Computational Vision Modeling for Target Detection

Dr. Grant Gerhart
Mr. Thomas Meitzler
US Army TARDEC
Warren, MI

Mr. Gary Witus
OptiMetrics Inc.
Ann Arbor, MI

ABSTRACT

The current DoD target acquisition models have two primary deficiencies: they use simplistic representations of the vehicle and background signatures, and a highly simplified description of the human observer. The current signature representation often fails for complex signature configurations and yields inaccurate detectability and marginal pay-off predictions for low signature vehicles. In addition it is not extensible to false alarms and temporal cues, and precludes applications to vehicle design guidance and diagnosis. The current human observer model is simplified to the same degree as the signature representation, and as such does not extend to high fidelity target/background signature representations

In answer to these deficiencies, we have developed the TARDEC Visual Model (TVM) that is based upon emerging academic computational vision models (CVM). Recent advances in CVM have made dramatic improvements in the understanding of early human vision processes. A model of neural receptive fields includes a generic image representation of the spatial processing characteristics for early vision cortical areas. An input image is first divided into its three color opponent components with each axis further decomposed into a set of band pass spatial frequency filters (Gabor or wavelet transform filters) with different center frequencies and orientations. Signal to noise statistics are then calculated on each channel, appropriately aggregated over all channels using signal detection theory to predict probabilities of detection and false alarm.

Present Assignment: Senior Research Physicist - GS15, Survivability Technology Center, US Army Tank-Automotive R&D Center (TARDEC), Warren, MI
Past Experience: 22 years as a TARDEC research scientist.
Degrees Held: BS Physics - Iowa State U. & Ph D -Physics - Wayne State U., Detroit, MI

Biography: Mr. Thomas Meitzler
Present Assignment: Research Physicist - GS13, Survivability Technology Center, US Army Tank-Automotive R&D Center (TARDEC), Warren, MI
Past Experience: Adjunct Professor at U. of Michigan and Henry Ford Community College (84-89) and Research Associate in the Neurology Dept at Henry Ford Hospital (86-87)
Degrees Held: BS and MS in Physics at Eastern Michigan U. and currently a Ph D candidate in EE at Wayne State U.

Biography: Mr. Gary Witus
Present Assignment: Senior Scientist, OptiMetrics Inc., Ann Arbor Michigan (92-94)
Past Experience: Engineering Chief at General Dynamics Land Systems Div. in Sterling Heights, MI (88-92) and Program Director at Vector Research Ann Arbor, MI (79-88).
Degrees Held: BS in Mathematics and MS in Industrial and Operations Engineering at the U. of Michigan, Ann Arbor MI

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **20 JUN 1994**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**
4. TITLE AND SUBTITLE **Computational Vision Modeling for Target Detection**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) **Dr. Grant Gerhart; Mr. Thomas Meitzler; Mr. Gary Witus**		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **US Army RDECOM-TARDEC 6501 E 11 Mile Rd Warren, MI 48397-5000**		8. PERFORMING ORGANIZATION REPORT NUMBER **18749**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) **TACOM/TARDEC** 11. SPONSOR/MONITOR'S REPORT NUMBER(S) **18749**
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release, distribution unlimited**		
13. SUPPLEMENTARY NOTES **Presented at the 19th Army Science Conference 20-24 June 1994, Orlando, Fl USA**		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**SAR**	**8**	

Computational Vision Modeling for Target Detection

Dr. Grant Gerhart
Mr. Thomas Meitzler
US Army TARDEC
Warren, MI

Mr. Gary Witus
OptiMetrics Inc.
Ann Arbor, MI

1. INTRODUCTION

There are many uses for accurate and realistic models describing human observer decision performance for image representations of target/background scenes. Military visual search and detection tasks are the most obvious uses, but important dual use applications exist including highway driving and flying. Good visual models have important utility in the design and evaluation of automotive passive collision avoidance technology and paint patterns for aircraft and highway barricades. In addition other applications such as machine vision and photo interpretation also require high fidelity predictive models of human vision system performance.

