A Prototype System for 3D Color Fusion and Mining of Multisensor / Spectral Imagery*

Allen Waxman, Jacques Verly, David Fay, Fang Liu, Michael Braun, Benjamin Pugliese, William Ross, and William Streilein

Sensor Exploitation Group
MIT Lincoln Laboratory
Lexington, MA, USA
waxman@ll.mit.edu

Abstract - We have developed a prototype system in which a user can fuse up to 4 modalities (or 4 spectral bands) of imagery previously registered to one another with respect to a 3D terrain model. The color fused imagery can be draped onto the terrain to support interactive 3D fly-through. The fused imagery, and its opponent-sensor contrasts, can be further processed to yield extended boundary contours and texture measures. Together, these layers of registered imagery and image features can be interactively mined for objects of interest. Data mining for infrastructure and compact targets is achieved using a point-and-click user interface in conjunction with a Fuzzy ARTMAP neural network for on-line pattern learning and recognition. Graphical user interfaces enable the user to control each stage of processing: image enhancement, image fusion, contour and texture extraction, 3D terrain characterization, 3D graphics model building, preparation for exploitation, and interactive data mining. The system is configured as a client-server architecture, enabling remote collaborative exploitation of multisensor imagery. Throughout, the processing of imagery and patterns relies on neural network models of spatial and color opponency, and the adaptive resonance theory of pattern processing. This system has been used to process imagery of a variety of geographic sites, in order to extract roads, rivers, forests and orchards, and performance has been assessed against manually determined “ground truth.” The data mining approach has been extended for the case of hyperspectral imagery of hundreds of bands. This prototype system has now been installed at multiple US government sites for evaluation by image analysts. We plan to extend this approach to include various non-imaging sensor modalities that can be localized to geographic coordinates (e.g., GMTI and SIGINT). We also plan to embed these image fusion and mining capabilities in commercial open software environments for image processing and GIS.

Keywords: Multisensor fusion, image fusion, data fusion, data mining, neural networks, pattern recognition, adaptive resonance theory

1 Introduction

In light of the rapid increase in number and type of imaging satellites becoming available to support both government and commercial interests throughout the world, we are motivated to develop paradigms and systems that can take advantage of these complementary and voluminous data sets. Space-based imaging of the earth’s surface is being conducted on a routine basis across the spectrum, employing panchromatic visible (EO), multispectral (MSI) visible through infrared (IR), hyperspectral (HSI) and polarimetric synthetic aperture radar (SAR) imaging, all of varying spatial resolution. It is clear that assembling and fusing multispectral and multi-modality data from multiple platforms will greatly improve our ability to detect, classify and map objects, land cover, surface features, and changes over time. There is an increasing need for systems that can aid image analysts to exploit this rich multi-modality data, in order to:

- Fuse it to enable high fidelity 2D and 3D visualization (particularly of non-literal modalities like thermal IR and SAR imagery);
- Combine spectral imagery with spatial image features (e.g., color boundary contours, textures and 3D features) into unified spatio-spectral signatures of objects and terrain;
- Support interactive image mining to discover information bearing layers of data particular to objects/features of interest;
- Learn the spatio-spectral signatures of objects (i.e., create Search Agents) and automate the search for them over vast geographic data sets;
- Use trained Search Agents to access fused imagery from databases based on object/feature content, and to assess changes in objects and terrain over time;
- Convert this exploited multisensor imagery into vector maps and hybrid raster-vector 2D/3D products
- Enable collaboration among remote analysts and rapid dissemination of customized products to users via web-based communications.

At the Fusion 2000 conference, we reported progress towards developing methods that support many of the

---

* This work was sponsored in part by the Air Force Office of Scientific Research, under Air Force Contract F19628-00-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and not necessarily endorsed by the U.S. Government.
above technical goals [1,2]. These methods for image fusion and mining that we typically apply off-line to archived imagery, are also applicable to real-time imaging in 2D using up to four passive sensors [3] as well as active 3D LADAR imaging [4]. Their early development has been reported in the context of color night vision [5,6], EO/IR/SAR fusion [7] and multispectral infrared fusion [8].

Our image fusion methods are motivated by biological examples of visible/IR fusion in snakes [9,10], and color processing in the human visual system [11-14]. We model these processes via spatial- and spectral-opponent neural networks embodied in center-surround receptive fields [15], oriented multiscale processing of textures and boudaries [16], and adaptive resonance models of learning and recognition [17,18].

2 Paradigms and a Prototype System

As introduced in [1,2], we adopt an image fusion paradigm based on a layered data structure comprised of multispectral and multi-modality imagery registered with respect to 3D site models. We employ digital terrain data to orthorectify and register imagery of terrain. An extension of these ideas applies separately to the rooftops and sides of buildings, when modeled. The 3D site model can be constructed directly from imagery using a variety of means (e.g., photogrammetry, radar mapping, ladar measurements).

