

Information Fusion for Image Analysis: Geospatial Foundations for Higher-Level Fusion*

Allen M. Waxman, David A. Fay, Bradley J. Rhodes, Timothy S. McKenna,
Richard T. Ivey, Neil A. Bomberger, Val K. Bykoski, and Gail A. Carpenter

Department of Cognitive and Neural Systems

Boston University

Boston, MA, USA

waxman@cns.bu.edu

Abstract - In support of the AFOSR program in Information Fusion, the CNS Technology Laboratory at Boston University is developing and applying neural models of image and signal processing, pattern learning and recognition, associative learning dynamics, and 3D visualization, to the domain of Information Fusion for Image Analysis in a geospatial context. Our research is focused by a challenge problem involving the emergence of a crisis in an urban environment, brought on by a terrorist attack or other man-made or natural disaster. We aim to develop methods aiding preparation and monitoring of the battlespace, deriving context from multiple sources of imagery (high-resolution visible and low-resolution hyperspectral) and signals (GMTI from moving vehicles, and ELINT from emitters). This context will serve as a foundation, in conjunction with existing knowledge nets, for exploring neural methods in higher-level information fusion supporting situation assessment and creation of a common operating picture (COP).

Keywords: Image fusion, hyperspectral, GMTI, ELINT, neural networks, data mining, information fusion, common operating picture.

1 Introduction

In support of the AFOSR program in Information Fusion [1], the CNS Technology Laboratory of Boston University's Department of Cognitive and Neural Systems has developed a partnership in research and technology transition with the Information Directorate of the Air Force Research Laboratory (AFRL/IF Rome Research Site), national laboratories (MIT Lincoln Laboratory and Los Alamos National Laboratory), and industry affiliates (ERDAS, Kodak, RSI/Kodak, ESRI, and Spectral Sciences). We are developing and applying neural models of image, signals, pattern, and information processing, to the domain of *Information Fusion for Image Analysis*, organized in a geospatial context. Insights into brain mechanisms of multisensor fusion, relevant to multiple levels of information processing, have recently been summarized [2]; (also see refs. in [3]). Our approach builds on substantial prior work conducted with our colleagues at MIT Lincoln Laboratory, on neural methods with human-in-the-loop guidance for multisensor image fusion and search agent training, supporting both 3D site

visualization and foundation feature extraction. This work was presented at the *Fusion 2001* and *Fusion 2000* conferences [3-7].

Our current research and development is conducted in the context of a challenge problem, an urban crisis involving multiple hot-spots unfolding in a domestic city. We are concerned with *preparation and monitoring of the battlespace* from available multisensor imagery and signals data, as well as the ongoing interpretation of the situation and assessment of the threat as the crisis unfolds. Fused imagery with integrated live signals data, 3D site models and extracted foundation features, will support 3D visualization of a *common operating picture (COP)* for mission (re)planning and situational awareness during operations. We are focusing our attention on an area in the city of Mobile, Alabama, including urban infrastructure, a rail yard, and a port area on a river. Our data set includes hyperspectral imagery from the HyMap sensor in conjunction with high-resolution color visible imagery from the CitiPix sensor, collected over Mobile in Spring 2001 (made available to us by our affiliate, Kodak Commercial and Government Systems unit). In addition, GMTI and ELINT signal data streams will be simulated (by AFRL/IF Rome Research Site) for this area, covering time periods before, during, and after the crisis.

Image and signals fusion at *Levels 0 & 1* will be developed as an extension of our methods for image fusion and interactive mining of geospatial data, training *search agents* to find foundation features (e.g., roads, intersections and buildings) and other objects/events (e.g., traffic choke points). Fusion at *Level 2* will be developed in the form of object and event groupings, associations and trends across space and time. Information fusion at *Levels 2+* will be realized in the dynamics of semantic information networks, sculpted by associative learning laws in conjunction with existing knowledge sub-nets, and human-in-the-loop guidance. More generally, we believe our approach to higher-level information fusion can be combined with other complementary approaches involving genetic, statistical, model-based, and rule-based methods, in order to create an information fusion system that addresses the needs of the decision maker. Sensors with adaptive tasking capabilities, and a distributed secure communications infrastructure will enable delivery of the right information to the right user at the right time. This

* *Acknowledgement:* Supported by the Air Force Office of Scientific Research (AFOSR grant F49620-01-1-0423).

paper summarizes our approach, example data sets, and progress to date.

2 Fusion Architecture & Data Sets

Our information fusion architecture extends the conceptual framework for multisensor image fusion described in our earlier work [3-5], and will be described below in the context of specific data sets being assembled for the urban site of interest.

2.1 Data & information fusion architecture

Figure 1 illustrates our system architecture for integrating multisensor data in a geospatial context. It conveys a way to organize surveillance imagery and signals taken from airborne and space-based platforms, terrain data, and 3D site models, in order to fuse the data for interactive 3D visualization, and enable analysts to rapidly train neural network *search agents* that find foundation features (e.g., roads, rails, rivers, buildings, parking lots, forests, orchards, etc.) for mapping, and targets (e.g., vehicles) for localization and tracking. By establishing a geospatial foundation for the site, live imagery and signals can be registered to that site model, and activity monitoring and analysis can be performed using the trained search agents. In this fashion, object groupings and trends can be detected, classified and fed to higher-levels of information fusion.

