Simultaneous Visualization of Spatial and Spectral Signatures of Targets

Dr. Melinda K. Higgins  
Georgia Tech Research Institute  
925 Dalney Street  
Atlanta, GA 30332  
404-894-4971/FAX 404-894-6285  
melinda.higgins@gtri.gatech.edu

Dr. Theodore J. Doll  
Georgia Tech Research Institute  
925 Dalney Street  
Atlanta, GA 30332  
404-894-0022/FAX 404-894-6199  
ted.doll@gtri.gatech.edu

Mr. Nickolas L. Faust  
Georgia Tech Research Institute  
925 Dalney Street  
Atlanta, GA 30332  
404-894-0021/FAX 404-894-6199  
nick.faust@gtri.gatech.edu

Mr. Tony Wasilewski  
Georgia Tech Research Institute  
925 Dalney Street  
Atlanta, GA 30332  
404-894-0133/FAX 404-894-6199  
tony.wasilewski@gtri.gatech.edu

Abstract - Traditional multispectral/hyperspectral image (MSI/HSI) data analysis focuses on single pixel-at-a-time analysis and pattern matching to known or trained spectral signatures. Nearest neighbor and object feature classification techniques are not fully exploited through traditional MSI/HSI approaches. Inputs from multiple sensors of varying formats, spectral resolution, and spatial resolution are beginning to be fused together. These new fused datasets make “target signatures” more complex thus necessitating more generalized approaches to target pattern recognition.

To date, the Georgia Tech Research Institute has been successful in exploiting fused datasets complied from different sensors of varying spatial resolution and spectral content. This work has leveraged the Georgia Tech Vision (GTV) model which is an artificial vision software system based upon human neurophysiology. This paper will cover the data analysis approach and the results from fused datasets for various applications.

Keywords: Signatures Analysis, Target Recognition, Visualization, Pattern Recognition, and Hyperspectral Imagery

1 Background

Recent progress at the Georgia Tech Research Institute has been made to exploit the fusion of datasets from different sensors of varying spatial resolution and spectral content. This work leverages the Georgia Tech Vision (GTV) model which is an artificial vision software system based upon human neurophysiology. The GTV system employs spatial and temporal frequency and chromatic (spectral) analysis for the discrimination and identification of features and/or targets within a scene. GTV has been successfully applied to many imagery sources including visual, FLIR, multispectral, and SAR. This system is currently used by the Army AMCOM to evaluate camouflage and IR signature suppression. It has also been applied to prediction of operator visual performance in air-defense systems, the evaluation of night-vision sensor performance; and evaluation of the dynamic effects of illumination changes on target recognition performance. GTV has also been deployed in automatic food products inspections and identification of tumors in biomedical imagery.

Prior applications of GTV have focused on single band or simple RGB (3-band composite) images. This paper will show the expansion of GTV to handle multiple bands from multiple data sources (e.g. CIB, IRS, Landsat, Positive Systems). For each input image, GTV produces multiple output images based on spatial frequency and orientation of objects within the scene. Thus, for each image input, GTV produces a data cube output consisting of x pixels by y pixels (for the image) by f frequency/orientation filter outputs. This process is continued for every input image in the fused sensor dataset, which produces a complete data cube x by y by f by λ (lambda) “bands” of spectral information. Each object's "signature" can then be represented by a four dimensional surface which captures not only the objects spectral signature, but also its spatial characteristics.

Over the past decade, the Georgia Tech Research Institute developed an end-to-end simulation of the human vision system called the Georgia Tech Vision (GTV) model. The “end-to-end” designation indicates
that GTV was designed to simulate all processes from image encoding to visual search and detection performance. GTV’s two most important capabilities are: [1]

A) the ability to generalize appropriately and
B) the ability to adapt to different and changing targets over time.

The algorithms employed in GTV are consistent with neurophysiological evidence concerning the organization and function of parts of the human vision system, from dynamic light adaptation processes in the retinal receptors and ganglia to the processing of motion, color, and edge information in the visual cortex. In addition, GTV models human selective attention, which is thought to involve feedback from the visual cortex to the lateral geniculate nucleus in the thalamus [1].

The GTV simulation is based on basic research in vision, attention, and perceptual decision making. The simulation incorporates findings from research on low-level visual processes, including computational vision models, and from the visual search, selective attention, color science, motion and flicker perception, and signal detection theory literatures. In the GTV simulation, these findings have been integrated into a single comprehensive simulation of visual performance [1].

