Benefits of photosimulation and sensor fusion for threat detection

E. Bankowski, Ph.D., D. Bednarz, Ph.D., D. Bryk, R. Jozwiak,
K. Lane, T. Meitzler, Ph.D., E.J. Sohn and J. Vala

U.S. Army TACOM
Survivability Technology Area
Warren, MI 48937

ABSTRACT

Visible, infrared (IR) and sensor-fused imagery of scenes that contain occluded camouflaged threats were compared for hit and miss differences in the probability of detection of objects. Response times were also measured. Image fusion was achieved using a Gaussian Laplacian pyramidal approach with wavelets for edge enhancement. Three types of images were also ranked in terms of better probability of detection of concealed weapons. Detecting potential threats that are camouflaged or difficult to see is important not only for military acquisition problems but, also for crowd surveillance as well as tactical use such as on border patrols. Imaging and display technologies that take advantage of photosimulation and sensor fusion are discussed in this paper.

I. BENEFITS OF SIMULATION IN VISUAL PERCEPTION LAB.

A method is described for using the TACOM photosimulation laboratory environment to detect threats and to evaluate the effectiveness of camouflage for military vehicles. There are distinct advantages to acquiring images at the field site and then bringing them back for observer testing in a laboratory environment. Laboratory testing provides a repeatable, secure, and low-cost way to generate realistic performance data for threat detection and vehicle evaluation for the purposes of signature testing, measurement of the effectiveness of camouflage relative to a baseline vehicle, and calibration and validation of target acquisition models.

Three tests are described by the authors. In the first test a baseline LAV is compared to a treated LAV in the TARDEC Photosimulation Laboratory using imagery collected from the field in the manner prescribed by an experimental design. Two tests were performed for RPG detection for homeland defense. Twenty five civilian subjects took the Rocket Propelled Grenade (RPG) detection test. The first RPG detection test consisted of 116. There were images of soldiers holding the RPG. There were also images of soldiers without RPG. The test was done on a flat panel monitor. There were three types of images: visual, IR and fused. The probability of detection (PD) of the RPG was estimated as a ratio of correct answers to the total number of pictures. The response time was also measured. The objective was to see the relationship between the probability of detection and the entropy metric. The second RPG detection test was ranking three types of images: visual, IR and fused. The test subjects were shown 32 sets of three images: visual, IR and fused (the total number of images was 96). The RPG was present in every picture. They had to rank images which image was better in terms of detecting the RPG. Using high- resolution graphics projectors, the imagery can be presented in the controlled environment of the lab in such a manner as to obtain observer data with confidence levels approaching 99%. It is the authors' opinion that, the benefits of the high degree of confidence achieved using the repeatability offered by the lab environment, far outweighs any so-called 'realistic' advantage of having multiple teams of observers present at a field location for threat detection and vehicle testing.

II. PREVIOUS LAV-25 TEST.

1. Introduction. The Light Armored Vehicle (LAV) Family of Vehicles (FOV) was developed to provide the Marine Corps with enhanced mobile warfare capabilities. The LAV FOV includes several variants which utilize light armor protection from small arms, light machine gun fire, artillery projectile fragments, and mine fragments. Each variant was designed for a specific mission function and was mounted on a common chassis.

2. Experiment. The test design implemented was an extension of visual detection requirements provided to our lab from PM LAV. Initially the PM requested only two range points at standard engagement ranges. We suggested having more ranges in between the critical range points to obtain a probability of detection versus range curve, as is more typical for these kinds of tests. A test matrix was developed in full-factorial form and 24-bit color imagery was collected using a Kodak 460 digital camera. The images were prepared for the photosimulation test and then presented to 30 subjects. The experimental factors and levels with their values are shown below in Table 1.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
22 AUG 2003	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Benefits of Photosimulation and Sensor Fusion for Threat detection	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
E. Bankowski, Ph.D.; D. Bednarz, Ph.D.; D. Bryk; R. Jozwiak; K. Lane; T. Meitzler, Ph.D.: E.J. Sohn; J. Vala	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
USA TACOM 6501 E 11 Mile Road Warren, MI 48397-5000	**13954**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	TACOM TARDEC
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**SAR**	**14**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

The photosimulation test in the lab was arranged so that a pixel IFOV subtended by the display was less than one minute of arc and the displayed image represented a unity magnification or 1X representation to the subject. The first test was meant to emulate naked-eye vision. Prior to the actual test, the subjects were instructed on the purpose of the test as well as required to take a pre-test in which they could become familiar with the imagery and software. None of the pictures used in the pre-test were used in the actual test, however, the images were from the same set. The test procedure was to display an image with a time-out of 30 seconds. The imagery is cropped so that no scrolling is required. The target can appear within one of five possible regions. The soldier must use the mouse to "click-on" what her or she thinks is a target, based on the training. The tests are done in a dark room in which the subjects are 'dark-adapted' to maximize contrast differences in the images.

