SENSOR FUSION AND IMAGE INTEPRETATION THROUGH INTEGRATED SPATIAL/SPECTRAL PATTERN RECOGNITION

June 2000

Dr. Melinda K. Higgins; Dr. Theodore J. Doll;
Mr. Nick L. Faust; Mr. Anthony A. Wasilewski
Georgia Tech Research Institute
Atlanta, GA 30332

ABSTRACT

Typical 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 only beginning to be fused together.

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

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (FROM - TO)
01-06-2000	Conference Proceedings	xx-xx-2000 to xx-xx-2000

4. TITLE AND SUBTITLE
Sensor Fusion and Image Intepretation Through Integrated Spatial/Spectral Pattern Recognition
Unclassified

5a. CONTRACT NUMBER
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Higgins, Melinda K. ;
Doll, Theodore J. ;
Faust, Nick L. ;
Wasilewski, Anthony A. ;

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME AND ADDRESS
Georgia Tech Research Institute
Atlanta, GA30332

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME AND ADDRESS
Director, CECOM RDEC
Night Vision and Electronic Sensors Directorate Security Team
10221 Burbeck Road
Ft. Belvoir, VA22060-5806

10. SPONSOR/MONITOR'S ACRONYM(S)
11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
APUBLIC RELEASE
,

13. SUPPLEMENTARY NOTES
See Also ADM201258, 2000 MSS Proceedings on CD-ROM, January 2001.

14. ABSTRACT
Typical multispectral/hyperspectral image (MSI/HSI) data analysis focuses on single pixel-at-atime 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 only beginning to be fused together.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	Public Release	20	Fenster, Lynn lfenster@dtic.mil
Unclassified	Unclassified	Unclassified			19b. TELEPHONE NUMBER International Area Code Area Code Telephone Number 703767-9007 DSN 427-9007

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.

1.0 INTRODUCTION/BACKGROUND

Over the past decade, the Georgia Tech Research Institute has 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]

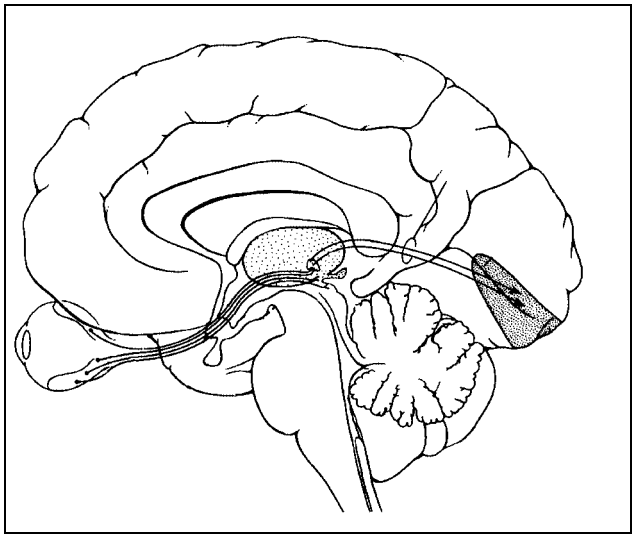


Figure 1. Information Flow in the Human Visual System from Retinal Ganglia to Thalamus to Visual Cortex

[1] Doll, T. J.; McWhorter, S. W.; Wasilewski, A. A.; Schmieder, D. E. “Robust, Sensor-Independent Target Detection and Recognition Based on Computational Models of Human Vision.” *Optical Engineering*, **37**(*7*), 2006-2021 (July 1998).

GTV accepts sequences of images representing a target moving through a cluttered background at 30 Hz or greater. It segments the images into objects and outputs predictions of the observer's search and detection performance over these objects. These predictions are generated at the rate of 3 Hz (in non-real, simulation time), which is the typical rate at which observers generate fixations during search of natural scenes. The prediction metrics include the probability that an observer fixates the eyes on each object and the probability that an observer will decide that an object is a target, given that it is fixated.[1] GTV also produces multiple-channel outputs, which characterize each segmented object in terms of spatial frequency, orientation, and chromaticity.[2]

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.[3]

The GTV model is an advance over earlier target acquisition models in that it generates accurate predictions for targets that are camouflaged or have a suppressed IR signature, are embedded in a cluttered scene, and/or are partially obscured by terrain or vegetation. Some of the earlier models were sensitive only to the average difference between the target and background luminances, and, therefore, could not handle pattern and color similarities and differences between the target and background. GTV uses computational models from spatial vision research to predict observers' ability to discriminate pattern and color differences. Most of those earlier models also assumed that observers search randomly through the scene. In GTV, the input scene drives search predictions. It estimates a probability of fixation for each object in the scene based on the object's contrast, color, motion, and similarity to both the target and background.[3]

Historically, target acquisition models have quantified the target signal by using first-order statistics, such as the mean and standard deviation of pixel intensities.[4,5] This method works well as long as the internal contrast within the target is negligible relative to the target-background contrast. However, these statistics are not adequate when variations in contrast or temperature within the target silhouette become the dominant cues for search and detection.[6] Accurate prediction of search and detection performance for low-observable (LO) targets requires that pattern or texture differences between the target and background be taken into account. A target may be immediately and easily perceptible on a single glimpse (i.e. it may "pop-out") even though it has exactly the same average luminance as its background.[5]

An additional limitation of existing models is that they don't explicitly address the effect of clutter on the pattern of eye fixations during visual search, but instead treat visual search as a random process. Most of the models

---

[2] Doll, T.J.; McWhorter, S. W.; Schmieder, D. E.; Wasilewski, A. A. "Simulation of Selective Attention and Training Effects in Visual Search and Detection," in *Vision Models for Target Detection and Recognition*, E. Peli, Ed., World Scientific Publishing, Singapore (1995).

[3] Doll, T. J. "An Integrated Model of Human Spatial Vision," *INSIGHT: The Visual Performance Technical Group Newsletter*, **20**(*2*), 1-4 (August 1998).

[4] Doll, T. J.; McWhorter, S. W.; Wasilewski, A. A.; Schmieder, D. E. *Georgia Tech Vision (GTV) Model, Version GTV96, Analyst's Manual*, Prepared for U.S. Army Aviation and Troop Command, Aviation Applied Technology Directorate (ATCOM/AATD), Ft. Eustis, VA 23604-5577, Under Contract No. DAAJ02-92-C-0044, January 1997.

[5] Doll, T. J.; Home, R. "Lessons Learned in Developing and Validating Models of Visual Search and Target Acquisition," *NATO/Research and Technology Organization (RTO) Meeting Proceedings 45: Search and Target Acquisition*, held in Utrecht, The Netherlands 21-23 June 1999, p. 1-1 to 1-8.