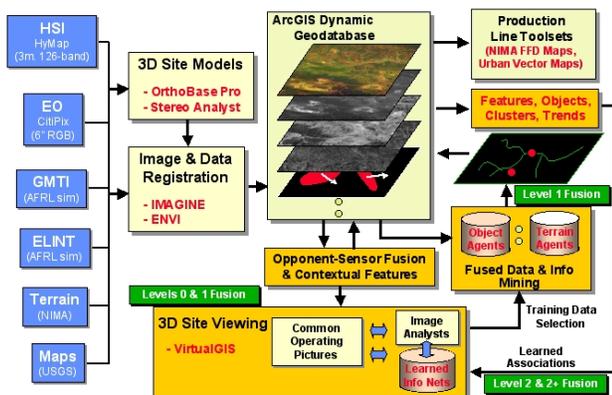


Figure 1. System architecture for multisensor and information fusion in a geospatial context.

2.2 Data sets and 3D models

We are concerned with the unfolding of a crisis brought on by a terrorist attack or other man-made or natural disaster in an urban environment. Therefore, it will be necessary to first establish a geospatial foundation for an urban data set. We have obtained multi-source imagery for such a data set from our affiliate, Kodak Commercial & Government Systems, for an area over Mobile, Alabama [21]. The area of interest covers approximately 1.5km x 6km as outlined in the map shown in Figure 2, and includes an urban grid, buildings, major and local roads, rails, a port facility on a river, and various kind of foliage.

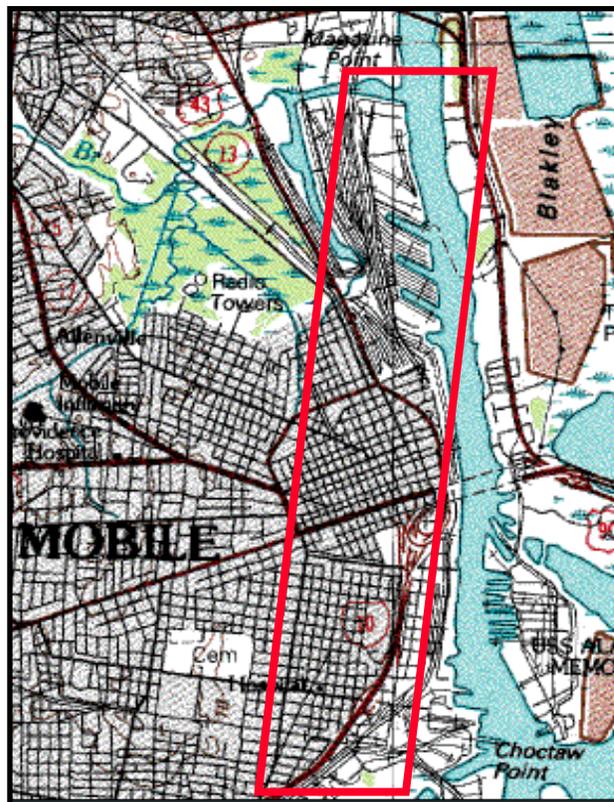


Figure 2. Map of urban area of interest in Mobile, AL.

Image data for this area consists of 6-inch resolution color visible *CitiPix* imagery (<http://kei.kodak.com>), and 3m resolution hyperspectral imagery from the *HyMap* sensor (www.aigllc.com). *HyMap* imagery consists of 126 bands in the 0.4 – 2.5 micron region of the spectrum, corrected for atmospheric and sensor effects using the ATREM and EFFORT methods [8]. An improved method of correcting for atmospheric absorption effects is FLAASH [9]. These image data sets were collected on an airborne platform in Spring 2001, at two different times, and so require georectification and registration. Example imagery and spectra are shown in Figures 3 & 4.

The *CitiPix* imagery is collected in overlapping frames, so we can use it as stereo imagery to construct a 3D site model (terrain and 20 buildings) for the area by first calibrating the sensor geometry. We have done this manually using the *OrthoBase Pro* and *Stereo Analyst* tools in *ERDAS Imagine* (www.erdas.com). For a large urban area of interest, it is essential to utilize automatic 3D site modeling tools like *RealSite* from Harris Corp. (www.govcomm.harris.com/realsite/index.html). Figure 5 illustrates two views of our limited Mobile site model (shown within *Imagine VirtualGIS*), one from the river looking into the port, and the other looking along a road towards the port. We have also indicated icons to suggest detections of moving vehicles from a GMTI sensor, and emissions as detected by an ELINT sensor. Such GMTI and ELINT data streams will be simulated for our crisis scenario by AFRL/IF Rome Research Site.



