

Biologically Inspired Approaches to Automated Feature Extraction and Target Recognition

Gail A. Carpenter, Siegfried Martens, Ennio Mingolla, Ogi J. Ogas, Chaitanya Sai
Department of Cognitive and Neural Systems
Boston University
677 Beacon Street
Boston, Massachusetts 02215 USA

[gail, sig, ennio, ogiogas, gsc] @cns.bu.edu

<http://cns.bu.edu/techlab>
<http://cns.bu.edu/celest>

AIPR 2004: 33rd Workshop on Applied Imagery Pattern Recognition
October 13-15, 2004, Washington, DC
Proceedings: Piscataway, NJ: IEEE

Technical Report CAS/CNS TR-2004-008
Boston, MA: Boston University

Acknowledgements: This work was supported by research grants from the Air Force Office of Scientific Research (AFOSR F49620-01-1-0423), the National Geospatial-Intelligence Agency (NMA 201-01-1-2016), the National Science Foundation (NSF SBE-0354378 and NSF BCS-0235398), and the Office of Naval Research (ONR N00014-01-1-0624); by postdoctoral fellowships from the National Geospatial-Intelligence Agency and the National Science Foundation for Siegfried Martens (NMA 501-03-1-2030 and NSF DGE-0221680); and by a Department of Homeland Security graduate fellowship for Ogi Ogas. Sucharita Gopal, Junchang Ju, and Mutlu Ozdogan provided remote sensing data, and Suhas Chelian and Brad Rhodes helped create the ground truth set for the Boston testbed.

Biologically Inspired Approaches to Automated Feature Extraction and Target Recognition

Gail A. Carpenter, Siegfried Martens, Ennio Mingolla, Ogi J. Ogas, Chaitanya Sai
Boston University, Department of Cognitive and Neural Systems
677 Beacon Street, Boston, MA 02215 USA
[*gail, sig, ennio, ogigas, gsc*]@cns.bu.edu

Abstract

Ongoing research at Boston University has produced computational models of biological vision and learning that embody a growing corpus of scientific data and predictions. Vision models perform long-range grouping and figure/ground segmentation, and memory models create attentionally controlled recognition codes that intrinsically combine bottom-up activation and top-down learned expectations. These two streams of research form the foundation of novel dynamically integrated systems for image understanding. Simulations using multi-spectral images illustrate road completion across occlusions in a cluttered scene and information fusion from input labels that are simultaneously inconsistent and correct. The CNS Vision and Technology Labs (cns.bu.edu/visionlab and cns.bu.edu/techlab) are further integrating science and technology through analysis, testing, and development of cognitive and neural models for large-scale applications, complemented by software specification and code distribution.

1. Introduction

Parallel research streams in the Boston University Department of Cognitive and Neural Systems (BU/CNS) have led to the development of families of biological models of early vision and attentive recognition. For vision, the Boundary Contour System and Feature Contour System (BCS/FCS) realize complementary design principles to perform long-range boundary completion and featural filling-in [1]. For recognition, Adaptive Resonance Theory (ART) models create stable recognition codes with fast or slow incremental learning and with supervised or unsupervised training [2,3]. Examples in this paper illustrate how recent research is moving toward the goal of providing open-source code for flexible, user-friendly, integrated vision and

recognition systems that bring the power of cognitive and neural computations to technological applications.

2. A default ARTMAP system for biologically based recognition learning

ART neural networks model real-time prediction, search, learning, and recognition. These systems have provided both computational models of human cognitive information processing (e.g., [4–9]) and neural methods for technology transfer (e.g., [10–12]). Sites of early and ongoing transfer of ART-based technologies include industrial venues such as the Boeing Corporation and government venues such as MIT Lincoln Laboratory. A recent report on industrial uses of neural networks [13] states: “[The] Boeing ... Neural Information Retrieval System [14] is probably still the largest-scale manufacturing application of neural networks. It uses [ART] to cluster binary templates of aeroplane parts in a complex hierarchical network that covers over 100,000 items, grouped into thousands of self-organised clusters. Claimed savings in manufacturing costs are in millions of dollars per annum.” At Lincoln Lab, a team led by Waxman developed an image mining system which incorporates several BU/CNS-based models of vision and recognition [15–17]. Over the years a dozen CNS graduates have contributed to this effort, which is now located at Alphatech, Inc.