[6] Walker, G. W.; McManamey, J. R. "The Importance of Second-Order Statistics for Predicting Target Detectability," *SPIE*, **vol. 1967**, 308-319 (1993).

treat clutter as a scale factor rather than as an important determinant of the search sequence. Perhaps the major effect of clutter on target acquisition is to reduce the probability of fixating on or near the target, given limited search time. Explicit modeling of the effect of clutter on the sequence of eye fixations is, therefore, crucial to accurately predicting target acquisition performance.[4]

Another shortcoming of existing models is that they ignore the perceptual decision making process. Most existing models assume a fixed threshold for detection. However, perceptual research has consistently shown that detection is not a simple threshold process. Rather, it involves decision making on the basis of uncertain evidence. The Theory of Signal Detectability (TSD) describes this process and provides a mathematical framework for predicting tradeoffs between observer detection and false alarm rates. This tradeoff is especially important to understanding observer performance in high clutter backgrounds.[7]

The GTV model is able to predict where observers will focus their attention when searching for targets (typically vehicles) in cluttered terrain backgrounds and accurately predicts the probability of detecting those targets. The model accounts for the effects of motion and chromaticity on visual search and detection under both photopic and scotopic conditions. In addition, the model predicts saccade-to-saccade search performance, and is, therefore, able to account for the effects of specific clutter patterns on search and detection performance. The model incorporates recently developed computational models of pattern perception (including "complex" or two-stage models), and is , therefore, able to predict the detection of subtle differences between targets and backgrounds. These same computational algorithms enable the model to account for the effects of target internal structure on detection performance.

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).

It is not apparent that there are any other models, including those that incorporate computational vision algorithms, that adequately model these processes, which are fundamental to visual search and detection performance and, therefore, to the design of camouflage and LO.

Preattentive processing enables human observers to process objects of potential importance over the whole field-of-view (FOV) in parallel. Perhaps the most widely studied example of preattentive processing is pop-out, the ability of persons with normal vision to perceive variations in pattern or texture in a single glimpse, without conscious effort. An object may pop-out from its background even when there is no difference in the average intensities or chromaticities of the object and the background.[8] The term "pop-out" is reserved for pattern differences that are perceived effortlessly, in contrast to those that must be discriminated, i.e. require close inspection or focused attention. Preattentive processing also provides inputs that can be used to allocate processing resources for further, more detailed, analyses (i.e. for attentive processing).

---

[7] Doll, T. J.; Schmieder, D. E. "Observer False Alarm Effects on Detection in Clutter," *Optical Engineering*, **32**, 1675-1684 (1993).

[8] Bergen, J.R. "Theories of Visual Texture Perception," in *Vision and Visual Dysfunction, Vol 10B of Spatial Vision*, D. Regan, ed., MacMillian publisher, New York, 114-134 (1991).

Preattentive processing is fundamental to visual search performance. There is substantial evidence that eye movements (saccades) during visual search are guided by preattentive processing of pattern information in peripheral vision. For example, recordings of eye movements over structured scenes reveal that the eye fixates on features such as edges and corners that are more likely to convey information than are plain surfaces.[9] In reading, the eyes of proficient readers search out larger words, which convey a high degree of meaning, rather than articles.[10] Visual search proficiency has even been used as a measure of peripheral visual acuity.[11,12,13]

Thus, it seems likely that preattentive processing of patterns outside the momentary area of focal attention plays an important role in determining where the observer will focus attention next. Several investigators have suggested that the "spotlight" of focal attention falls on locations in the visual field according to their conspicuity, which is a combination of features derived from early visual processing.[14,15] The GTV model uses a similar formulation. The probability that focal attention is directed to a given object or location is directly related to the extent to which it "pops-out," which is termed conspicuity.

In GTV, conspicuity depends on luminance contrast, chromatic contrast, temporal modulation, and texture differences. Perceptual segregation of textures is simulated using complex spatial frequency channels models, such as those developed by Graham and Wilson and their colleagues.[16,17,18] Since focal attention is closely related to fixation (though not necessarily identical), the probability that focal attention is directed to an object is called its probability of fixation, P_{fix}.

The second property of the visual system mentioned earlier, selective attention, focuses processing resources on a limited region of the visual field for purposes of further analysis, such as object recognition. A number of investigators have argued that the complexity of the processing implied by tasks such as object recognition would make it prohibitive for a biological processor to perform such operations in parallel over the whole visual field.[14,15] Selective attention allows the observer to adaptively allocate limited resources available for attentive processing to different parts of the visual field at different times.

But what rule is used to select one region of the visual field for further processing and exclude others? Koch and Ullman (1985)[14] have suggested that the outputs of the preattentive stage (low-level vision properties) are differentially weighted, and that the "spotlight" of attention is directed to the region of the visual field with the greatest weighted output, or conspicuity. In GTV, the computation of weights that predict conspicuity is performed by the *selective attention/training unit.*

Part of the adaptability of the human visual system derives from the third topic area mentioned earlier, perceptual learning, or the modification of selective attention through learning. Human observers learn "what to look for" and greatly improve their performance in the course of searching for particular targets in a given type of

---

[9] Gould, J. D. "Looking at Pictures," in *Eye Movements and Psychological Processes*, R. A. Monty & J. W. Senders, eds., Lawrence Erlbaum Associates publishers, Hillsdale, NJ (1976).
[10] Rayner, K. "Foveal and Parafoveal Cues in Reading," in *Attention and Performance, Vol 7*, J. Requin, ed., Lawrence Erlbaum and Associates publishers, Hillsdale, NJ (1978).
[11] Johnston, D. M. "Search Performance as a Function of Peripheral Acuity," *Human Factors*, **7**, 528-535 (1965).
[12] Erikson, R. A. "Relation Between Visual Search Time and Visual Acuity," *Human Factors*, **6**, 165-178 (1964).
[13] Bellamy, L. J.; Courtney, A. J. "Development of a Search Task for the Measurement of Peripheral Visual Acuity," *Ergonomics*, **24**, 497-509 (1981).
[14] Koch, C; Ullman, S. "Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry," *Human Neurobiology*, **4**, 219-227 (1985).
[15] Sandon, P. A. "Simulating Visual Attention," *Journal of Cognitive Neuroscience*, **2**, 213-231 (1989).
[16] Graham, N. "Complex Channels, Early Local Nonlinearities, and Normalization in Texture Segregation," in *Computational Models of Visual Processing*, M. S. Landy and J. A. Movshon, eds., MIT Press publishing, Cambridge, MA (1991).
[17] Graham, N.; Beck, J.; Sutter, A. "Nonlinear Processes in Spatial-Frequency Channel Models of Perceived Texture Segregation: Effects of Sign and Amount Contrast," *Vision Research*, **32**, 719-743 (1992).
[18] Wilson, H. R.; Richards, W. A. "Curvature and Separation Discrimination at Texture Boundaries," *Journal of the Optical Society of America, A*, **9**, 1653-1662 (1992).

background.[19,20] The results of such studies suggest that learning modifies the extent to which various low-level vision properties contribute to the conspicuity or salience of objects.

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 on pop-out 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).[14] 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.

2.0 APPROACH/ ALGORITHM DESCRIPTION

The GTV algorithm includes five major components: (see Figure 2)

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.

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 3).

[19] Neisser, U; Novick, R.; Lazar, R. "Searching for Ten Targets Simultaneously," *Perceptual and Motor Skills*, **17**, 955-961 (1963).
[20] Schneider, W.; Shiffrin, R. "Controlled and Automatic Human Information Processing I: Detection, Search, and Attention," *Psychological Review*, **84**, 1-66 (1977).