Figure 3a. CitiPix imagery of the port area in Mobile, AL



Figure 4a. HyMap imagery of the port area in Mobile, AL

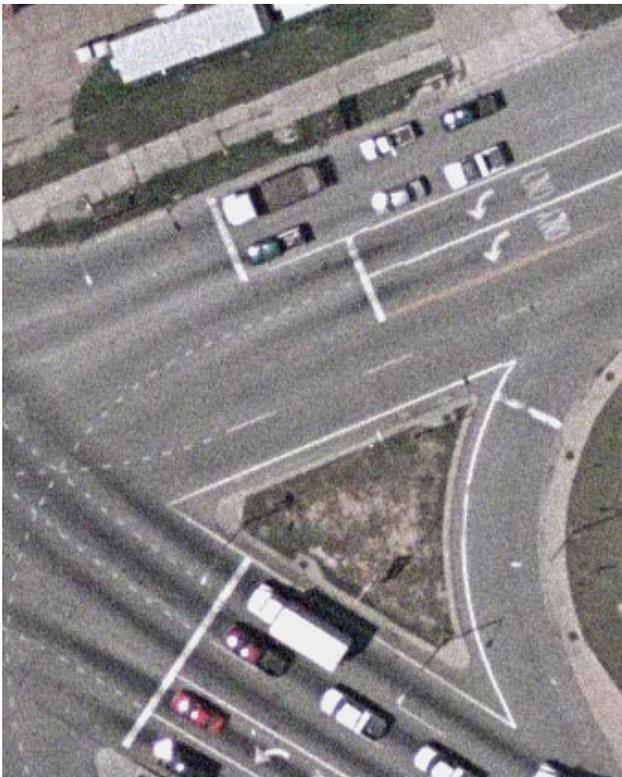


Figure 3b. Full resolution CitiPix imagery of Mobile, AL

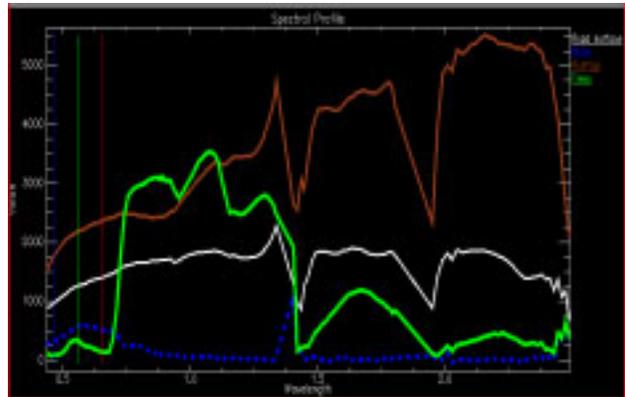


Figure 4b. 126-band spectra sampled from pixels in the above HyMap image (brown=rooftop, gray=road, green=tree, blue=water). HyMap imagery and spectra are easily viewed using *ENVI* (www.rsinc.com).

A 3D site model such as this, created from multi-sensor archived imagery, is useful for extracting roads, buildings, and other foundation features. It also provides a substrate against which to register live imagery collections using an automatic registration algorithm such as *HART* (High Accuracy Registration Technique) from BAE Systems [10]. It also provides context for interactive 3D viewing (using *Imagine VirtualGIS*) of live data streams, tracks, and statistics such as those derived from GMTI and ELINT (signals) sensors. We will illustrate this in a future presentation and paper.

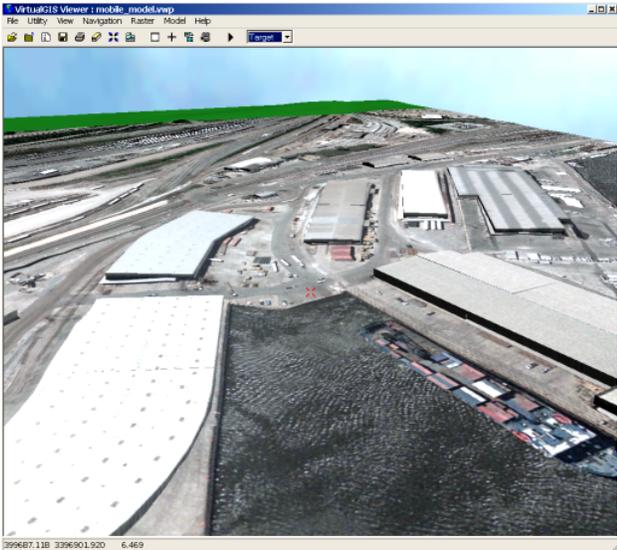


Figure 5a. Local 3D site model of the port facility in Mobile, AL, constructed from overlapping CitiPix data. It is constructed and viewed with tools in ERDAS *Imagine*.



Figure 5b. Local 3D site model of Mobile, AL, viewed from a road towards the port facility. Icons superimposed are suggestive of GMTI detections (red ellipses) and ELINT detections (yellow stars).