Design principles derived from scientific analyses and design constraints imposed by targeted applications have jointly guided the development of many variants of the basic ART networks, including fuzzy ARTMAP [18], simplified fuzzy ARTMAP [19], ART-EMAP [20], ARTMAP-IC [21], Gaussian ARTMAP [22], and distributed ARTMAP [23]. Across the variations of these models, a neural computation central to both the scientific and the technological analyses is the *ART matching rule* [2], which represents the interaction between top-down learned expectation and bottom-up

sensory input. This interaction creates a focus of attention which, in turn, determines the nature of stored memories.

While the earliest unsupervised ART [2] and supervised ARTMAP networks [3] feature winner-take-all code representations, many of the networks developed over the past dozen years incorporate distributed code representations. Comparative analyses of these systems have led to the specification of a *default ARTMAP* network, which features simplicity of design and robust performance in many application domains [24]. Like ART-EMAP, default ARTMAP uses winner-take-all coding during training and distributed coding during testing. Distributed test outputs have helped improve various methods for categorical decision-making. One such method, in a map production application, compares a baseline mapping procedure, which selects the class with the largest total output, with a procedure that enforces *a priori* output class probabilities and another that selects class-specific output thresholds, via validation [25]. Distributed coding supports each method, but the ultimate prediction is still one output class per test input (Figure 1, upper right). This paper also specifies a canonical cross-validation training / testing method, which partitions the area in question into four vertical or horizontal strips. A given simulation takes training pixels from two of these strips; uses the validation strip to choose parameters, if necessary; and tests on the fourth strip. Learning methods are thus compared with training and test sets that are not only disjoint but drawn from geographically distinct locations. This separation tests for generalization to new regions, where class distributions could typically be far from those of the training and validation sets.

3. A default BCS/FCS system for biologically based image reconstruction

Over the past two decades, increasingly detailed analyses of emerging bodies of scientific data from visual psychophysics and neuroscience have, as in the ART domain, produced a variety of BCS/FCS models [e.g., 26,27]. The many variations on mechanism, anatomy, and parameter selection for these systems have impeded the transition of their full computational capabilities for large-scale, general-purpose image processing. Following the default ARTMAP paradigm, a new *default BCS/FCS* defines a clean, systematically analyzed core algorithm. Default BCS/FCS performs the basic visual functions of boundary finding, long-range grouping, and featural filling-in of connected components. Model analyses have

also fed back to the science, introducing new hypotheses concerning key neural mechanisms of early vision.

The Boundary Contour and Feature Contour Systems realize computationally complementary processes of early vision. For example, BCS stages of long-range cooperation among orientation-sensitive units are balanced by short-range competition among units of different orientational preference at nearby positions. The BCS thus seeks to find the best groupings of local contrasts relative to the degree of variability specified by image regions. Note that each stage of the BCS can also be computed in several scales, as in pyramid architectures.

Outputs of the BCS are used to contain filling-in, whereby signals from FCS contours diffuse through image areas that are not blocked by BCS boundary signals. The combined result of BCS/FCS processing is a form-sensitive set of representations of featural information, such as brightness or color, where each location's value of a featural quantity has been derived from both local bottom-up signals and from a context-sensitive comparison of these signals with inputs from the surrounding area.