A key feature of GTV is that it is an integration of many different computational vision algorithms. The model simulates the chain of visual processes that leads to visual search and a detection decision, starting with dilation of the pupil and responses of the retinal receptors, and including subtractive and multiplicative luminance adaptation, sensitivity to temporal modulation, and color opponency. GTV also includes spatial pattern processing characteristic of simple and complex cortical cells, selective attention, the results of task-specific perceptual learning, and decision processes [2].

An important feature of the GTV model is the ability to predict sequential dependencies in observer fixations during search. This part of GTV is called the “systematic search model.” It accounts for observer behavior during prolonged viewing of the same scene. Specifically, when observers visually inspect clutter objects in a scene, they often learn to reject some of them as possible targets. This learning process reduces the effective clutter level for that observer, and increases the probability that the observer detects a target when one comes in view. The addition of the systematic search algorithms to GTV allows it to better predict search and detection performance in field test conditions and, therefore, makes the model easier to validate.

Perhaps the most significant aspect of the GTV model is the fact that it models three important, closely inter-related properties of the human vision system:

1) the ability to process large amounts of stimulus information to a limited extent in parallel (preattentive processing);
2) the ability to select regions and/or features in the field of view for further processing (selective attention); and
3) the modification of selective attention and search performance with training (perceptual learning).

Many everyday tasks, like military target acquisition and diagnostic inspection of medical imagery, involve extensive practice. Therefore, it is important to model the effect of learning about conspicuity and visual search performance. One way of modeling the effect of learning is to change the relative weights of the low-level vision properties that contribute to conspicuity, as suggested by Koch and Ullman (1985) [3]. The GTV model includes a routine that models what observers learn as a result of experience with particular sets of targets and backgrounds. This routine, which is based on discriminant analysis, automatically modifies the weighting of low-level properties in the computation of object conspicuities.

2 Approach/Algorithm Described

The GTV algorithm includes five major components: (see Figure 1)

1) front-end;
2) preattentive processing;
3) attentive processing;
4) selective attention/training; and
5) performance modules.

Front-End: The front-end module simulates the initial processing stages of the human visual system, including receptor pigment bleaching, pupil dilation, receptor thresholds and saturations, color opponency, and the dynamics of luminance adaptation and sensitivity to motion, flicker, and transient flashes. The inputs to this module are images with the spectral characteristics of the retinal receptors. The outputs are color-opponent photopic and scotopic signals that include effects due to receptor thresholds and saturations. The temporal, spatial, and intensity characteristics of these output signals also reflect the effects of time-varying luminance adaptation processes. Signal intensities of individual areas of these output images are enhanced due to effects of motion, flicker, and variations in luminance level within the image.
defined by segmenting the preattentive output image, which is done by the selective attention/training unit. Search performance is quantified in terms of a probability of fixation, $P_{\text{fix}}$, for each blob in the preattentive output image. Discrimination performance is quantified in terms of a probability that the observer indicates “yes, the blob is a target,” given that it is fixated, $P_{\text{yesfix}}$. Additional detail of GTV outputs can be found in the VISEO User’s Manual [4].

3 Results

To date, GTV has been applied successfully to modeling human observer performance in the recognition of targets over various sensor platforms; to analyzing the conspicuity of targets; to characterizing complex objects; and to analyzing higher-level fused datasets. Some results from these applications are described below.

Reducing the Conspicuity of a Target: Based on the colors, texture, and spatial frequency pattern of the background, GTV was used to help design a camouflage pattern to reduce the conspicuity of this helicopter. The top image in Figure 3 shows the initial camouflage paint design applied based on color and background clutter patterns. This initial design reduced the conspicuity of the original black paint only slightly. An additional step was applied to reduce the shadows on the sides and bottom by increasing the intensity of the paint, which changed the reflectivity, thus, reducing the conspicuity even further (Figure 3, bottom).

Discrimination of Multiple Objects: Figure 4 shows the GTV outputs for a single input image (mid-wave infrared, MWIR) of a face. The outputs are shown in order from lowest spatial frequency to highest spatial frequency. All outputs shown are for 0 deg orientation. However, channel outputs for 45 deg, 90 deg, and 135 deg orientations were also generated. All of these spatial frequency and orientation channel outputs were then used to discriminate 7 “faces.” The results of this discrimination is shown at the bottom of Figure 4.