3 Fusion & Mining for Roads

In the context of an urban crisis, intelligence preparation of the battlespace would consist of providing up-to-date electronic maps indicating the locations (and relevant info) of foundation features like roads, rails, buildings, rivers, ports, foliage, and the full suite of urban infrastructure. In addition, a 3D site model would be needed to support multisensor data registration (using standard photogrammetry techniques), as well as interactive 3D visualization for operations planning and situational awareness. This information would easily be represented

within a geospatial information system like *ArcGIS* from ESRI (www.esri.com). It will also be necessary to update the status of foundation features and infrastructure from sensor data, as the crisis unfolds. Thus, tools which are able to learn to extract such features as necessary, directly from fused multisensor data, can prove valuable.

Our prior work on multisensor image fusion and mining [3-5] can be used to train search agents which learn to find foundation features in hyperspectral imagery from the HyMap sensor. We utilize both spectral features and multi-resolution spatial features in the imagery, and have the learning system discover which image features are sufficient to conduct the search. In future work we will utilize both the HyMap spectral imagery and the CitiPix high-resolution visible imagery for this purpose. Here we report results for road extraction using only HyMap data.

Figure 6 illustrates the recovered reflectance spectra for a variety of road pixels. Removing the strong atmospheric absorption bands leaves 100 bands of useful data from 0.4 – 2.5 μ . Due to varying terrain orientation on a small scale, as well as partial blockage of road pixels (by dirt, tar, etc.), we can see that the spectral shapes (Fig. 6-upper) are quite similar, though they are scaled versions of one another. Using the normalized spectra instead, their similarity becomes more apparent (Fig. 6-lower).

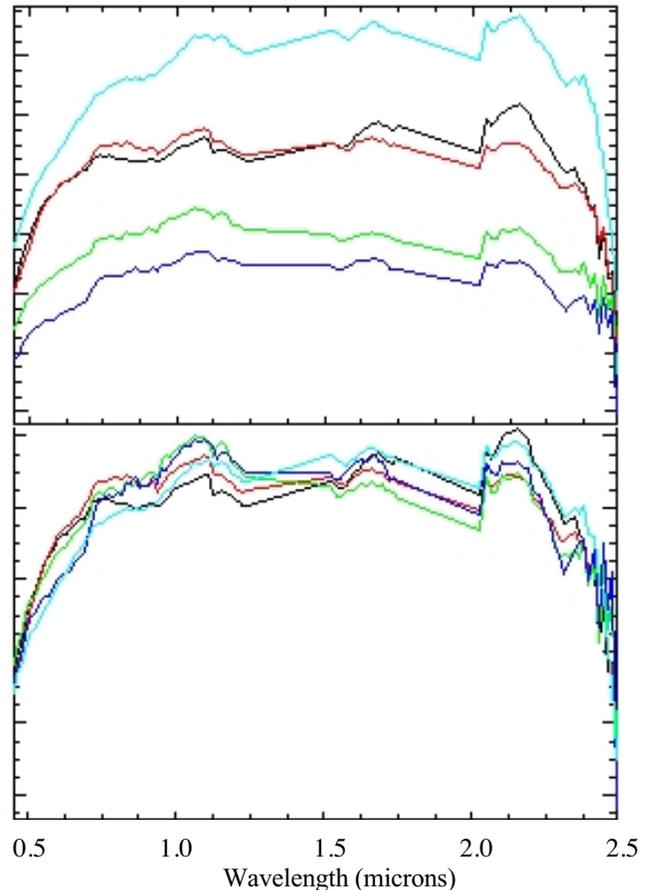


Figure 6. Sample road spectra from HyMap imagery
Upper: Unnormalized; Lower: Normalized spectra

Mining the hyperspectral HyMap imagery for *roads* is accomplished by training search agents to recognize patterns [11] across prepared layers of processed data, as explained in [2-5]. Hyperspectral processing here consists only of normalizing the 100 remaining reflectance values at each pixel (following the atmospheric correction and deletion of strong absorption bands). Neural models of visual (retinal and cortical) processing are applied on three scales to the HyMap imagery after it is first aggregated into four broad bands: visible ($0.4 - 0.7\mu$), near-IR ($0.7 - 1.0\mu$), SWIR-1 ($1.0 - 1.75\mu$), and SWIR-2 ($2.0 - 2.5\mu$). These four broad-bands are used to create two opponent-color channels (visible vs. near-IR, and SWIR-1 vs. SWIR-2), each producing single and double-opponent layers based on center-surround shunting dynamics [12, 6]. Each broad-band image is then processed using twelve oriented Gabor filters that compete across orientation and local image space, but cooperate across scales in a manner that approximates the Boundary Contour System [13, 14]. These oriented filter responses are also used to create various measures of image texture, and they are then combined with the opponent-color and contour imagery on three scales into a stack of 108 layers. Finally, the 100 normalized spectral layers can be combined with the 108 spatial layers to create a spatio-spectral stack, where each pixel corresponds to a feature vector of 208 elements.