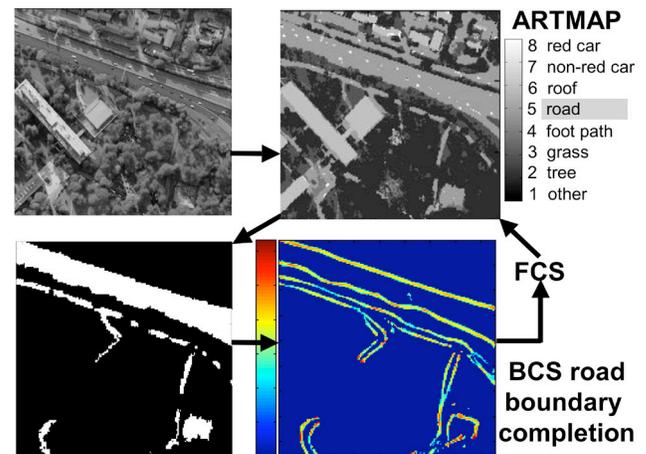


Figure 1. Pixels of an aerial image (upper left) are first assigned labels from one of eight classes (upper right) by a default ARTMAP algorithm [25]. Pixels labeled *road* produce collinear sections that are interrupted by gaps caused by overhanging trees in the original image (lower left). With this road pixel image serving as input, default BCS performs boundary finding and long-range completion (lower right). Subsequent filling-in by the default FCS produces a more meaningful labeling of the image than was possible from local computations.

4. Integrating ARTMAP and BCS/FCS models for image understanding

Figure 1 illustrates computations of a system that integrates ARTMAP recognition with BCS image processing. In this example, a Lincoln Lab image mining system [15-17] produced 20-dimensional input vectors from local neighborhoods of pixels in the image shown at the upper left. A default ARTMAP system produced a labeled map (upper right) after training on a subset of labeled feature vectors [25]. This procedure typifies a traditional approach to combining vision and recognition systems, with preattentive computations of early vision producing preprocessed inputs for learned recognition.

The integrated vision / recognition system illustrated here begins with these steps, but then feeds the results of local (pixel) recognition back to the “early” vision system for further processing. Figure 1 indicates how a default BCS model can complete road boundaries across sections occluded by overhanging trees, then, via the FCS, return newly reconstructed featural information to the recognition system, to improve image understanding.

Figure 2 shows details of long-range completion by default BCS, starting with pixel inputs from one of the occluded road segments in Figure 1.

5. A cognitive approach to information fusion

A second type of simulation example next illustrates how a cognitive and neural approach can define, as well as provide solutions for, novel problem classes.

Image fusion has been defined as “the acquisition, processing and synergistic combination of information provided by various sensors or by the same sensor in many measuring contexts.” [28, p. 3] When multiple sources provide inconsistent data, fusion methods are called upon to select the accurate information components. As quoted by the International Society of Information Fusion (www.inforfusion.org/terminology): “Evaluating the reliability of different information sources is crucial when the received data reveal some inconsistencies and we have to choose among various options.” For example, independent sources might label a small area as *residential* or *industrial* or *park*, as in Figure 3 (left). A fusion method could address this problem by weighing the confidence and reliability of each source, merging complementary information, or gathering more data. In any case, at most one of these answers is correct.

The methods illustrated here address a complementary

and previously unexamined aspect of the information fusion problem, seeking to derive consistent knowledge from sources that are inconsistent – but accurate. This is a problem that the human brain solves very well. A young child who hears the family pet variously called *Spot*, *puppy*, *dog*, *dalmatian*, *mammal*, and *animal* is not only not alarmed by these labels but readily uses them to infer functional relationships. An analogous problem for information fusion methods seeks to classify the terrain and objects in an unfamiliar territory based on