Current DoD target acquisition models use ad hoc signature metrics[1,2] to quantify and describe specific cue features inherent in complex imagery. These metrics typically model scene content using 1st and 2nd order statistical parameters. Figure 1 compares two vehicle thermal images and corresponding Fourier transform (FT) images. The amplitude and phase of the FT images have switched before inverse transforming back to the spatial domain. At the bottom are the two resulting images which look similar to the original except for some noise degradation. In both cases the phase dominates in importance over the amplitude information as evidenced by the primary cue features which originate from the corresponding phase image. Ad hoc signature metrics typically use parameters such as the mean ΔT and standard deviation of the target and local background. These metrics do not contain any phase information and therefore cannot be accurately used to predict human performance.

Figure 2 compares the performance of the mean ΔT and root mean square (RMS) metrics for a standard thermal scene of a M60 tank. Each column contains a series of images each filtered by a one octave band pass filter with a center frequency FS corresponding to feature sizes of 4, 8, 16, etc. pixels. The right hand column contains the original scene with no target present. The next column contains both the original target and background with no modifications. The two columns on the left contain modified imagery where the target image is modified relative to the background by performing a correction which subtracts either the mean or RMS ΔT value from each pixel. The modified target should be indistinguishable from the background if in fact either metric provides a good measure of target/background matching.

A good match occurs for feature sizes (i.e. 32 pixels) approximately the size of the target dimensions. The target virtually disappears for the 32 pixel feature size filter while becoming quite visible for the other high spatial frequency bandpass filters. Particularly note the high contrast horizontal edges which become quite visible at the higher spatial frequencies. It is evident that a simple metric description of the target/background scene is not sufficient to eliminate complex cue features.

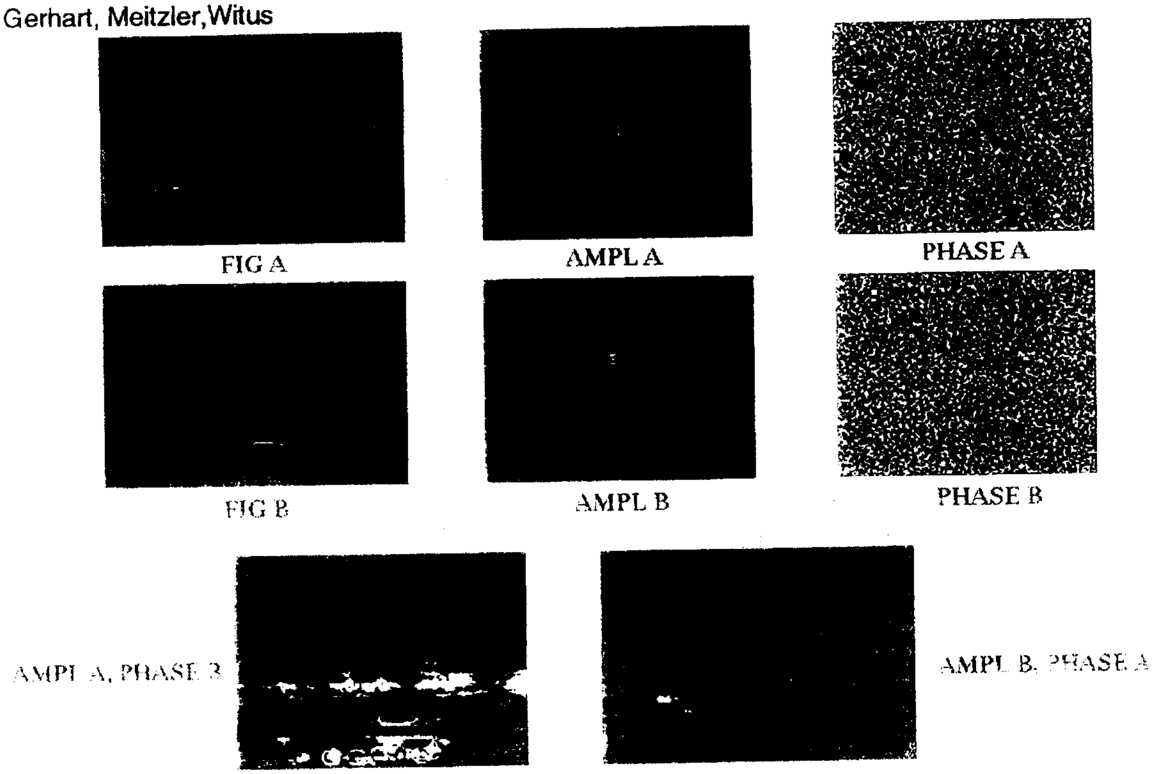

Figure 1 Switching the amplitude and phase of the Fourier transform of two images