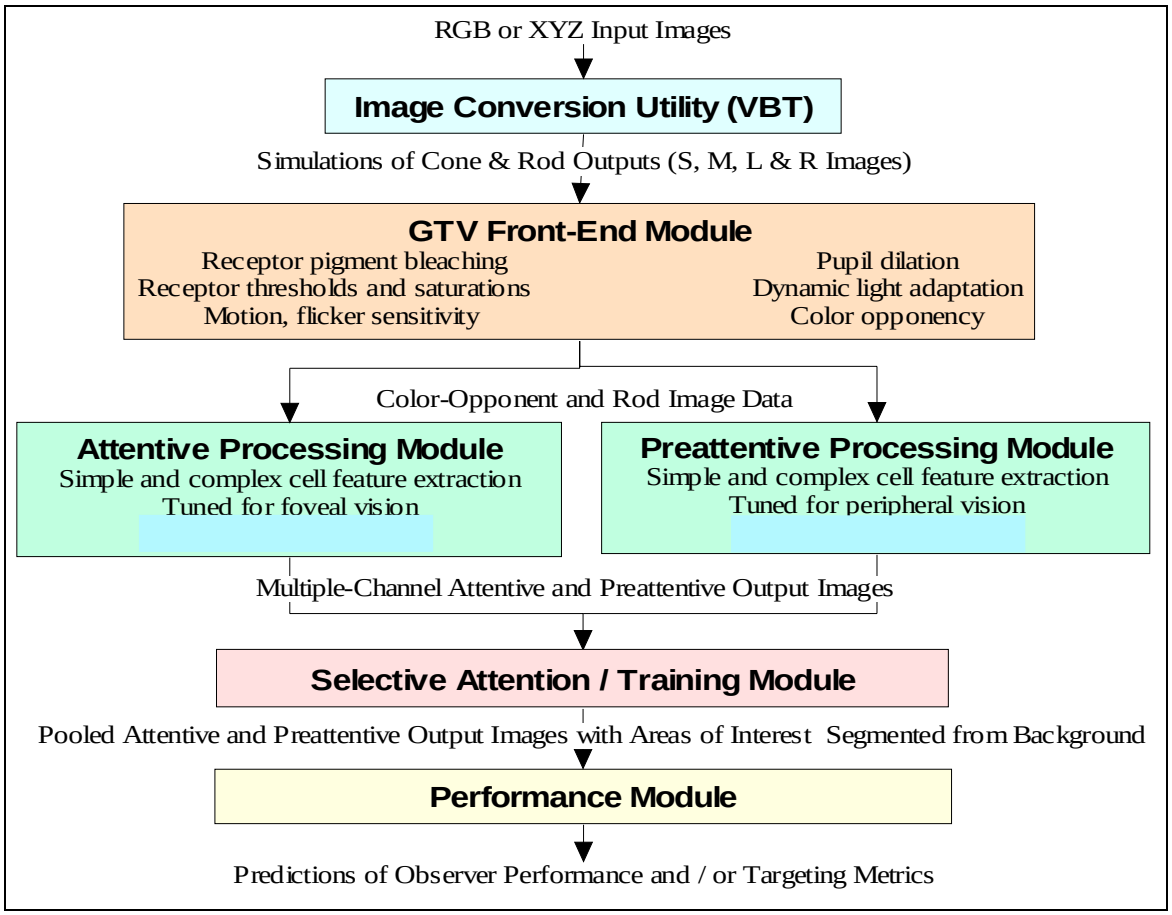


Figure 2. High-Level Overview of GTV Model Algorithms

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 3). 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 disciminability 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.

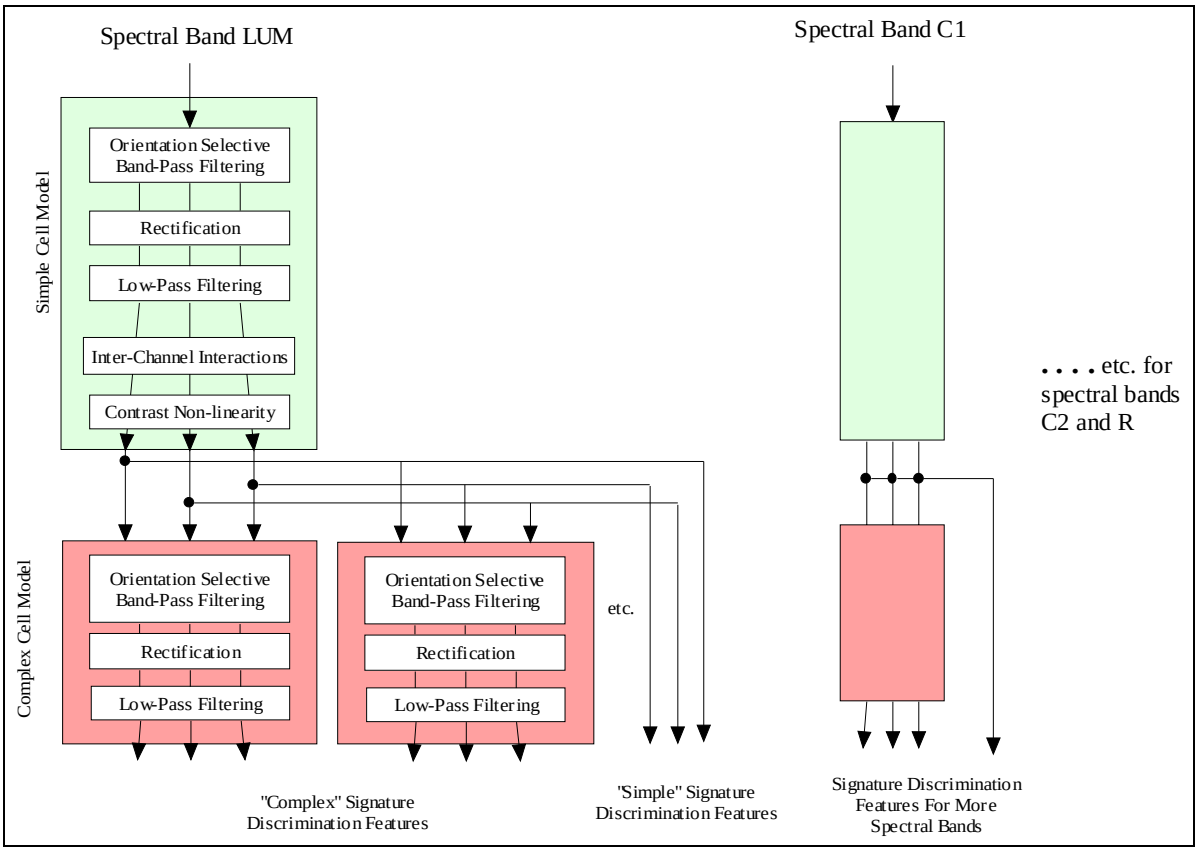


Figure 3. GTV Feature Extraction Algorithms

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 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_{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, $P_{yes/fix}$. Additional detail of GTV outputs can be found in the VISEO User's Manual.[21]

[21] McWhorter, S. W.; Doll, T. J.; Hetzler, M. C.; Wasilewski, A. A.; Stewart, J. M.; Schmieder, D. E.; Owens, W. R.; Scheffer, A. D.; Galloway, G. L.; Harbert, S. L. *Visual/Electro-Optical (VISEO) Detection Analysis Systems User's Manual*, Prepared under contract no. DAAJ02-92-C-0044 with the Army Aviation and Troop Command, Aviation Applied Technology Directorate (ATCOM/AATD), Georgia Tech Research Institute, Atlanta, GA (1997).