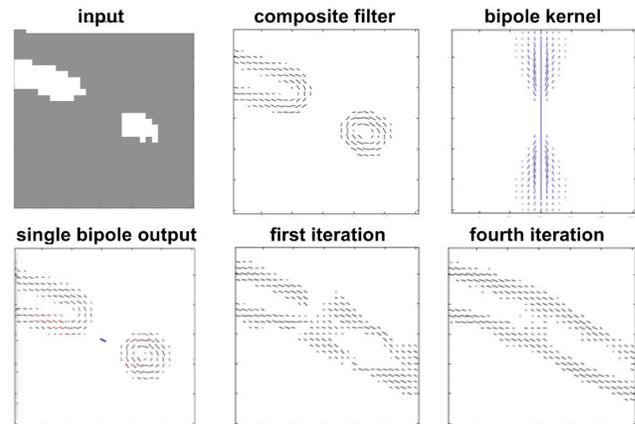


Figure 2. Details of default BCS processing for a small region of the image of Figure 1. The upper left square shows labeled road pixels (white). Local oriented-contrast filtering (upper middle) is represented with each segment’s length proportional to the degree of activation of units tuned to that segment’s defining orientation. The bipole kernel (upper right) displays the set of weights that express the connection strength among units of various orientations. The central unit of the bipole shown here would support completion of vertical contours if evidence from several nearby locations were sufficiently strong. The lower left square shows the result of one such completion event for a single bipole of diagonal orientational preference whose center is located midway between the pools of activation. By the first iteration of bipole processing (lower middle), the system has started to produce long-range completion and strengthening of orientations that are statistically coherent across space. This process converges quickly (lower right). Note that completion is influenced at the lower right of each square by other road pixels that are beyond the border of this detail.

intelligence supplied by several reliable experts, as in Figure 3 (right). Each expert labels a portion of the region based on sensor data and observations collected at specific times and based on individual goals and interests. Across experts, a given area might be correctly, though inconsistently, labeled *residential* and *built-up* and *man-made*. A human mapping analyst would, in this case, be able to apply a lifetime of experience to resolve the paradox by placing objects in a knowledge hierarchy, and a rule-based expert system could be constructed to codify this knowledge.

6. Deriving consistent knowledge from inconsistent information

An ARTMAP neural network can act as a self-organizing expert system to derive hierarchical knowledge structures from inconsistent training data. This ability is implicit in the network's learning strategy, which creates one-to-many, as well as many-to-one,

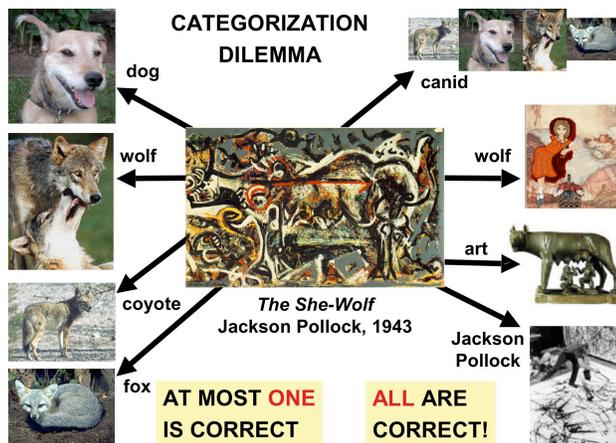


Figure 3. Classifying novel terrain or objects from sparse, complex data may require the resolution of conflicting information from sensors working at different times, locations, and scales, and from sources with different goals and situations. Information fusion methods can help resolve inconsistencies, as when evidence variously suggests that an object's class is *dog*, *wolf*, *coyote*, or *fox* (left). The methods summarized here consider a complementary problem, supposing that information from sensors and experts is reliable though inconsistent, as when evidence suggests that an object's class is correctly described as *canid*, *wolf*, *art*, and a *Jackson Pollock* (right).

maps of the input space. During training, the system can learn that disparate pixels map to the output class *residential*; but, if similar or identical pixels are later labeled *built-up* or *man-made*, the system can learn to associate numerous output classes with a given input (Figure 4). During testing, distributed code activations predict multiple output class labels. A rule-production algorithm uses these distributed outputs to derive a knowledge hierarchy for the output classes. The resulting diagram of the relationships among classes can then guide the construction of consistent layered maps.