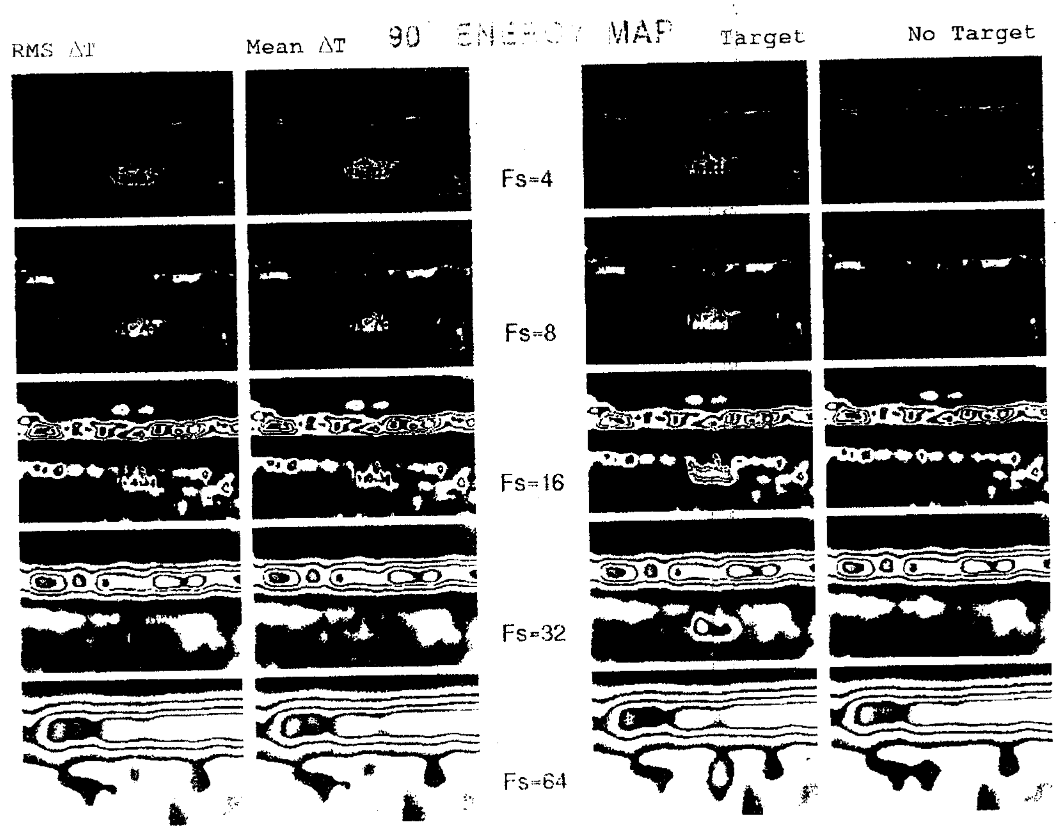

Figure 2 A comparison of the mean ΔT and RMS signature metrics different band pass filters

2. TARDEC VISUAL MODEL (TVM)

This section will describe a more complex, but robust model of the human vision system. The human vision system begins with the retina, the early vision channels and ends with higher level image understanding. The early vision system[3,4,5] consists of a set of parallel channels which analyze image characteristics such as motion, color, spatial frequency content, orientation, etc., by a complex mapping of the photoreceptor cells in the retina through the lateral geniculate nucleus onto the visual cortex. The higher level image understanding is not currently well understood and additional research needs to be done to establish a predictive model. The research community has, however, done a good deal of work in developing computational models of the early vision system.