In each case, we utilize the prototype system for image mining we developed at Lincoln Laboratory [3-5], where a user can select (with the mouse) examples and counter-examples of *target* pixels, and rapidly train a search agent consisting of 5 ARTMAP neural networks [15, 16], which then searches the entire image (stack) for other *target* pixels. Voting among the 5 ARTMAP networks is used to establish a confidence level for pattern recognition. The results can then be displayed as *detections* and post-processed into vectors for mapping purposes. In addition, we had developed a method to reduce the large feature space into a small *sub-set* of features *sufficient* to correctly classify all the selected training data, and thereby accelerate the search process.

Figure 7 shows the results of mining the 100-layer stack of normalized spectral data for *roads*, displaying results for 53% and above confidence. It also shows the user interface where examples and counter-example pixels are selected. The ARTMAP networks learn a maximum of only 7 categories (2 *road* and 5 *non-road*), and the feature space is reduced to a maximum of 11 spectral layers with good consistency across the 5 ARTMAP nets. Figure 8 illustrates *road* detections obtained by mining the 108-layer spatial feature stack, displaying results for 75% and above confidence, having learned a maximum of only 4 categories and selected only 7 feature layers. The best results are obtained by mining the 208-layer spatio-spectral stack, as shown in Figure 9. In this case a maximum of 6 categories are learned (4 *road* and 2 *non-road*), and a reduced sub-set of only 16 features are selected, about half of which are spectral and the other half being spatial across all three scales. Results are shown (in white) for the 100% confidence level only.

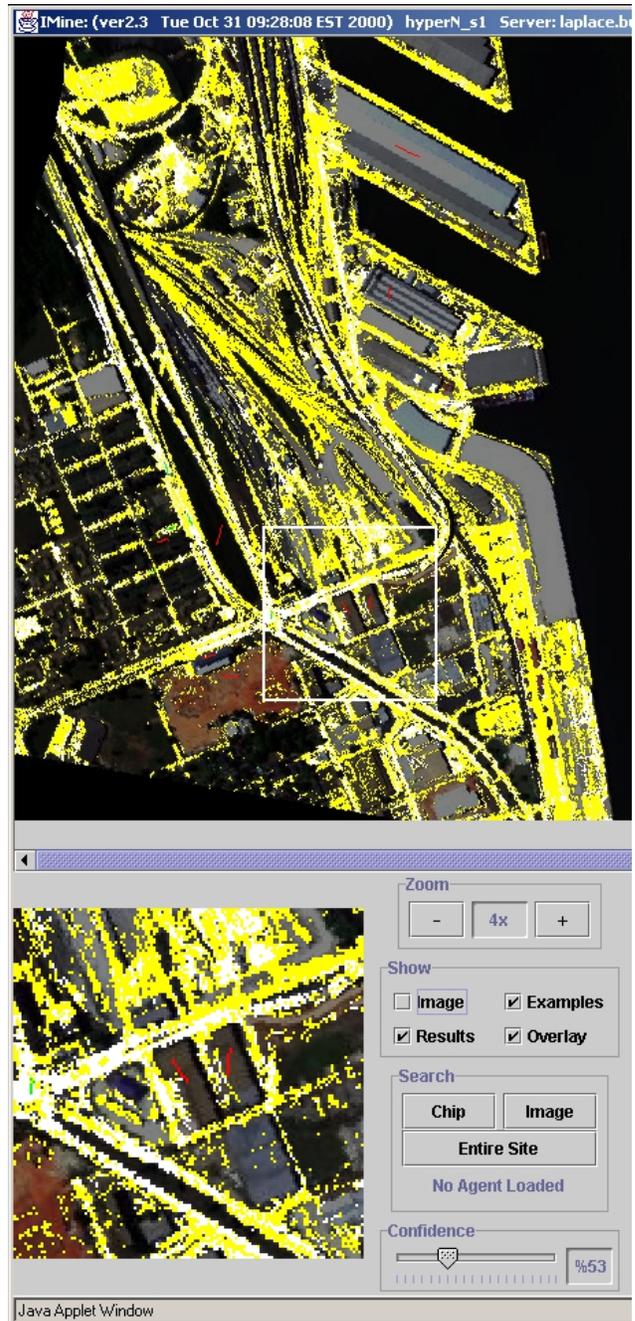


Figure 7. Mining for *roads* in HyMap normalized spectral imagery (100 bands). Results displayed at 53% confidence and above are in yellow, with 100% confidence in white.

Future work will make use of registered high-resolution CitiPix imagery for the multi-resolution spatial contours and texture signature, in conjunction with contrast enhanced spectral signatures from HyMap. A key objective is to incorporate the non-imaging modalities of GMTI and ELINT detections, tracks and statistics, as additional layers of registered data, enabling for example, a search for roads carrying unusual traffic patterns at night. We will soon address this challenge by utilizing simulated GMTI and ELINT data streams in conjunction with the multi-sensor IMINT.