3.0 IMAGE EXPLOITATION EXAMPLES

This section provides examples of the image analysis products from GTV. These examples include:

- Complex Texture Segregation
- Clutter Rejection Performance
- Reducing the Conspicuity of a Target
- Discrimination of Multiple Objects
- Higher-Level Image Analysis (Multispectral Imagery)
- Simultaneous Spatial/Spectral Signatures: Analysis and Discrimination
- Sensor Fusion: Analysis and Discrimination

Complex Texture Segregation

Figure 4 shows the ability of GTV to segregate complex textures utilizing complex-cell cortical filters within the preattentive processing module described above. The center image in Figure 4 shows the output from only simple cortical cell filters of the input image (left). Notice that there is no differential signal that distinguishes the center irregularity (center). However, the right image in Figure 4 shows the output from multiple second stage filters applied to the first stage "simple" filter output (i.e. the center image).[5] GTV uses the results from the second stage filters to determine texture boundaries.

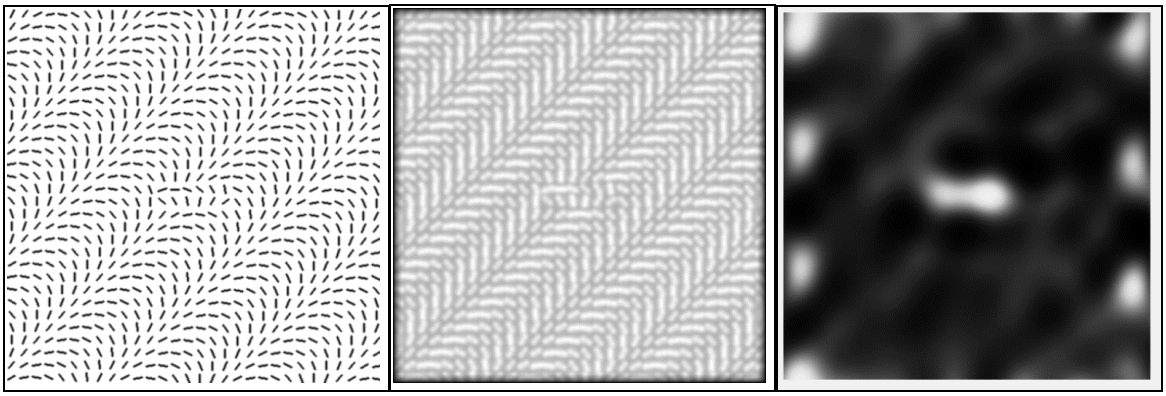
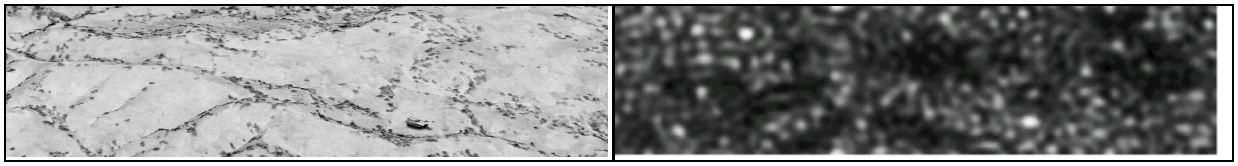


Figure 4. Complex Texture Segregation Example: (left) Input Image; (center) Simple- Cell Cortical Filters (single-stage) Output; (right) Complex-Cell Cortical Filters (second-stage)[22]

Clutter Rejection Performance

It has been shown that with extensive practice military observers are often able to immediately pick out targets in cluttered scenes that novice observers must search for painstakingly.[19,20] These observers have evidently learned to preattentively process the target. One way to model this "pop-out" and visual search performance is to differentially weight the filter-channel outputs before pooling them into a single salience map.[5] GTV uses this method by employing a discriminant analysis routine to compute the weights (within the selective attention algorithm), which is highly effective in rejecting clutter as shown in figure 5.



[22] Northdurft, H. C. "Research Note: Texture Segmentation and Pop-Out From Orientation Contrast," *Vision Research*, **31**, 1073-1078 (1991).

Figure 5A. With No Selective Attention Algorithm (Left is input image; Right is model output).

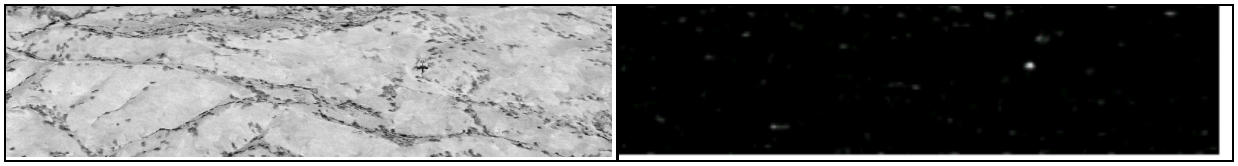


Figure 5B. With Selective Attention Algorithm (Left is input image; Right is model output).

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 left image in Figure 6 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 6, right).

Figure 6. GTV Camouflage Design: (left) Color Matching to Background;
(right) Additional Correction for Lighting Intensities (i.e. the shadows on the side and bottom were reduced)

Discrimination of Multiple Objects

Figure 7 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 in Figure 8.