The TVM model assumes that defeating the early vision system is sufficient to defeat the human vision for target detection applications relating to low signature vehicles. Figure 3 outlines the TARDEC Computational Vision Model (CVM) starting with the color opponent black-white, red-green and yellow-blue decomposition. The next step consists of a spatial frequency decomposition of each color opponent channel into one octave wide band pass filters with different center frequencies and orientations. The model defines a signal to noise ratio for each channel and aggregates these over all channels to define d' human performance parameter. Signal detection theory uses the d' parameter to predict single glimpse probabilities of detection (P_d) and false alarm (P_{fa}). The model includes both fovea vision and peripheral off axis eccentricity effects in calculating d'. A search model then aggregates the single glimpse model for calculating time dependent P_d's and P_{fa}'s.

2.1 TVM Signature Vector

The human visual system[5] samples contrast gradients with a series of Gaussian filters approximating directional spatial derivatives. Each filter performs a band pass operation in one spatial dimension and a low pass operation in the orthogonal direction. The filters are implemented sequentially with a simple three pixel kernel in each of two directions. The set of band pass filters are implemented at six distinct center frequencies differing by a factor of two in spatial frequency, and they are implemented as a pyramidal hierarchy of filters in which the image input to the next lower spatial frequency is obtained as a residual of the filtering operations from the next higher spatial frequency.

A typical configuration is a set of 36 band pass filters consisting of three color opponent channels each divided into two orientations with six center frequencies. Each filter output is an image which contains some particular subset of the original image content and cue features. A measure of image content is the "contrast modulation energy" which is the target outline projected onto each image with the average energy computed over the target area and its immediate local background. Figure 4 illustrates the process which is remarkably similar to conventional amplitude modulation signal detection. To extract the "energy envelop function" from the contrast modulation function one first squares the original signal followed by a low pass operation for each band pass filter.

The difference in mean energy and variance between the target and background are two measures of the contrast modulation function. A single signature metric combines these two measures for each band pass filtered image using the following mathematical operations:

$$\Delta\, SIGNAL = \sqrt{\left|\sigma_{tgt}^2 - \sigma_{bkg}^2\right| + \left[\mu_{tgt} - \mu_{bkg}\right]^2} \qquad (1)$$