![Figure 1: Multisensor 3D fusion paradigm](image)

In Fig. 1 we illustrate our paradigm of multisensor fusion and data mining. Multiple imaging (and possibly non-imaging) modalities are registered by means of a 3D site model, forming a layered georeferenced database. (This data structure will be expanded to include processed imagery below.) The imagery is then fused and hosted in a client-server web-based environment for interactive 3D visualization and data mining by multiple collaborating analysts.

We have developed a prototype system that supports each of these stages of multisensor fusion and exploitation. It can be organized as a processing chain, as shown in Fig. 2. The processing chain consists of a set of software modules, some of which employ commercial tools (shown in blue: for data ingest, orthorectification, cross-modality image registration, vector product generation, and 3D viewing), and others that we have developed (shown in red) specifically for image conditioning, opponent-band image fusion, spatio-spectral context feature extraction, interactive data mining, and mined feature display. The approach is extensible, as additional imaging sensors can provide source data for fusion, and additional image features can be added to provide additional context for subsequent data mining. A variety of change detection methods (indicated in green) can be incorporated. Each additional data layer (source, feature, or change) must be registered using the 3D site model.

![Figure 2: Processing chain for image fusion & mining](image)
The resulting fused imagery and source modalities can be combined with the 3D terrain in the Exploitation Preparation module of the control panel in Fig. 3, for interactive stereo visualization using a viewer built on top of SGI Inventor or TGS Open Inventor viewer (in an Irix or Windows NT operating system, respectively). An example is shown in Fig. 7 for Tuzla, Bosnia, based on 30m resolution DTED, and textured with the fused imagery. The user can switch among the modalities to change textures.
Textures module from the control panel in Fig. 3. The GUI shown in Fig. 8 allows the user to select any conditioned band or opponent-band contrast imagery created, and select among a variety of image texture processing routines based on oriented multi-scale Gabor filtering, local variance and auto-correlation, and boundary contour extraction based on BCS theory [16]. This creates additional layers of spatio-spectral data that is registered to the processed sensor imagery. This module has been augmented to enable terrain geometry processing for ground slopes and curvatures.

The entire set of conditioned source imagery, opponent-band imagery, spatio-spectral textures and contours, and terrain features, is then assembled into a layered data stack (or deepfile) in the Exploitation Preparation module launched from the Session Manager. This layered data stack can be augmented with other kinds of registered data (not necessarily imagery), and is ideal for fused image mining.

3 Fused Image Mining

The layered data stack is a useful data structure, and provides a mechanism to augment the data mining approach with other source imagery, non-imaging geospatial data like GMTI and associated tracks, SIGINT localizations and tracks, contextual features extracted by other algorithms, image change layers, spatial and topological analyses, etc. Such a layered stack is illustrated in Fig. 9. As indicated there, a vertical line through the stack selects a set of registered data and features corresponding to a pixel or geo-spatial location (the features capture information about the neighborhood of the location on the ground). This data forms a feature vector suitable for classification and mining.

Fused image mining is accomplished through the use of an interactive data selection GUI (shown in Fig. 10), and a Fuzzy ARTMAP neural network for pattern learning (to create search agents) and search [17,18] (shown in Fig. 11).
The user selects (with the mouse) in the lower left zoom box of the GUI, examples and counter-examples of target pixels, as shown in Fig. 10. Any layer in the data stack can be viewed in the upper panel, yet selection of a pixel actually selects the entire corresponding feature vector from the data stack. All feature vectors are labeled as either target or non-target class, and are fed to the Fuzzy ARTMAP net.

![Figure 11: Fuzzy ARTMAP neural network for learning search agents that discriminate targets from non-targets](image1)

The ARTMAP architecture shown in Fig.11 is simplified and consists only of a lower ART module, which performs unsupervised clustering of feature vectors into categories, and an upper layer in which the learned categories form associations with one or the other class for a target of interest. This approach enables the network to learn a compressed representation of the target class in the context of non-targets in the scene. That is, it learns to discriminate the target from the context, using the multisensor data and spatio-spectral features. A target of interest is typically represented by a few learned categories, as are the non-targets. For a target of interest, the user goes through the process of example/counter-example selection, local search in a selected area, display of search results, and possible refinement by selecting additional pixels as examples of missed detections (targets) or false alarms (non-targets) and then relearning the representation. This interactive cycle typically takes seconds. Again, the user interacts with an applet running on a (possibly) light client computer, while the computations run on a remote server computer. This kind of interactive training of search agents is enabled by the on-line learning capabilities of ART architectures, and the simplicity of the Fuzzy ART computations. In addition, we have developed a fast procedure for isolating the subset of features (or layers) necessary to actually discriminate the target from the non-target [2] in the selected training data, which then accelerates the search over the image session or entire site. For each target class we actually train five Fuzzy ARTMAP networks with randomized order of selected training data, and derive a confidence value for target detections from the voting consensus across all five nets.

