C. E., ART networks for remote sensing: vegetation classification from Landsat TM and terrain data, *IEEE Transactions on Geoscience and Remote Sensing*, 1996.
- [2] Casasent, D. P., & Neiberg, L., Time-sequential neural net classifier and pose estimator, *Proceedings of the World Congress on Neural Networks*, 345-350, 1996.

- [3] Chen, K., Xie, D., & Huisheng, C., Speaker identification based on input/output HMMs, *Proceedings of the World Congress on Neural Networks*, 37-40, 1996.
- [4] Koch, M. W., Moya, M. M., Hostetler, L. D., & Fogler, R. J., Cueing, feature discovery, and one-class learning for synthetic aperture radar automatic target recognition, *Neural Networks*, **8**, 1081-1102, 1995.
- [5] Kohonen, T., & Oja, E., Fast adaptive formation of orthogonalizing filters and associative memory in recurrent networks of neuron-like elements, *Biological Cybernetics*, **21**, 85-95, 1976.
- [6] Moya, M. M., & Hostetler, L. D., One-class generalization in second-order backpropagation networks for image classification, *Proceedings of the International Joint Conference on Neural Networks*, **II**, 221-224, 1990.
- [7] Moya, M. M. & Hush, D. R., Network constraints and multi-objective optimization for one-class classification, *Neural Networks* **9**, 436-474, 1996.
- [8] Murshed, M. A., Bortolozzi, F., & Sabourin, R., A fuzzy ARTMAP-based classification system for detecting cancerous cells, based on the one-class problem approach, *Proceedings of the International Conference on Pattern Recognition*, 1996.
- [9] Parra, L., Deco, G., & Miesbach, S., Statistical independence and novelty detection with information preserving nonlinear maps. *Neural Computation*, **8**, 260-269, 1996.
- [10] Petsche, T., Marcantonio, A., Darken, C., Santoso, I., & Hanson, S. J., A neural network autoassociator for induction motor failure prediction. *Proceedings of the Neural Information Processing Systems Conference*, 1996.
- [11] Rubin, M. A., Automatic target recognition from radar range profiles using fuzzy ARTMAP. *Proceedings of the World Congress on Neural Networks*, **III**, 291-298, 1994.
- [12] Wang, L., Chen, K., & Chi, H., Text-independent speaker identification by combining classifiers with different features, *Proceedings of the World Congress on Neural Networks*, 59-62, 1996.
- [13] Waxman, A. M., Seibert, M. C., Gove, A., Fay, D. A., Bernardon, A. M., Lazott, C., Steele, W. R., & Cunningham, R. K., Neural processing of targets in visible, multispectral IR and SAR imagery, *Neural Networks*, **8**, 1029-1051, 1995.
- [14] Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., & Rosen, D. B., Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps, *IEEE Transactions on Neural Networks*, **3**, 698-713, 1992.
- [15] Carpenter, G. A., Grossberg, S., & Rosen, D. B., Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system, *Neural Networks*, **4**, 759-771, 1991.
- [16] Carpenter, G. A., & Ross, W. D., ART-EMAP: A neural network architecture for object recognition by evidence accumulation, *IEEE Transactions on Neural Networks*, **6**, 805-818, 1995.
- [17] Borden, B., Problems in airborne target recognition, *Inverse Problems*, **10**, 1009-1022, 1993.
- [18] Hudson, S., & Psaltis, D., Correlation filters for aircraft identification from radar range profiles, *IEEE Transactions on Aerospace and Electronic Systems*, **19**, 741-748, 1993.
- [19] Smith, C. R., & Goggans, P. M., Radar target identification, *IEEE Antennas and Propagation Magazine*, **35**, 27-38, 1992.
- [20] Botha, E. C., Barnard, E., & Barnard, C. J., Feature-based classification of aerospace radar targets using neural networks, *Neural Networks*, **9**, 129-142, 1996.
- [21] Rubin, M. A., Application of fuzzy ARTMAP and ART-EMAP to automatic target recognition using radar range profiles. *Neural Networks*, **8**, 1109-1116, 1995.
- [22] Helstrom, C. W., *Elements of Signal Detection and Estimation*, Prentice Hall, Englewood Cliffs, NJ, 1995.
- [23] Masters, T., *Practical Neural Network Recipes in C++*, Academic Press, San Diego, 1993.