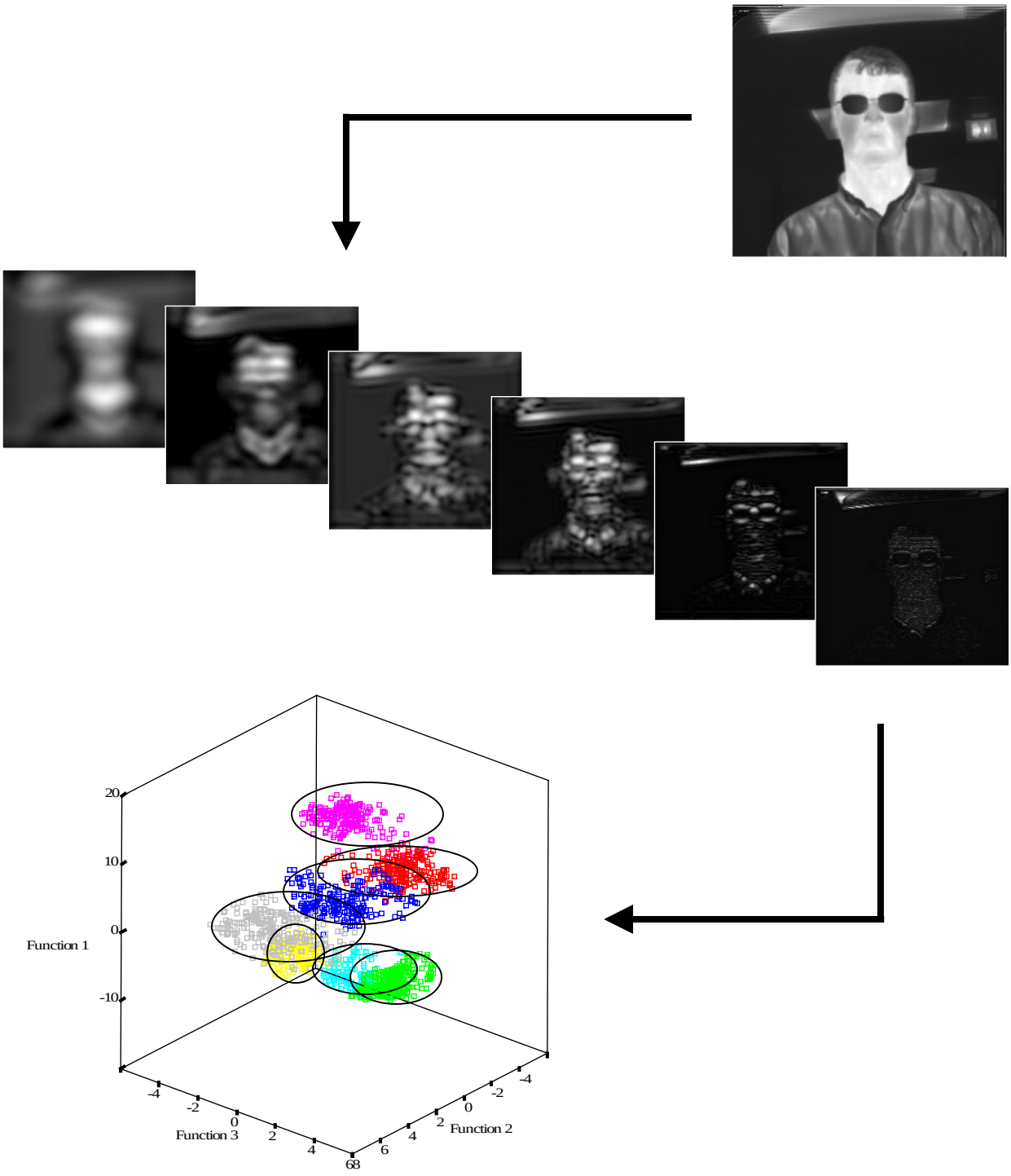


Figure 7. GTV Analysis of Face Images in the MWIR (The multiple spatial frequency and orientation filter output channels were differentiated for 7 different faces, shown in the discriminant analysis plot in the lower left.)

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 8 below shows a multi-band IR test image (RGB composite image) and GTV's output (right) 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.).

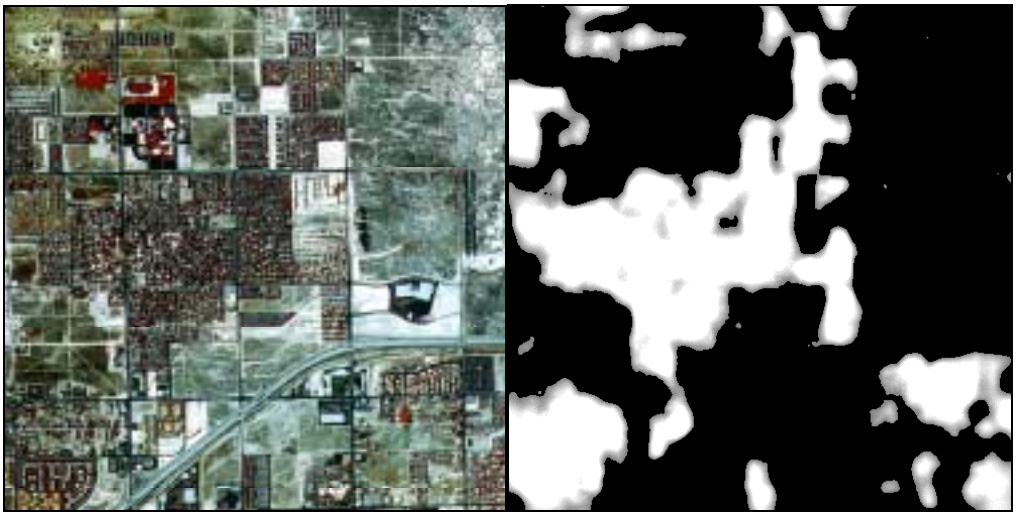


Figure 8. GTV Recognition of GIS Features in Multispectral Imagery
(left) RGB Composite Image of Multispectral Input Image; (right) GTV Output Recognizing Subdivisions

Simultaneous Spatial/Spectral Signatures: Analysis and Discrimination

So far, the examples have shown the target discrimination capabilities of GTV using single band images or RGB composites. This example and the next demonstrate the results of running GTV on multiple spectral bands and types of sensor system outputs. Normally, the spectral bands used by GTV represent the outputs of the three types of cones in the retina. 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 9. [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 x 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 11 below). Figure 10 shows some of the objects of interest selected for discrimination: at the top of image and scattered through out, 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 ("ungabled") were selected. Each of these groups of pixels were then discriminated using these spatial/spectral signature surfaces (Figure 12).

Figure 9. Positive Systems Data: 4 bands (VIS-NIR)

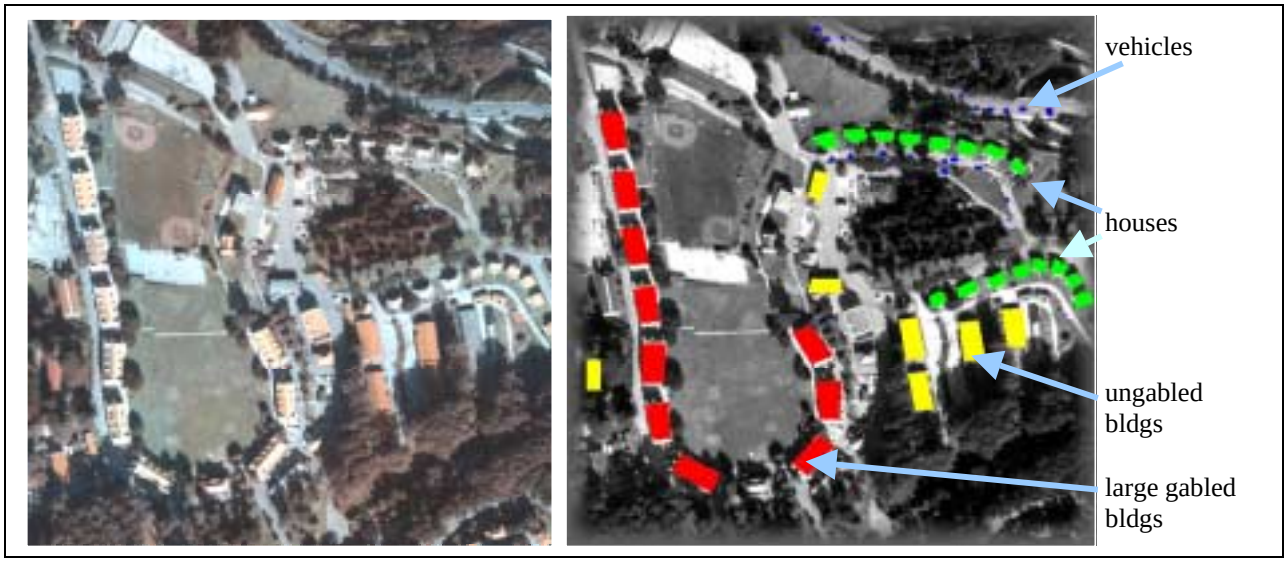


Figure 10. Selected Objects for Simultaneous Spatial/Spectral Discrimination:
(left) “Color Composite” of 3 Visible Bands; (right) Objects of Interest Highlighted

Notice in the following surface plots the discriminating “features” which stand out. The upper two quads are from buildings of similar footprints (areas), however, the upper left quad is 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). Similar higher-frequency “features” are 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. Better understanding of these spatial/spectral signature correlations and feature highlights is currently being pursued.