Figure 1: GTV Model Algorithms – Overview

Preattentive Processing: The preattentive module simulates pattern perception in the peripheral visual field, which directs the focus of attention during visual search. The outputs of these preattentive module are images of the same dimensions as the input. There are up to 208 different images, each representing the result of filtering the input with a different filter. The filters for each of these 208 channels have differing spatial frequency/orientation bandpass characteristics. They also represent different color-opponent signals and the various types of retinal receptor outputs (see Figure 2).

Attentive Processing: The attentive processing module simulates close visual inspection and its outputs are multiple images of the same dimensions as the inputs (see Figure 2). These images are combined into a pooled attentive output image by the selective attention/training module. The signal for the target in this pooled image is a measure of its discriminability from background clutter. The signal values of non-target blobs in the pooled attentive output image are used to calculate the probability that the observer “false alarms” to each object. This computation is done by the GTV performance module.

Selective Attention/Training: The selective attention/training module uses the preattentive output images, in both the training mode and subsequent analysis runs, to autonomously segment the input images and discriminate the target from clutter. In the training mode, this routine collects data on what channel outputs characterize targets and clutter. In the analysis mode, it uses a discriminant function, based on that data, to segment the scene into objects or “blobs” that are target candidates. This module outputs a pooled preattentive image that identifies the conspicuities of objects in the filed of view, i.e. the extent to which the objects attract the observer’s attention.

Performance Module: The performance module computes a probability of fixation and a probability of detection or false alarm for each “perceptual object” in the field of view. These computations are based on the output images from the preattentive and attentive processing modules. Perceptual objects, or “blobs,” are
Higher-Level Image Analysis (Multispectral Imagery): The following example shows GTV’s capability to recognize particular features in GIS imagery. In this example, the system was trained to recognize housing sub-divisions on one set of imagery, and then tested on a second set. Figure 5 below shows a multi-band IR test image (RGB composite image) and GTV’s output (bottom) showing the areas which were classified as sub-divisions from the test image. This example shows the capability of GTV for higher-level image interpretations, where “objects” are not classified simply by themselves from their color and shape characteristics, but also by their relationships to other objects (i.e., sub-divisions consist of typically smaller buildings (houses), cars, closer compacted roads and driveways, etc.).
In this study, these were replaced by image data collected by the Positive Systems 4-band sensor (3 bands in the visible and 1 band in the near IR), Figure 6. [Note: This data was provided under the “Multi-Modality and Image Fusion” study sponsored by Eastman Kodak through the National Reconnaissance Office (NRO000-98-D-2132 Task #7), October 1999.]

Each band was input separately into GTV and 24 filter channel output images were generated: the 24 channels consisted of 6 spatial frequencies and 4 orientations (0 deg, 45 deg, 90 deg, 135 deg). Thus, the resulting “hyper-data” cube was 490 pixels by 490 pixels (the image size) by 4 “bands” by 24 channels. Pixels within this hyper-data cube, thus, had “signatures” consisting of 4 spectral bands by 24 spatial filter channels or 96 values which represented that pixel’s “signature.” A better way to show the pixel’s signature instead of plotting all 96 values on a 2D plot, was to plot a 4D signature surface consisting of 4 bands by 6 frequencies by 4 orientations by their intensities (shown in Figure 8 below). Figure 7 shows some of the objects of interest selected for discrimination: at the top of image and scattered throughout, pixels from vehicles were selected; pixels from within the two rows of houses on the right side of the image were also selected; pixels from the “U-turn” arrangement of larger buildings with “textured” or “gabled” roof structure were also selected; and pixels from similarly larger buildings without textured roofs (“un-gabled”) were selected. Each of these groups of pixels were then discriminated using these spatial/spectral signature surfaces (Figure 9).

![Figure 6: Positive Systems Data: 4 bands (VIS-NIR)](image)

![Figure 7: Selected Objects for Simultaneous Spatial/Spectral Discrimination: (top) “Color Composite” of 3 Visible Bands; (bottom) Objects of Interest Highlighted [blue=cars; red= large gabled buildings; yellow=un-gabled buildings; green=houses](image)