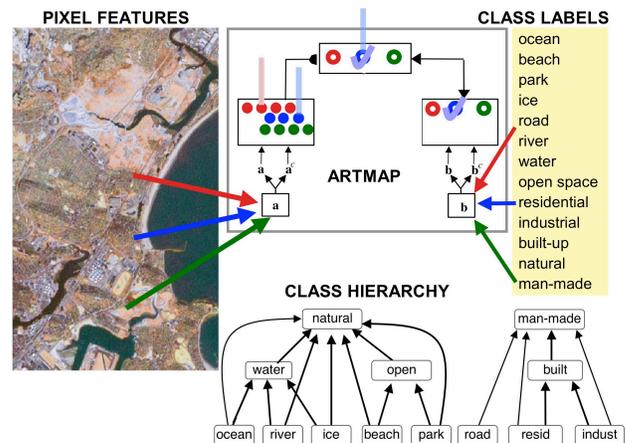


Figure 4. During training by an ARTMAP information fusion system, individual pixels from the Boston image testbed (left) learned associations with individual class labels (right). Underlying relationships among objects are assumed to be unknown to the automated system or the human user. The ARTMAP information fusion system uses distributed code representations that exploit the neural network's capacity for one-to-many learning in order to produce self-organizing expert systems that discover hierarchical knowledge structures.

The Boston testbed was derived from a Landsat 7 Thematic Mapper (TM) image acquired on the morning of January 1, 2001. The region includes portions of northeast Boston and suburbs. The resolution of the Boston image is $30m^2$ in six TM bands, $60m^2$ in two thermal bands, and $15m^2$ in one Panchromatic band. The image encompasses mixed urban, suburban, industrial, water, and park spaces. Landsat 7 spectral band values were acquired from the Earth Resources Observation System (EROS) Data Center, U.S. Geological Survey, Sioux Falls, SD (edc.usgs.gov). Dimensions: 180×300 pixels $\cong 5.4 \text{ km} \times 9 \text{ km}$.

Inputs for the Boston testbed example shown in Figure 4 were preprocessed by a version of the Lincoln Lab image mining system [15-17], called the *Neural Fusion Module*, which was developed by Waxman and colleagues working in the CNS Technology Laboratory during 2001-2002 [29,30]. For each pixel in the Boston image, this Module, implemented on an ERDAS Imagine (gis.leica-geosystems.com) platform, produced a 41-dimensional input vector representing local contrast, color, and texture attributes.

The ARTMAP fusion system provides a canonical procedure for labeling an arbitrary number of output classes in a supervised learning problem. A critical aspect of the embedded default ARTMAP network is the distributed nature of its internal code representation, which produces continuous-valued test set predictions distributed across output classes.

The information fusion techniques applied here modify the ARTMAP baseline mapping procedure by allowing the system to predict more than one output class during testing. A given pixel predicts the N classes receiving the largest net system outputs. Here the pixel-specific number N of predicted classes depends on the profile of distributed output predictions made by the feature vector for each given pixel. In this way, the information fusion method illustrated here improves upon the first version of the ARTMAP fusion method [31], which used a validation procedure to select a global number of predicted output classes (or a global activation threshold) for all test pixels. This original method works well when all pixels predict approximately the same number of classes in an underlying hierarchy, but performance would deteriorate if the correct number of ground truth output classes varied widely across test pixels.

Information implicit in the distributed predictions of a trained ARTMAP network generates the hierarchy of output class relationships. To accomplish this, each test pixel first produces its set of output class predictions. The resulting list of predictions across all test set pixels then determines a list of rules $x \Rightarrow y$, which define relationships between pairs of output classes, with each rule carrying a confidence value. The rules are used to assign classes to *levels*, with rule antecedents x at lower levels and consequents y at higher levels. Classes connected by arrows that codify the list of rules and confidence values form a graphical representation of the knowledge hierarchy.

The new information fusion methodologies are not limited to the image domain illustrated here, and could be applied, for example, to infer patterns of drug resistance from medical data or to improve marketing suggestions to individual consumers.

7. CLASSifier Simulation ManagER

CLASSER (CLASSifier Simulation ManagER) is a new modular set of software tools that provide a user with classifier implementations while handling details of data management and collection of test results. CLASSER provides a high-level system interface for learning applications, allowing the user to work with entire data sets at a time instead of individual points, and automating the collection of output results. The software facilitates neural algorithm implementations in both the user's application setting and in the Leica ERDAS Imagine environment.