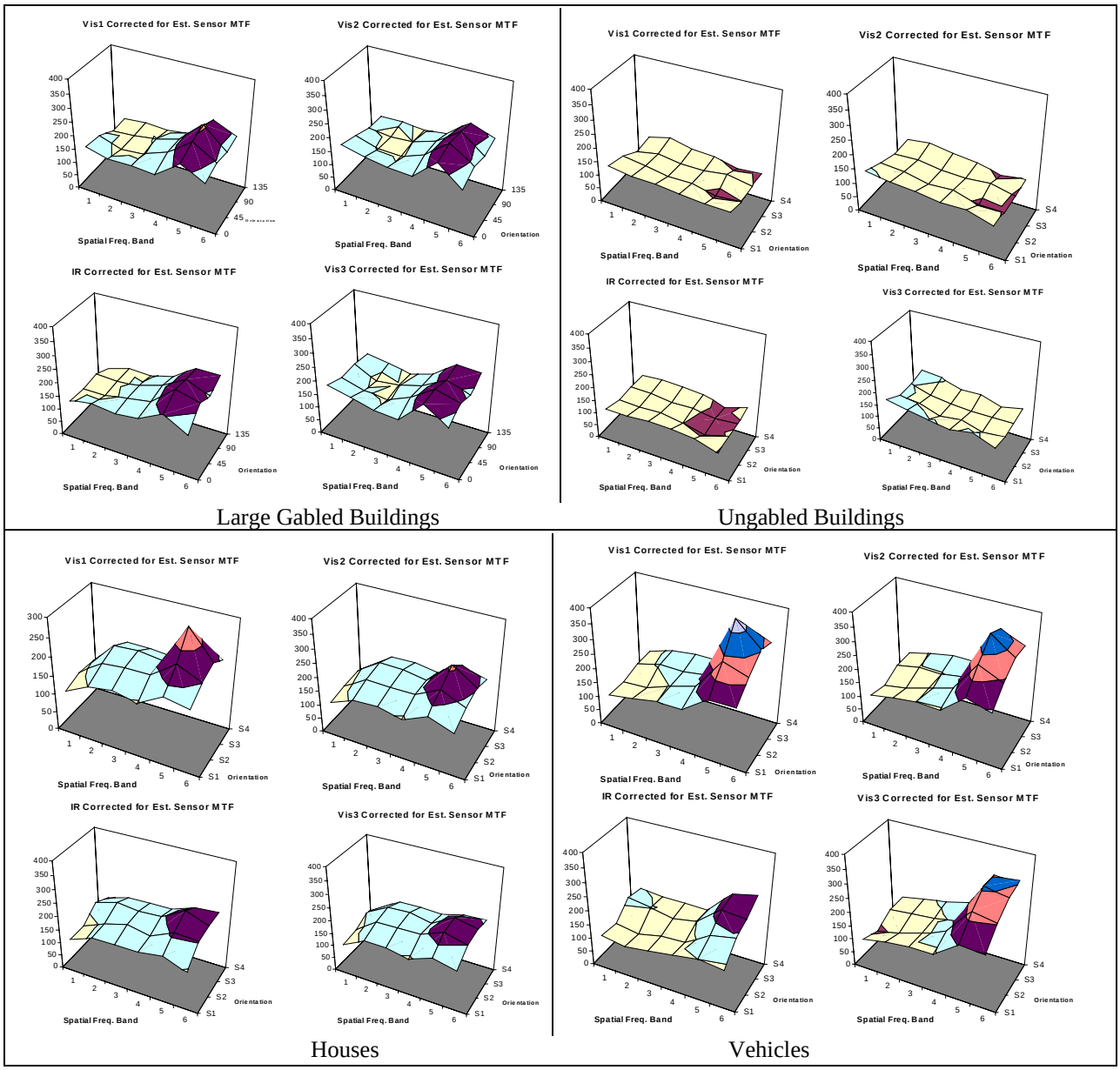


Figure 11. Spatial/Spectral Signature Surfaces for Objects of Interest:
(upper left) Surfaces for Large Gabled Buildings; (upper right) Surfaces for Ungabled Buildings;
(lower left) Surfaces for Houses; (lower right) Surfaces for Vehicles.
[Note: Within each quad surface plot, the upper left is from visible band 1;
upper right is visible band 2; lower left is the near IR band; and the lower right is visible band 3.]

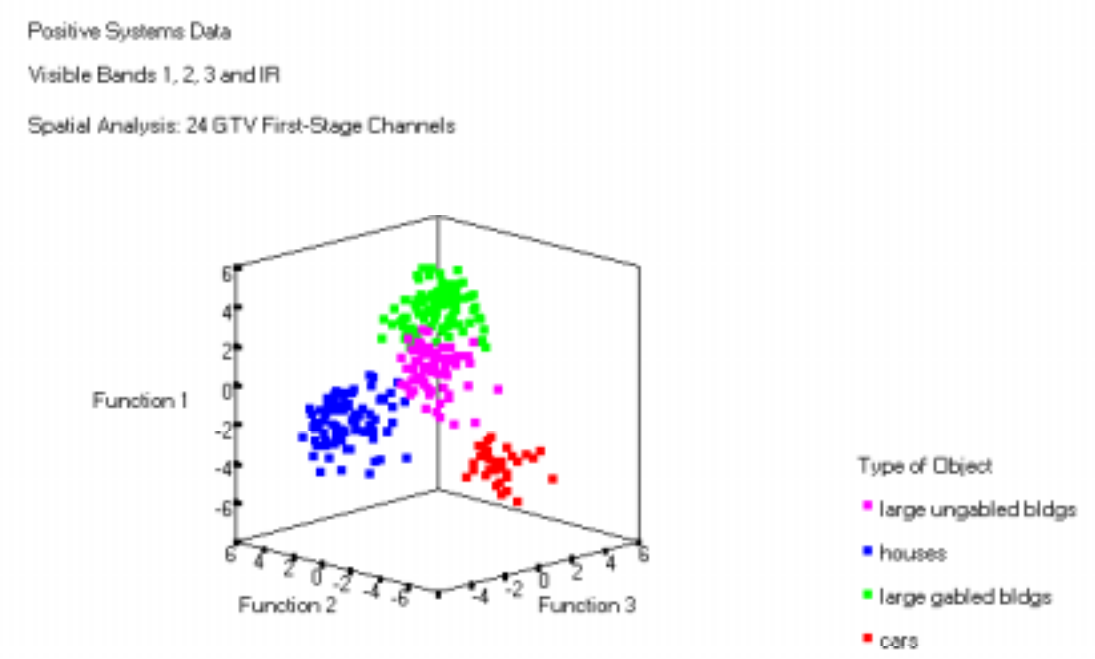


Figure 12. Discriminant Analysis Results of Selected Objects of GTV Output "Hyper-data" Cube Spatial/Spectral Signatures from 4-band Positive Systems Data.