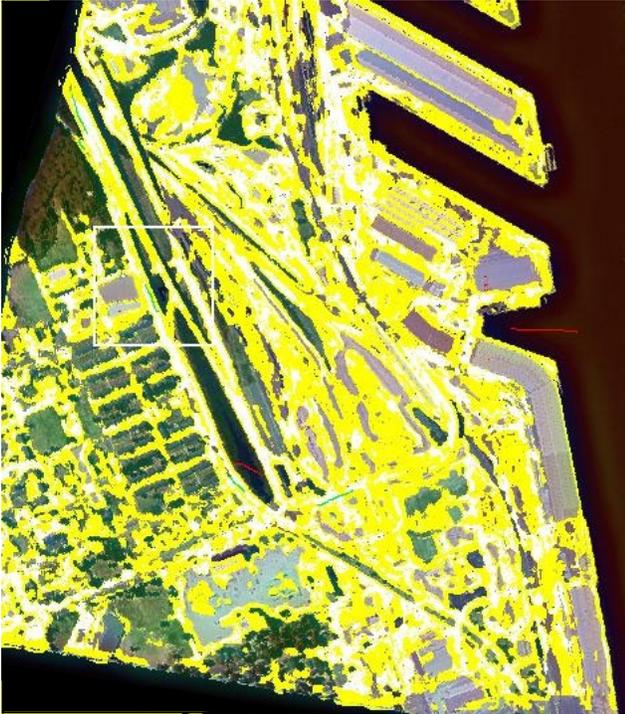


Figure 8. Mining for roads in HyMap imagery processed as 4-band multi-resolution (108-layer) data. Search results are displayed at 75% confidence and above.

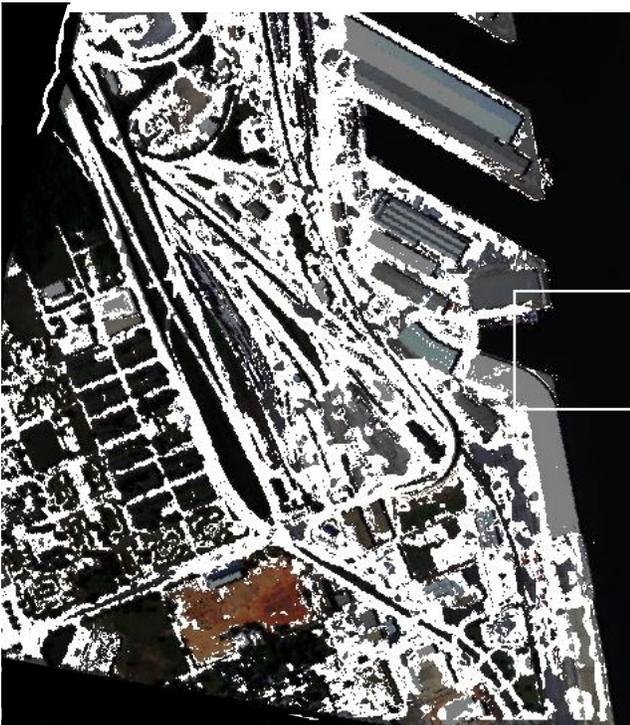


Figure 9. Mining for roads in HyMap imagery of Mobile, processed as multi-resolution spatio-spectral data (208 layers). Detection results are displayed at 100% confidence in white. A reduced set of 16 features is found sufficient, and only 6 categories are learned (4 road and 2 non-road).

4 Fusion & Mining in *Imagine*⁺

In order to enable widespread dissemination and use of our neural methods for multisensor/spectral image fusion and mining for foundation features and targets, we have refined and enhanced our prototype system [3], and reimplemented it within an extensible commercial image exploitation environment, ERDAS *Imagine*. This allows us to take advantage of the significant software infrastructure and capabilities of the *Imagine* suite, while supporting technology transfer to the National Imagery and Mapping Agency, the Air Force Research Laboratory AFRL/IF, and our other affiliates.

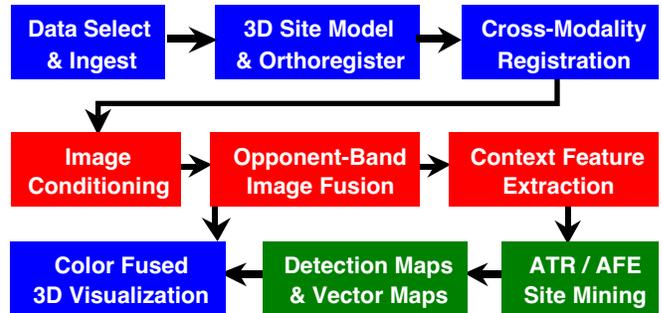


Figure 10. Workflow for multisensor/spectral image fusion (red boxes) and mining (green boxes), integrated into the ERDAS *Imagine* environment (blue boxes).

The process workflow for image fusion and mining is shown in Figure 10, where the red and green boxes correspond to our own modules, and the blue boxes are modules developed from existing *Imagine* functionality. Figure 11 illustrates the opponent-color image fusion user interface, with an area in Monterey, CA, imaged in four spectral bands (red, green, blue, near-IR) and a panchromatic band, being fused into a color presentation.



Figure 11. User interface for opponent-color image fusion of four bands/sensors, integrated into ERDAS *Imagine*.

⁺ This component of our work has been supported by the National Imagery and Mapping Agency (NIMA/NTA).