Version 1 of CLASSER is now in beta testing, along with its first interface, CLASSER Script. The interface is batch-mode and script-driven, which allows the user to command a series of simulations from short scripts in a high-level language focused on data sets and classifier parameters, while automating the details of train/test protocols. The planned result of this software development effort will be a family of interconnected open-source tools for supporting research science, technology development, and applications of many varieties of cognitive and neural systems.

Acknowledgements

This work was supported by research grants from the Air Force Office of Scientific Research (AFOSR F49620-01-1-0423), the National Geospatial-Intelligence Agency (NMA 201-01-1-2016), the National Science Foundation (NSF SBE-0354378 and NSF BCS-0235398), and the Office of Naval Research (ONR N00014-01-1-0624); by postdoctoral fellowships from the National Geospatial-Intelligence Agency and the National Science Foundation for Siegfried Martens (NMA 501-03-1-2030 and NSF DGE-0221680); and by a Department of Homeland Security graduate fellowship for Ogi Ogas. Sucharita Gopal, Junchang Ju, and Mutlu Ozdogan provided remote sensing data, and Suhas Chelian and Brad Rhodes helped create the ground truth set for the Boston testbed.

References

- [1] S. Grossberg and E. Mingolla, "Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations", *Perception and Psychophysics*, 38(2), 1985, pp. 141-171.
- [2] G.A. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine", *Computer Vision, Graphics, and Image Processing*, 37, 1987, pp. 54-115.
- [3] G.A. Carpenter, S. Grossberg, and J.H. Reynolds, "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network", *Neural Networks*, 4(5), 1991, pp. 565-588.
- [4] G.A. Carpenter, "Distributed learning, recognition, and prediction by ART and ARTMAP neural networks", *Neural Networks*, 10(8), 1997, pp. 1473-1494.
- [5] G.A. Carpenter and S. Grossberg, "Normal and amnesic learning, recognition, and memory by a neural model of cortico-hippocampal interactions", *Trends in Neuroscience*, 16(4), 1993, pp. 131-137.
- [6] S. Grossberg, "How does a brain build a cognitive code?", *Psychological Review*, 87, 1980, pp. 1-51.
- [7] S. Grossberg, "The link between brain, learning, attention, and consciousness", *Consciousness and Cognition*, 8, 1999, pp. 1-44.
- [8] S. Grossberg, "How does the cerebral cortex work? Development, learning, attention, and 3-D vision by laminar circuits of visual cortex", *Behavioral and Cognitive Neuroscience Reviews*, 2(1), 2003, pp. 47-76.
- [9] M. Page, "Connectionist modelling in psychology: A localist manifesto", *Behavioral and Brain Sciences*, 23, 2000, pp. 443-512.
- [10] R.K. Aggarwal, Q.Y. Xuan, A.T. Johns, F. Li, and A. Bennett, "A novel approach to fault diagnosis in multicircuit transmission lines using fuzzy ARTMAP neural networks", *IEEE Trans. on Neural Networks*, 10(5), 1999, pp. 1214-1221.
- [11] S. Gopal, C. Woodcock, and A. Strahler, "Fuzzy ARTMAP classification of global land cover from the 1 degree AVHRR data set", *Remote Sensing of Environment*, 67, 1999, pp. 230-243.
- [12] N. Griffith and P.M. Todd, editors, *Musical Networks: Parallel Distributed Perception and Performance*, MIT Press, Cambridge, MA, 1999.
- [13] P. Lisboa, "Industrial use of safety-related artificial neural networks", *Contract Research Report 327/2001*, John Moores University, Liverpool, UK, 2001.