3. Results. Analysis of the first test results has shown that most subjects obtained a score of only 20 % detection. This is not unreasonable given the difficulty of the imagery. The ranges are not unusual for such a test, however the high degree of clutter and in particular the height of the grass on the terrain makes it difficult for the unaided eye to detect common cure features of the vehicle. A second test was arranged at a power of 3X. The imagery from the field was of sufficient resolution so that there was no noticeable increase in pixilation of the imagery an in increase in magnification. The presentation in the lab was randomized, this is a very good reason to use the lab.

Table 1: Factor matrices for the visual detection test

Region	
1	Top-Left
2	Top-Right
3	Lower-Left
4	Lower-Right
5	Center
Vehicle Type	
1	Baseline (old LAV)
2	SLEP + ADCAM
3	SLEP + ADCAM - ADCAM bowplane
Aspect angle	
1	Front
2	30 degree
3	Side
Lighting	
1	Front Lit
2	Back Lit
Weather condition	
1	Clear
2	Overcast
Range (km)	
1	1
2	1.5
3	2
4	2.5
5	3

When making inferences about differences in a factor in a perception experiment in the laboratory we want to make the experimental error as small possible. This requires that we remove the variability between subjects

from the experimental error. The design we use to accomplish this is a factorial experiment run in a randomized complete block. By using this design with the subjects as blocks we form a more homogeneous experimental unit on which to compare different factors. This experimental design improves the accuracy of the comparisons among the different factors by eliminating the variability among the subjects. Within a block, the order in which the treatment combinations are run is randomly determined. It is usually not possible to implement this experimental design in the context of a traditional field test.

The pictures below in Fig. 1 through Fig. 6 were used for training observers as to what kind of vehicles they would be looking for in the test.

Fig. 1: Baseline side

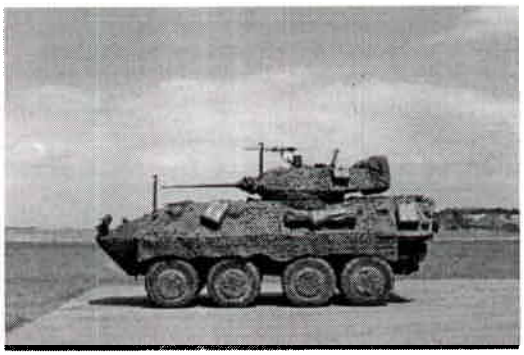

Fig. 2: ADCAM side

The charts in Fig. 3 below show the results of measuring the X and Y chromaticity values of the monitors that were used in the test. The values measured were compared to standard values and found to be virtually identical.

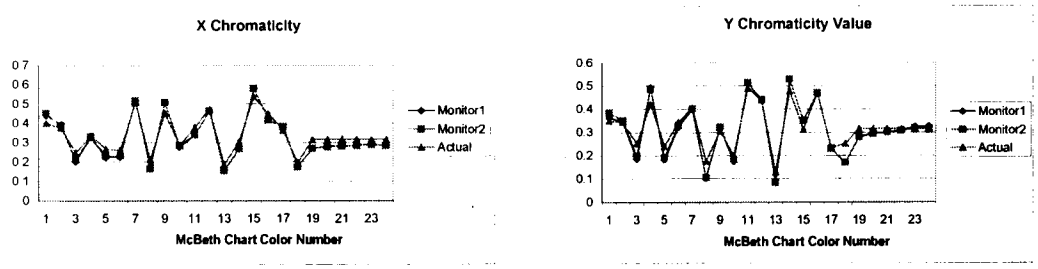

Figure 3: CIE X and Y monitor chromaticity calibration charts

4. Analysis. Below in Table 2 is the ANOVA table for the baseline vehicle and the other experimental factors. The treated vehicle has been excluded because of security classification. The power of the experimental design methodology is shown here in that the significance of individual factors and of their interactions are available. Using this kind of a test, one can obtain not only a model curve of the detection probability versus any factor in the test, but, one can also obtain the relative importance of the individual factors.

Table 2 shows that the aspect angle was the least important factor in the experiment. Figure 4 shows the model generated logistic curve of the probability of detection versus the distance from LAV, or range. The model shows a good fit.