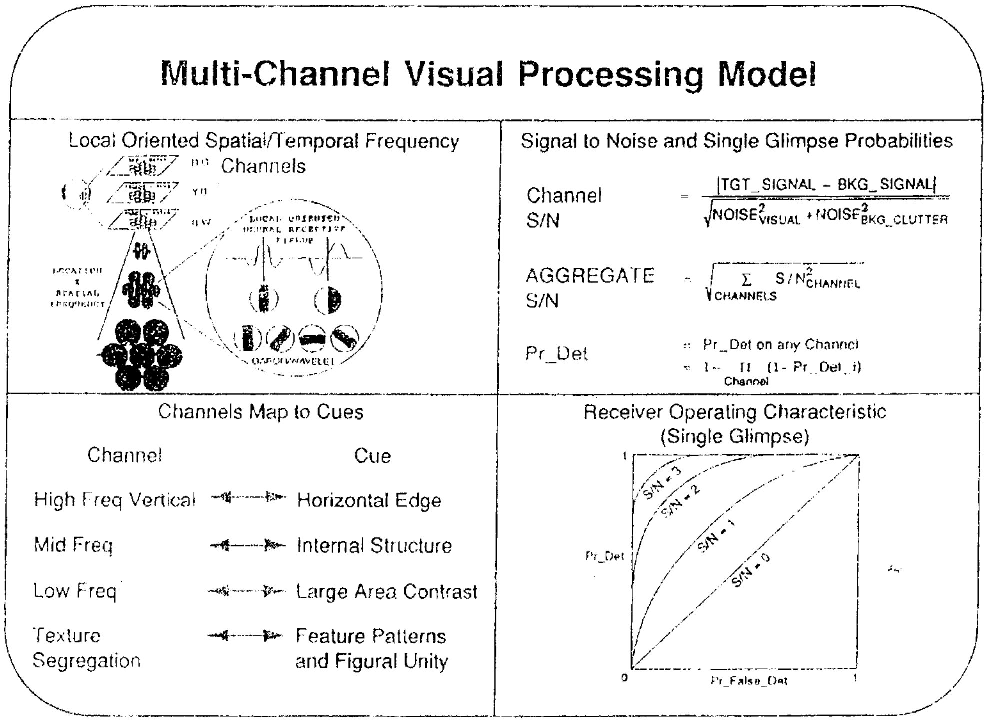

Figure 3 Flow chart for the TVM model

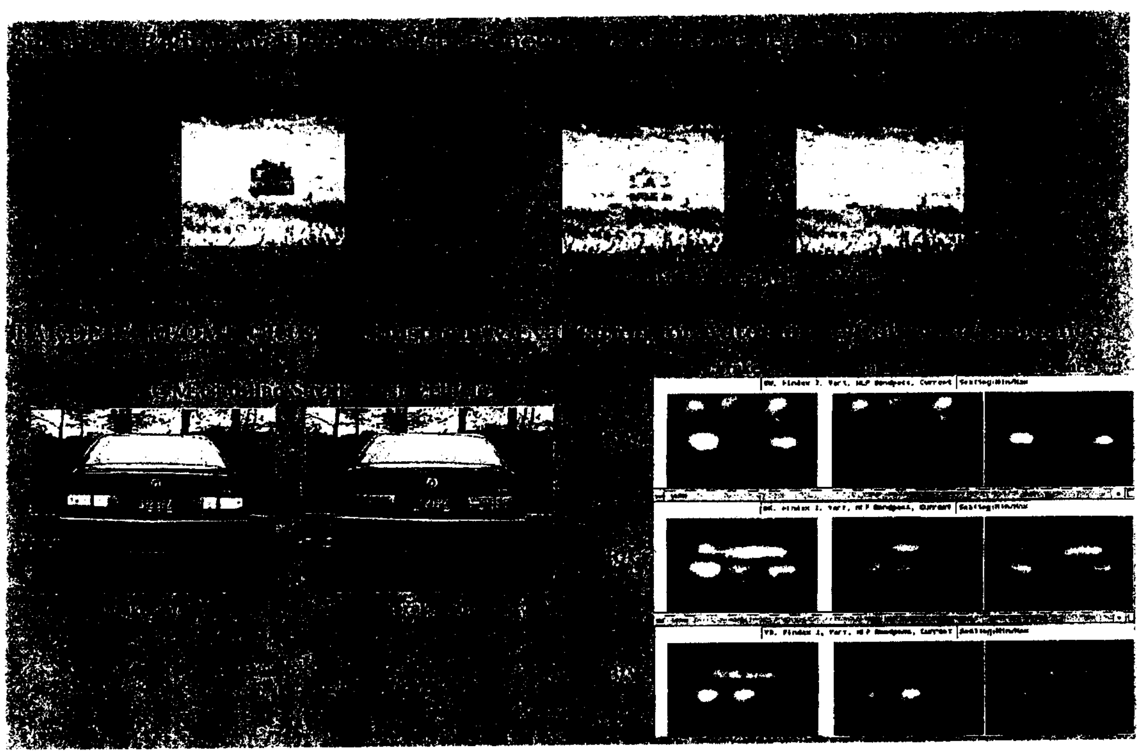

Figure 5 Dual use applications of the TVM model

$$= \sqrt{\left|\textit{Change in Energy Variance}\right| + \left(\textit{Change in Energy}\right)^2} \quad (2)$$

$$\textit{Noise} = \sqrt{\textit{ICN noise power} + \textit{BKG noise power}} \quad (3)$$

Equations 1-3 define the necessary parameters for a single channel signal to noise ratio (S/N). The noise term includes internal eye noise which is a function of illumination level and a clutter noise term which is estimated from the background statistics of the contrast modulation energy.

The signature vector consists of 36 S/N ratios - one for each channel output. A single "detectability" metric weights each signature vector component by the number of receptive fields for that particular band pass filter. Next the model "pools" these elements using a cortical model by Watson which includes the density of receptive fields sensitive to any spatial frequency inversely proportional to the eccentricity distance from the fovea. Equations 4-6 contain a mathematical description of the process where the c,a,f indices refer to a specific signature vector component with a particular color opponent component, orientation, and center frequency. The quantity d is the "detectability" metric which has a log-linear relationship to d'. The latter predicts a particular ROC curve (i.e. Fig. 3). The exponent QNSR is approximately two corresponding to an ideal observer model for Signal Detection Theory.