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 13 shows some of the output images from the CIB data input only. Each of the 4 orientation filter outputs (0, 45, 90, 135 degrees) from the lowest and highest spatial frequencies are shown. Figure 14 then shows a) some of the objects selected for discrimination within this data set and b) the other input images from IRS and Landsat.

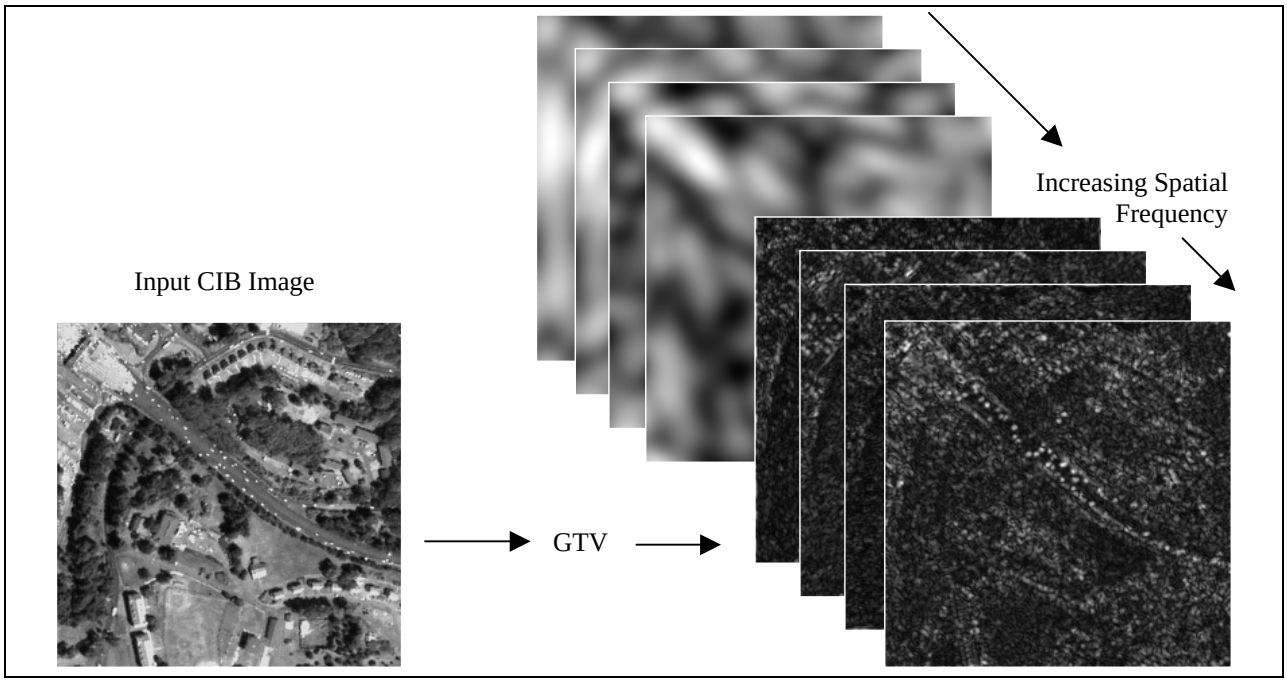
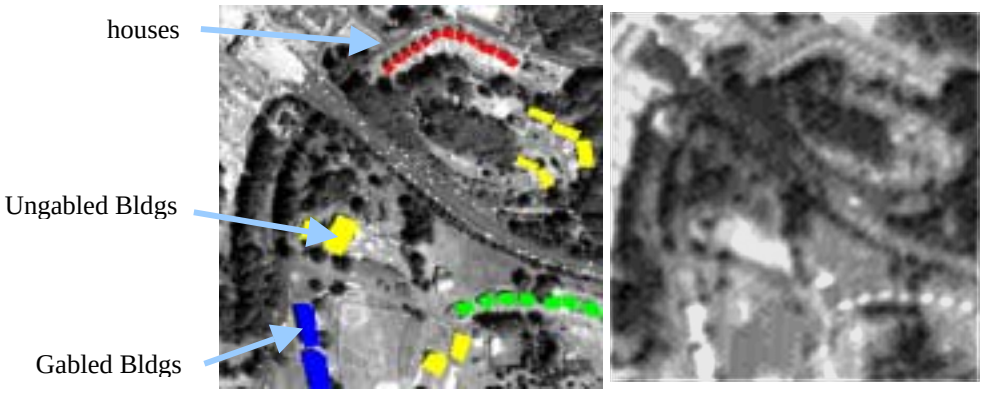


Figure 13. GTV Channel Filter Outputs:
[Showing the 4 orientation outputs for the lowest and highest spatial frequencies.]



CIB (1m) IRS (approx. 6m)

Figure 14A. CIB Input and IRS Input: Selected Objects are Highlighted

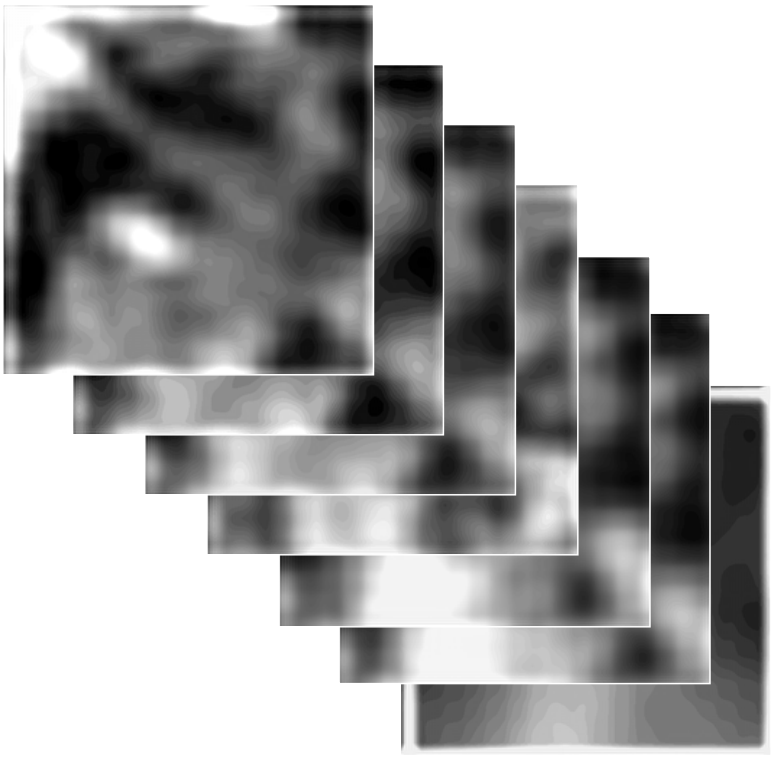


Figure 14B. Landsat Data 7-bands (6 VIS/NIR, 1 Thermal IR)

The resulting “hyper-data” cube is 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 15). [Note that Figure 15 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 were used in the discrimination step.] The discrimination of the selected objects is shown in Figure 16.