Figs. 12 & 13 illustrate search results for roads and buildings in the fused example shown in Fig. 7.

![Figure 12: Search for roads in fused EO, IR, SAR](image2)

![Figure 13: Search for buildings in fused EO, IR, SAR](image3)

This same approach to fused image mining can be applied to hyperspectral imagery (HSI) of hundreds of bands, though our current prototype system is not designed to accommodate so many bands. Still, the same methods can be applied directly, to derive opponent-band contrast, extract contours and textures from select (opponent) bands, and create a layered data stack for interactive mining. For purposes of display and target pixel selection, we can aggregate bands into four broad spectral bands and create a fused visualization. This was demonstrated on AVIRIS imagery in [2], and here we illustrate results for HYDICE.
imagery of an area of Ft. Hood, Texas, consisting of 210 bands in the spectral range 0.4-2.5 microns (visible through short-wave infrared). Figs. 14-16 show search results for rivers, roads/trails, and forested areas, respectively.

Figure 14: Search for rivers in HYDICE fused HSI

Figure 15: Search for trails in HYDICE fused HSI

Figure 16: Search for forested areas in HYDICE HSI (detected areas are delineated by their boundaries)

In these examples of mining fused HSI, though the deep-file of feature vectors exceeds two-hundred layers, the same Fuzzy ARTMAP architecture can easily process it, and discover that only four layers are necessary to find all three classes of targets searched for. The salient layers involve opponent-band contrasts and contours. Searches which discriminate forests from orchards discover that texture is (not surprisingly) a key feature, since the spectral content is so similar. Reducing a hyperspectral image search down to only a few layers accelerates the process by two orders of magnitude. This approach is quite different from the more conventional principal components decomposition.

Figure 17: Performance summary (percent detection vs. false alarms) for fused image mining on six sites (different symbols) for 4 classes of target features
In Fig. 17 we summarize the performance of fused image mining for four classes of targets (roads, rivers, forests, orchards) extracted from six sites at four different geographic locations using different mixtures of sensor imagery. We determine the percent detection and percent false alarm by comparison with manually delineated target features. We see here the potential to achieve in excess of 90% detection with less than 5% false alarms in single-pass search for such cultural features. This suggests that multisensor fused image mining may provide a tool to aid geospatial image analysts. Converting these feature detections from raster to vector format, enables vector editing and map generation using existing COTS software (as indicated in Fig. 2).

4 Collaborative Map-Image Interface

We’ve noted that our prototype fusion and mining system is structured as a client-server architecture that supports web-based communications. This provides a simple means to support collaborative exploitation by remotely located analysts, and a mechanism for rapid dissemination of results to remote users. In fact, users can easily select among the search results relevant to them, and possibly conduct further analysis such as spatial relations among detected targets. To enable such collaboration, we’ve prototyped a map-image interface which is displayed on the client and supports selective display of search results, annotation, and read-out of latitude & longitude for the cursor location. In fact, it is this same GUI which is used to select the area of interest for image mining, and to launch the mining GUI of Fig. 10.

An example of client-server based exploitation is shown in Fig. 18, for four-band multispectral imagery (MSI) collected with the Positive Systems Inc., airborne ADAR sensor over the Naval Postgraduate School, Monterey, CA, to support the Urban Warrior Experiment of March 1999.

Results of mining for two classes of targets (red cars and buildings) within a selected area by two analysts are each saved to the server, then down-loaded into a single map-image interface, and displayed as an applet within Netscape by a remote user. This is illustrated (with an enlarged area) in Figs. 19 & 20, where the user can toggle between an underlying fused image and an (old) map.
5 Summary and Future Directions

We have developed a prototype integrated system for multisensor image fusion and mining, and shown its performance on a variety of data sets. It builds on the underlying science of neural models of visual computation, learning and recognition, in the form of opponent-color center-surround shunting networks, Gabor multi-scale oriented filtering, Boundary Contour System theory, and Fuzzy ARTMAP pattern processing. These neural methods, in conjunction with graphical user interfaces, are combined with client-server web-based communication technologies to enable remote, collaborative exploitation of fused sensor imagery. Application to multi-modality medical imagery has already begun and looks very promising [19].

We intend to extend this fusion approach by incorporating non-imaging geospatial data, like GMTI and SIGINT, and enhancing search with additional features and multi-pass methods that treat prior search results as context. But the real challenge for the future is higher levels of information fusion for image analysis, and that too will follow from neural models of information fusion in the brain.

References