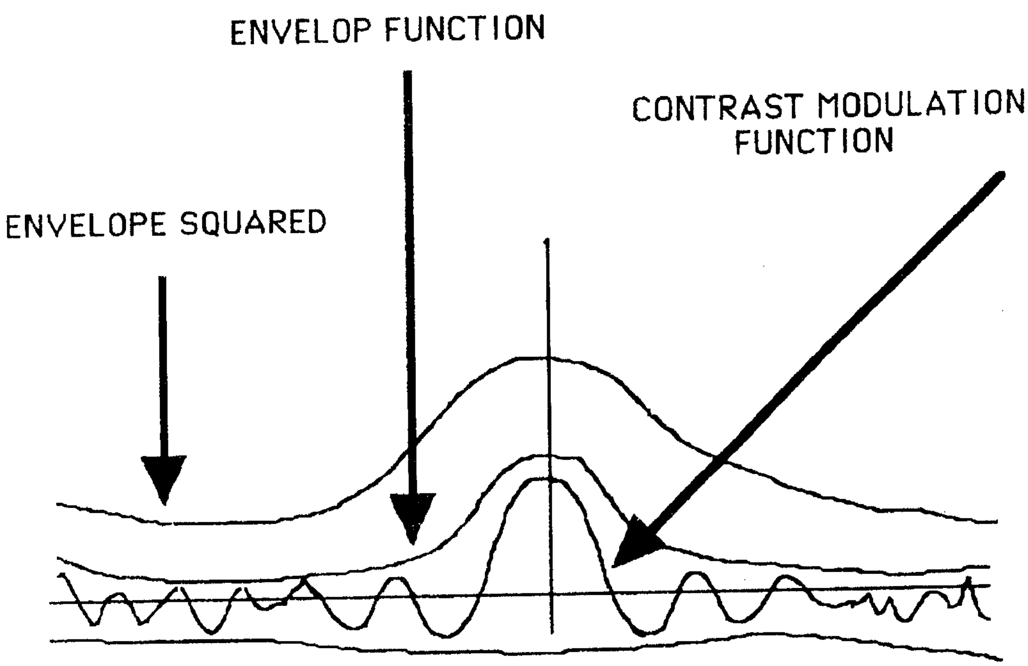

Figure 4. The process for extracting the energy envelop function from a band pass filter contract modulation function.

$$SNR_{c,a,f} = \frac{\Delta' signal_{c,a,f}}{Noise_{c,a,f}} \tag{4}$$

$$d(\varepsilon) = \sum_{c,a,f} NRF_{c,f}^{QSNR}(\varepsilon) SNR_{c,a,f}^{QSNR} \tag{5}$$

$$d'(\varepsilon) = \psi_1 d(\varepsilon)^{\Psi_2} \tag{6}$$

3. TVM CALIBRATION TO OBSERVER PERFORMANCE

TVM uses a distribution of observers because individuals have different thresholds corresponding to varying probabilities of detection and false alarm. The parameters Ψ_1 and Ψ_2 are empirical and are calibrated using observer field test data. The pilot study used 10 military observers viewing six sets of images, representing six separate target scenes. The imagery consisted of low contrast target/background scenes near detectability thresholds which were presented 20 times to each observer in random order.

The hypothesis that the Ψ_1 and Ψ_2 parameters are approximately constant for the observer population was well supported by the data where $\Psi_1 = 3.0 \pm 0.75$ and $\Psi_2 = 0.12 \pm 0.016$. Though a good deal of additional calibration testing needs to be done, the preliminary results are quite encouraging because the values for Ψ_1 and Ψ_2 were approximately constant for several observers over several thousand observations.