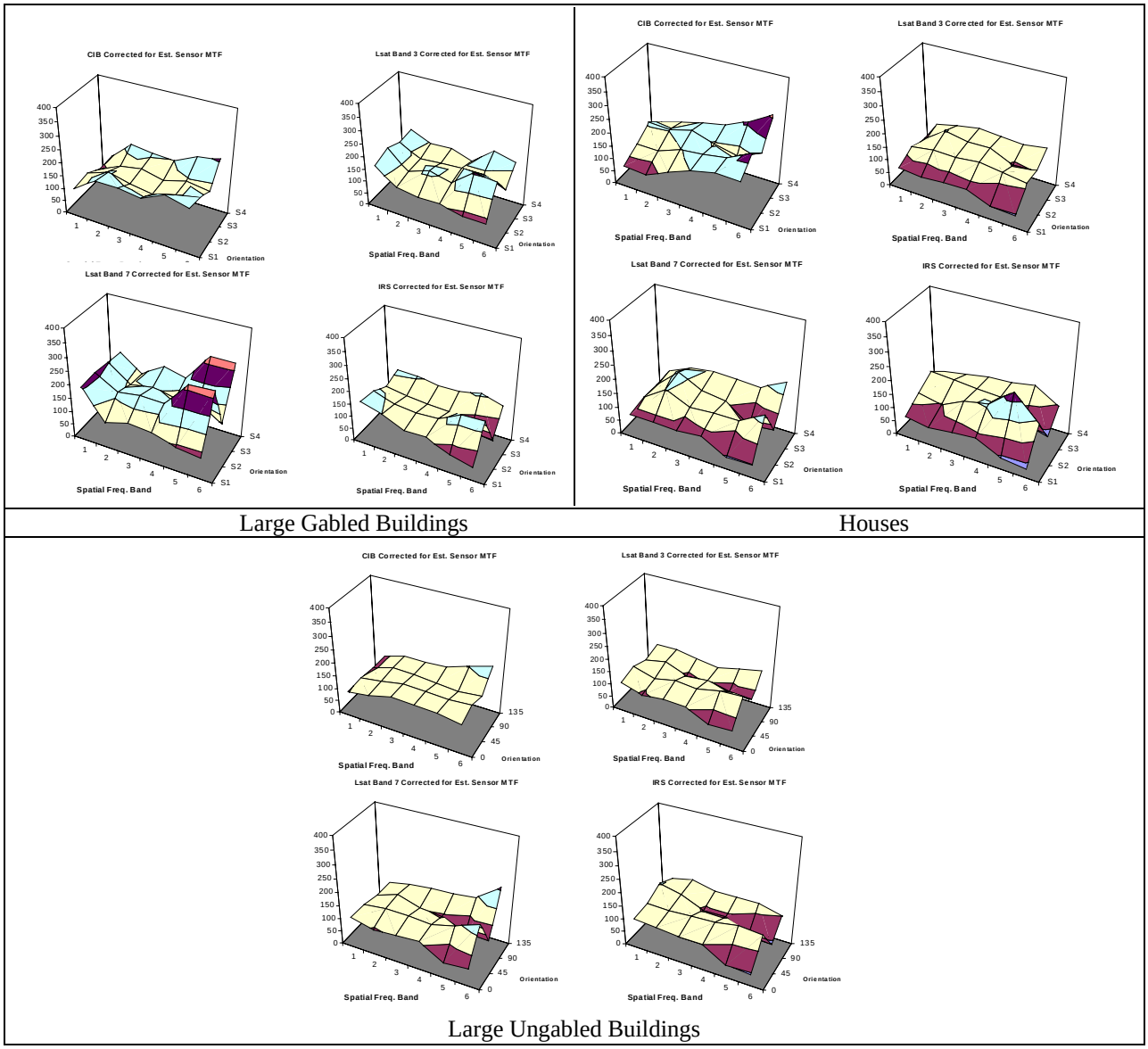


Figure 15. Spatial/Spectral Signature Surfaces for Objects of Interest:
(upper left) Surfaces for Large Gabled Buildings; (upper right) Surfaces for Houses;
(lower center) Surfaces for Ungabled Buildings
[Note: Within each quad surface plot, the upper left is from CIB;
upper right is Landsat band 3; lower left is Landsat band 7 and the lower right is IRS. Also note that there are 9 "3D surface" plots which were considered in the discriminant analysis, but are not shown in order to simplify the display.]

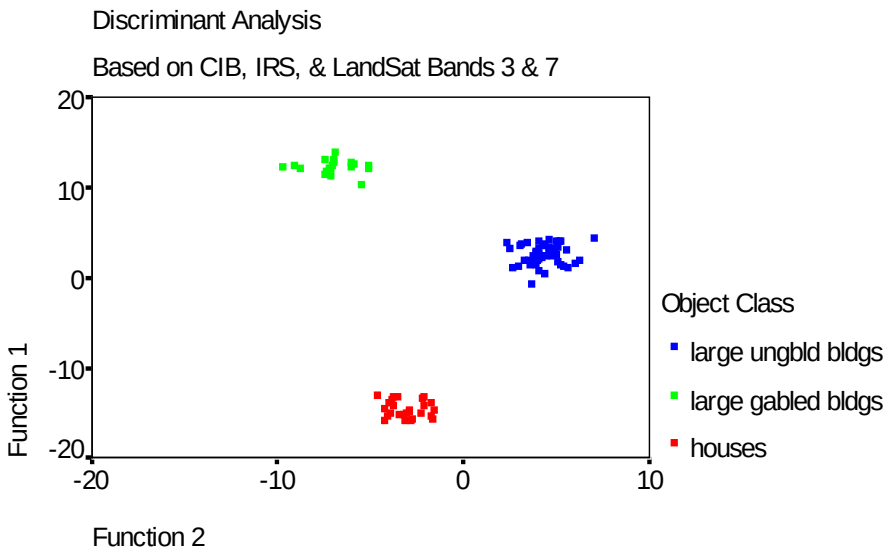


Figure 16. Discriminant Analysis Results of Selected Objects of GTV Output "Hyper-data" Cube Spatial/Spectral Signatures from 9-band CIB/IRS/Landsat Fusion

4.0 FUTURE WORK

Georgia Tech is pursuing a number of lines of research to refine, extend, and apply the integrated spatial/spectral pattern recognition tools discussed in this paper. Applications include face recognition, recognition of tumors in biomedical imagery, evaluation of image quality, and identification of features and objects in reconnaissance imagery. 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.

5.0 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.

For further information on the work presented here and on the GTV model, please contact:

Dr. Melinda Higgins or Dr. Theodore Doll
Georgia Tech Research Institute
925 Dalney Street
Atlanta, GA 30332
404-894-4971 (Higgins)/ 404-894-0022 (Doll)
melinda.higgins@gtri.gatech.edu
ted.doll@gtri.gatech.edu
http://eoeml-web.gtri.gatech.edu/tdoll/