Figures 12 and 13 illustrate the image mining interface, whereby a search agent (5 ARTMAP networks) is easily trained to find targets pointed out by example and counter-example. Also, a subset of *sufficient feature layers* is discovered that is able to support 100% correct categorization of the selected training examples and counter-examples, and the user is then informed of these important data layers. As the user points out mistakes or missed targets in subsequent search areas, the mining system learns to improve its performance and refines the trained search agent.

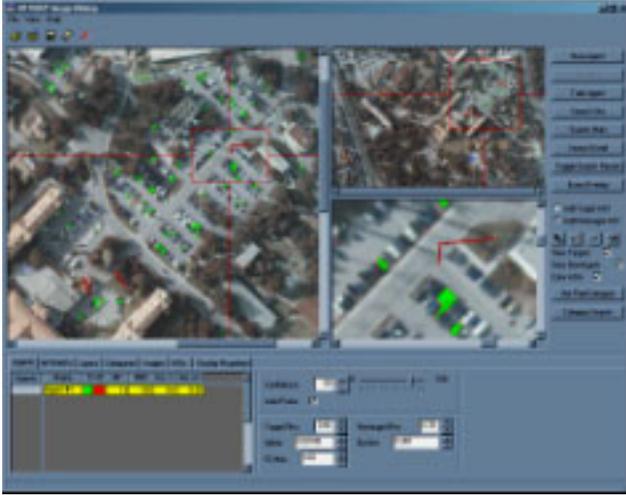


Figure 12. User interface for ARTMAP Image Mining integrated into ERDAS *Imagine*, shown with an agent trained to find *red cars* (detections marked in green).

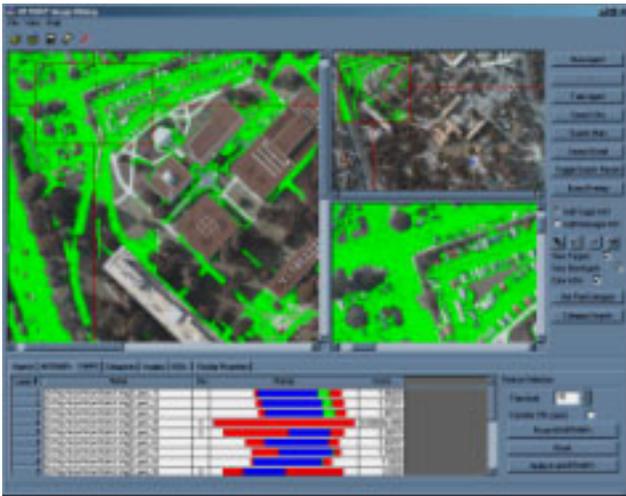


Figure 13. Training a search agent to find *roads*, and discovering a *sufficient subset* of feature layers.

5 Higher Levels of Fusion

The methods described above, indicative of *Levels 0 & 1* fusion, will lead to the rapid mapping and updating of urban infrastructure from airborne and space-based spectral imaging platforms. The results will provide context for extending this analysis to live signals data, corresponding

to moving vehicles and radio emitters. *Level 2* fusion will also be approached from a neural modeling standpoint, in which associative groupings will be established among objects across space and time (e.g., vehicle groups with certain emission characteristics in proximity to certain buildings during select time intervals). Such groupings and trends will be determined from the data streams, both to establish normalcy models and to detect anomalies that may become associated with other events. Similarly, such groupings, trends and events can become associated with established knowledge structures, also in the form of networks.

Recalling that trained *search agents* transform data, initially in the form of feature layers, into categories and class hierarchies, provides a means to convert *signals* to *symbols*. These category and class *nodes* (i.e., *symbols*) may then develop associations (both excitatory and inhibitory) with one another and with pre-constructed knowledge nets, by means of modified Hebbian learning [11, 12, 17]. Groupings of these nodes will form *concepts*, and higher-order associations will establish a network of *semantic* knowledge. Again, we will enable human-guided input (activating *concept nodes*) at these higher-levels of information fusion.

The information networks we envision here are more than the traditional notion of a *semantic network*, as they will be dynamic. Neural models of coupled networks of excitatory and inhibitory nodes, activated by input data, are known to exhibit oscillations and resonant states [12, 17, 18]. The inclusion of spiking behavior in these neural models can then lead to the emergence of synchronous spiking activity among sub-nets within the larger network [19, 20]. It is tempting to interpret these synchronies as states of transient binding among data-driven categories, concepts and knowledge. This would then represent a hypothesis or assessment of the situation, a *Level 2+* fusion of information.

6 Conclusions

The architecture shown in Figure 1, and the methods described in Section 3, have been applied to the extraction of *roads* from HyMap hyperspectral imagery. We will extend this in order to establish an urban mapping capability from fused HyMap and CitiPix imagery, and merge these foundation features with the interactive 3D visualization capabilities illustrated in Figure 5. We will soon augment our imagery data set with simulated GMTI and ELINT streams that span the time period before, during and after an urban crisis unfolds. This dynamic data will augment our 3D visualization, as well as create additional feature layers that will support mining for movers, emitters, tracks, and utilized infrastructure.