4. SEARCH

Since the pilot study used observer tests where the subject viewing times were relatively short, the model produces single glimpse probabilities of detection and false alarm. Time dependent detection and false alarm probabilities are obtained by aggregating multiple glimpses through some particular search strategy. The search process can be modeled in several ways depending upon one's particular knowledge of the search process. If eye tracking data is available, for example, then a search strategy can be devised which is a function of the probability of eye fixation and its position in the scene. For combat models, however, this type of information is usually not appropriate and other types of search models must be used. The TVM model uses a Markov process which consists of three transient states: (1) the observer is cued to a target, (2) the observer is cued to a background object, and (3) the observer is not cued to any specific scene object. The result is a set of transition rates for detection, false alarm and quitting, involving single glimpse probabilities, eye integration time, the probability of the target being in the field of view and various other integral expressions over the search field of regard.

5. TVM APPLICATIONS

The original model development objective was to develop better ground vehicle requirements and specifications for low signature vehicles and camouflage, concealment, and deception (CCD) applications. Previous acquisition models for man in the loop imaging systems used a series of ad hoc signature metrics which averaged over an ensemble of targets and consequently are not sensitive to specific target cue features and characteristics. Figure 5 illustrates the successful application to the Requirements Translation problem. The top portion of shows three images of a high contrast vehicle in a color scene. The two images in the upper right show different levels of signature suppression as a function of different S/N ratios using the computational vision model outlined in the previous text. Note that when the S/N=0, very little residual signature is left indicating that the multi-channel band pass filter model captures nearly all of the relative target/background difference information. Similar comparisons using conventional ad hoc signature metrics such as contrast ratios for visible imagery or ΔT for thermal imagery would not work nearly as well.

The bottom portion of Fig. 5 illustrates a dual use application of TVM to passive collision avoidance. This work is part of a joint TARDEC/GM CRDA to develop various metrics for evaluating the conspicuity of automobiles in collision avoidance applications. Note the difference between the on/off brake light configurations for the different color opponent channels with a specific high spatial frequency band pass filter sensitive to horizontal edges and the subsequent large contributions to overall conspicuity made by the Cadillac Seville "third tail light" which consists of a horizontal strip of red LEDs. The left hand column shows the three color channels with the bake lights on, the second column with the brake lights off and the third column is the difference between the two conditions. GM statistics show that automobile collisions caused $135 billion damage exclusive of pain and suffering in FY90. Passive collision avoidance technology is much cheaper than its active counterpart. It is quite effective as evidenced by the success of the "third tail light" which is estimated to have reduced collision accidents by at least 3% resulting in a several billion dollar savings to the American consumer. The joint effort with GM begins in FY94 and will develop and validate conspicuity metrics based upon elderly driver test data for intersection scenarios. Elderly drivers represent a particularly high risk category in this particular scenario. TARDEC will provide a dual use facility for conducting the observer experiments to develop conspicuity metrics and GM will make available its extensive automotive proving ground and test facilities for final model validation and accreditation. Additional dual use applications include photo interpretation, machine vision and automatic object recognition.

SUMMARY

The first phase of the TVM model development program will be completed at the end of FY94 including final documentation and validation for both pop-out and low signature targets. The model will treat both stationary and dynamic targets in complex background scenes, analyze color scenes, provide a visualization module (i.e. Fig 5), and have an acquisition module for insertion into the combat models. A follow on phase beginning in FY95 will incorporate higher levels of target/background discrimination for IFF applications and object recognition.

BIBLIOGRAPHY

1. G.R. Gerhart, "Target Acquisition for Visual and IR Countermeasures," Proc. 29th IRIS Symposium on Infrared Countermeasures (Session D6) (1990).
2. G.R. Gerhart, E. Lapinski, and S. Russell, "Visual Target Acquisition Models," Proc. 2nd Ground Target Modeling and Validation Conference," TACOM Technical Report #13553 (1991).
3. *Computational Models of Visual Processing,* M.S. Landy and H.A. Movshon, eds., The MIT Press, Cambridge, 1991.
4. L.A. Olzak and J.P. Olzak and J.P Thomas, "Seeing Spatial Patterns," in the *Handbook of Human Performance,* 1986.
5. A.B. Watson, "Detection and Recognition of Simple Spatial Forms," *NASA Technical Memorandum 84353,* NASA Ames Research Center, Moffett Field, 1983.