Notice in the following surface plots (Figure 8) the discriminating “features” which stand out. The two sets of quads (i.e. the set of 4 surface plots) are from buildings of similar footprints (areas), however, the upper set of 4 plots are from the buildings with textured roofs, which is reflected by the higher intensities of the higher spatial frequencies (spatial frequency increases going down the bottom left axis on each plot; orientation increases going up the right axis). The lower set of 4 surface plots (from the un-gabled buildings) do not show these higher frequency features. Similar higher-frequency “features” were also noted for both the houses (which were smaller in area as well as having some “points” to their roofs) and vehicles, which had the smallest footprints (neither are shown here). Better understanding of these spatial/spectral signature correlations and feature highlights is currently being pursued.
Figure 8: Spatial/Spectral Signature Surfaces for Objects of Interest: Surfaces for Large Gabled Buildings; (top 4 plots) Surfaces for Un-gabled Buildings; (bottom 4 plots) Surfaces for Houses and Vehicles not shown. [Note: Within each set of 4 surface plots, the upper left is from visible band 1; upper right is visible band 2; lower left is the near IR band; and the lower right is visible band 3.]

Figure 9 below shows the results from a discriminant analysis of the combined spectral/spatial “signature” from each “target of interest” [i.e. Large Gabled Buildings; Un-gabled Buildings, Houses, and Vehicles].

Sensor Fusion: Analysis and Discrimination: In a manner analogous to the previous example, GTV was run on image data from CIB, IRS (© [1999] Space Imaging L. P.), and Landsat sensors (all geo-registered). Figure 10 shows a) some of the objects selected for discrimination within this data set, the input layer from IRS and b) the other input images from the 7-bands of Landsat.

Figure 9: Discriminant Analysis Results of Selected Objects of GTV Output “Hyper-data” Cube Spatial/Spectral Signatures from 4-band Positive Systems Data.

Figure 10a: CIB Input Image (top) with Selected Objects Highlighted and IRS Input Image

[CIB Image (1m): yellow=Un-gabled Buildings; blue=Gabled Buildings; red and green=Houses]

[IRS © 1999 Space Imaging L. P. (approx. 6m)]
The resulting “hyper-data” cube was 256 pixels by 256 pixels (image size) by 9 “bands” (1 CIB + 1 IRS + 7 Landsat) by 24 spatial filter channels. From this cube the spatial/spectral signature surfaces were generated (Figure 11). [Note that Figure 11 only shows 4 of the 9 “surfaces” for the CIB, IRS, and Landsat band 3 and band 7; the other 5 Landsat bands are not shown, but they were used in the discrimination step.] The discrimination of the selected objects is shown in Figure 12.

4 Conclusions/Future Directions

To date, GTV has been applied successfully to modeling human observer performance in the recognition of targets over various sensor platforms; to analyzing the conspicuity of targets; to characterizing complex objects; and to analyzing higher-level fused datasets. However, Georgia Tech is involved in a number of lines of research to refine, extend, and apply the integrated spatial/spectral pattern recognition tools discussed in this paper. Additional applications include the face recognition (preliminary results provided above), recognition of tumors in biomedical imagery, evaluation of image quality, and further identification of features and objects from fused reconnaissance sensors.

Some of these refinements include optimization of the software to reduce run-time, and the addition of unsupervised classification algorithms. Extensions include the addition of algorithms to simplify the training process when multiple targets are of interest,
and implementation of additional capabilities and features of the human visual system, such as stereopsis, accommodation, and additional aspects of visual cognition related to event understanding and active inference-making. The emulation of other biological vision systems are also being explored, such as birds or prey which can “see” in up to 5 spectral bands with high spatial acuity; and insects, some of which “see” well in the UV spectrum and/or perceive light polarization differences.

5 Acknowledgements

The primary support for the development of the GTV model was provided by the Army Aviation and Troop Command, Aviation Applied Technology Directorate (AMCOM/AATD) under the Visual/Electro-Optical (VISEO) Detection Analysis program (Contract No. DAAJ02-92-C-0044). Additional support was provided by the Air Force Special Operations Command (AFSOC) through the Air Force Warner Robbins Air Logistics Center (WRALC/LNXEA) at Robbins Air Force Base, GA and the Army Aviation and Troop Command (ATCOM/AMSAT-B-Y) at Saint Louis, MO.

The Georgia Tech Research Institute would also like to thank the National Imagery and Mapping Agency (NIMA) for providing the CIB, IRS. Landsat, and Positive Systems imagery shown above. This data was provided under the “Multi-Modality and Image Fusion” study sponsored by Eastman Kodak through the National Reconnaissance Office (NRO000-98-D-2132 Task #7), October 1999.

References