We are quite confident about our neural systems approach to generating the geospatial foundations for higher-level information fusion, and its potential to support fused IMINT, GMTI and ELINT. The challenge is at *Levels 2 & 2+*, whereby neural models of spatio-

temporal grouping will be applied to event detection (expected and anomalous events), and trend learning and prediction. Still more challenging is the establishment of a dynamic network of semantic information, combining data-driven category learning with knowledge networks via learned associations. We expect that activation of these networks, by input data streams and human-guided priming, will establish synchronous dynamics, oscillations and spiking, among sub-nets that represent multiple hypotheses about the urban situation and assessment of the threat.

References

- [1] J.F. Tangney, *AFOSR programs in higher levels of information fusion*, in 5th International Conference on Information Fusion, Annapolis, 2002.
- [2] A.M. Waxman, *Sensor fusion: Biology to technology*, in **The Handbook of Brain Theory and Neural Networks**, 2nd edition, (M.A. Arbib, Ed.), Cambridge, MA: MIT Press, 2002 (in press).
- [3] A.M. Waxman, J.G. Verly, D.A. Fay, F. Liu, M.I. Braun, B. Pugliese, W. Ross, and W. Streilein, *A prototype system for 3D color fusion and mining of multisensor/spectral imagery*, in 4th International Conference on Information Fusion, Montreal, 2001.
- [4] W. Ross, A. Waxman, W. Streilein, J. Verly, F. Liu, M. Braun, P. Harmon, and S. Rak, *Multisensor image fusion for 3D site visualization and search*, in 3rd International Conference on Information Fusion, Paris, July, 2000.
- [5] W. Streilein, A. Waxman, W. Ross, F. Liu, M. Braun, J. Verly, and C.H. Read, *Fused multisensor image mining for feature foundation data*, in 3rd International Conference on Information Fusion, Paris, July, 2000.
- [6] D.A. Fay, A.M. Waxman, M. Aguilar, D.B. Ireland, J.P. Racamato, W.D. Ross, W.W. Streilein, and M.I. Braun, *Fusion of multisensor imagery for night vision: Color visualization, target learning and search*, in 3rd International Conference on Information Fusion, Paris, July, 2000.
- [7] D.A. Fay, A.M. Waxman, J.G. Verly, M.I. Braun, J.P. Racamato, and C. Frost, *Fusion of visible, infrared and 3D LADAR imagery*, in 4th International Conference on Information Fusion, Montreal, August, 2001.
- [8] F.A. Kruse, J.W. Boardman, A.B. Lefkoff, J.M. Young, K.S. Kierein-Young, T.D. Cocks, R. Jenssen, and P.A. Cocks, *The 1999 AIG/HyVista HyMap group shoot: Commercial hyperspectral sensing is here*, in Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI, SPIE-4049, 210-217, 2000.
- [9] A. Berk et al., *Exploiting MODTRAN radiation transport for atmospheric correction: The FLAASH algorithm*, in 5th International Conference on Information Fusion, Annapolis, 2002.
- [10] D. Lewis, *Multisensor georegistration using HART (High Accuracy Registration Technique)*, in 5th International Conference on Information Fusion, Annapolis, 2002.
- [11] G.A. Carpenter, *Neural network models for pattern recognition and associative memory*, Neural Networks, 2, 243-257, 1989.
- [12] S. Grossberg, *Nonlinear neural networks: Principles, mechanisms and architectures*, Neural Networks, 1, 17-61, 1988.
- [13] S. Grossberg and E. Mingolla, *Neural dynamics of perceptual grouping: Textures, boundaries and emergent segmentations*, Perception & Psychophysics, 38, 141-171, 1985.
- [14] A.M. Waxman, M.C. Seibert, A. Gove, D.A. Fay, A.M. Bernardon, C. Lazott, W.R. Steele, and R.K. Cunningham, *Neural processing of targets in visible, multispectral IR and SAR imagery*, Neural Networks, 8, 1029-1051, 1995.
- [15] G.A. Carpenter, S. Grossberg, and D.B. Rosen, *Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system*, Neural Networks, 4, 759-771, 1991.
- [16] 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, pp. 698-713, 1992.
- [17] D.S. Levine, **Introduction to Neural and Cognitive Modeling**, 2nd edition, Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
- [18] D.S. Levine, V.R. Brown, and V.T. Shirey, (editors), **Oscillations in Neural Systems**, Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
- [19] W. Maass and C.M. Bishop (eds.), **Pulsed Neural Networks**, Cambridge, MA: MIT Press, 1999.
- [20] S. Grossberg, W. Maass, and H. Markram, (editors), **Spiking Neurons in Neuroscience and Technology**, special issue of Neural Networks, vol. 14, 2001.
- [21] A.D. Mirzaoff, S.G. Vogler, and J. Coss, *Kodak Multi-Sensor IMINT Data Collection*, in 5th International Conference on Information Fusion, Annapolis, 2002. in 5th International Conference on Information Fusion, Annapolis, 2002.