- [14] T.P. Caudell, S.D.G. Smith, R. Escobedo, and M. Anderson, "NIRS: Large scale ART 1 neural architectures for engineering design retrieval", *Neural Networks*, 7(9), 1994, pp. 1339-1350.
- [15] W.D. Ross, A.M. Waxman, W.W. Streilein, M. Aguilar, J. Verly, F. Liu, M.I. Braun, P. Harmon, and S. Rak, "Multi-sensor 3D image fusion and interactive search", *Proc. Third Int. Conf. Information Fusion*, Paris, vol. I, 2000.
- [16] W. Streilein, A. Waxman, W. Ross, F. Liu, M. Braun, D. Fay, P. Harmon, and C.H. Read, "Fused multi-sensor image mining for feature foundation data", *Proc. Third Int. Conf. Information Fusion*, Paris, vol. I, 2000.
- [17] A.M. Waxman, J.G. Verly, D.A. Fay, F. Liu, M.I. Braun, B. Pugliese, W.D. Ross, and W.W. Streilein, "A prototype system for 3D color fusion and mining of multisensor / spectral imagery", *Proc. Fourth Int. Conf. Information Fusion*, Montreal, vol. I, 2001, pp. 2-10.
- [18] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds, and D.B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps", *IEEE Transactions on Neural Networks*, 3(5), 1992, pp. 698-713.
- [19] T. Kasuba, "Simplified fuzzy ARTmap", *AI Expert*, 8(11), 1993, pp. 18-25.
- [20] G.A. Carpenter and W.D. Ross, "ART-EMAP: a neural network architecture for object recognition by evidence accumulation", *IEEE Transactions on Neural Networks*, 6(4), 1995, pp. 805-818.
- [21] G.A. Carpenter and N. Markuzon, "ARTMAP-IC and medical diagnosis: Instance counting and inconsistent cases", *Neural Networks*, 11(2), 1998, pp. 323-336.
- [22] J.R. Williamson, "Gaussian ARTMAP: A neural network for fast incremental learning of noisy multidimensional maps", *Neural Networks*, 9(5), 1998, pp. 881-897.
- [23] G.A. Carpenter, B.L. Milenova, and B.W. Noeske, "Distributed ARTMAP: a neural network for fast distributed supervised learning", *Neural Networks*, 11(5), 1998, pp. 793-813.
- [24] G.A. Carpenter, "Default ARTMAP", *Proc. Int. Joint Conf. on Neural Networks (IJCNN'03)*, Portland, Oregon, 2003, pp. 1396-1401.
- [25] O. Parsons and G.A. Carpenter, "ARTMAP neural networks for information fusion and data mining: map production and target recognition methodologies", *Neural Networks*, 16(7), 2003, pp. 1075-1089.
- [26] E. Mingolla, W.D. Ross, and S. Grossberg, "A neural network for enhancing boundaries and surfaces in synthetic aperture radar images", *Neural Networks*, 12, 1999, pp. 499-511.
- [27] G. Guy and G. Medioni, "Inferring global perceptual contours from local features", *International Journal of Computer Vision*, 20(1/2), 1996, pp. 113-133.
- [28] G. Simone, A. Farina, F.C. Morabito, S.B. Serpico, and L. Bruzzone, "Image fusion techniques for remote sensing applications", *Information Fusion*, 3, 2002, pp. 3-15.
- [29] A.M. Waxman, D.A. Fay, B.J. Rhodes, T.S. McKenna, R.T. Ivey, N.A. Bomberger, V.K. Bykoski, and G.A. Carpenter, "Information fusion for image analysis: geospatial foundations for higher-level fusion", *Proc. Fifth Int. Conf. Information Fusion*, Annapolis, 2002.
- [30] D.A. Fay, R.T. Ivey, N. Bomberger, and A.M. Waxman, "Image fusion and mining tools for a COTS environment", *Proc. Sixth Int. Conf. Information Fusion*, Cairns, Australia, 2003, pp. 606-613.
- [31] G.A. Carpenter, S. Martens, and O.J. Ogas, "Self-organizing hierarchical knowledge discovery by an ARTMAP image fusion system", *Proc. 7th Int. Conf. Information Fusion*, Stockholm, 2004, pp. 235-242.