Table 2: ANOVA of test factors

Tests of Between-Subjects Effects

Dependent Variable: RANK of RESPONSE

Source	Type IV Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Corrected Model	185215009[b]	89	2081067.514	26.782	.000	2383.607	1.000
Intercept	1159759015	1	1159759015	14925.408	.000	14925.408	1.000
SKY_COND	2918024.068	2	1459012.034	18.777	.000	37.553	1.000
RANGE	161301308	9	17922367.58	230.650	.000	2075.852	1.000
ASPECT	944347.896	2	472173.948	6.077	.002	12.153	.887
SKY_COND * RANGE	5751990.053	18	319555.003	4.112	.000	74.025	1.000
SKY_COND * ASPECT	2459473.720	4	614868.430	7.913	.000	31.652	.998
RANGE * ASPECT	6204854.010	18	344714.112	4.436	.000	79.853	1.000
SKY_COND * RANGE * ASPECT	5397720.861	36	149936.691	1.930	.001	69.465	1.000
Error	128521871	1654	77703.671				
Total	1641367780	1744					
Corrected Total	313736880	1743					

a. Computed using alpha = .05

b. R Squared = .590 (Adjusted R Squared = .568)

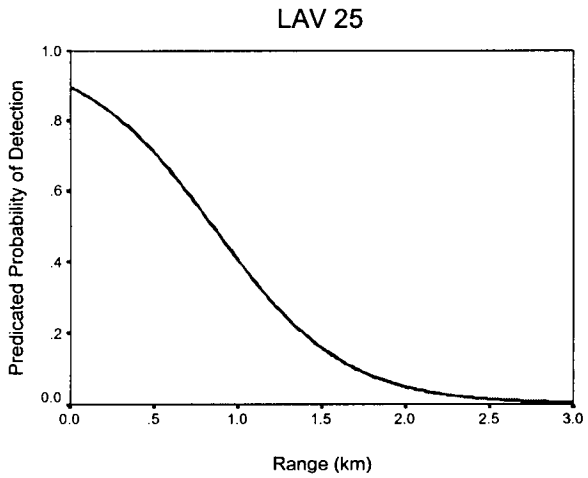

Figure 4. Logistic curve fit to the model from the subject responses

Figure 5 shows the plot of the entropy versus the probability of detection.

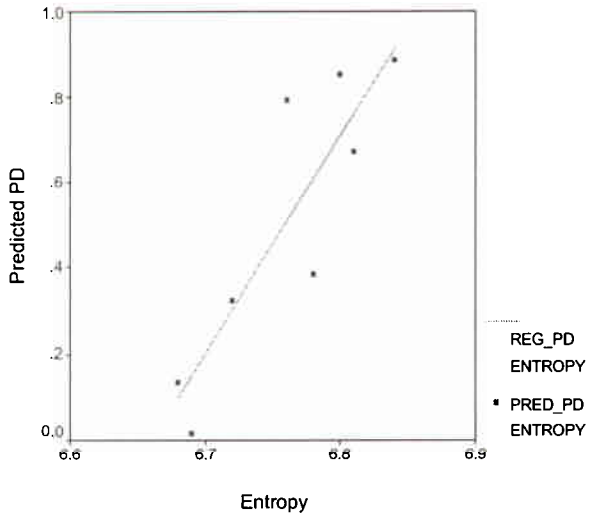

Figure 5. The plot of entropy versus the probability of detection.

There was a correlation between the entropy and the probability of detection, as it was observed in the LAV test.

III. THE RPG DETECTION TEST FOR HOMELAND DEFENCE.

1. Experimental Test Procedure.

Twenty five civilian subjects were picked up randomly from the employee population of TARDEC. Each subject took the Rocket Propelled Grenade (RPG) detection test, and ranking the quality of images test.

Test No. 1 was the RPG detection test. It consisted of 116 images in a random order. There were images of soldiers holding the RPG. There were also images of soldiers without RPG. The test was done on a flat panel monitor. There were three types of images: visual, IR and fused. Some images had brightness, contrast and noise adjusted. Each image was shown for one second. After viewing each image for one second, the test subject was asked to click "Yes" or "No", depending whether he could see the RPG in the picture, then go to the next image. The probability of detection (PD) of the RPG was estimated as a ratio of correct answers to the total number of pictures. The response time was also measured. The objective was to see the relationship between the probability of detection and the entropy metric. Figure 6 shows three identical images of men with the RPG: the image on the left side is the visual image, the image in the middle is the infra-red (IR) image and the image on the right is the fused image.

Figure 6. Images of men with the RPG: visual image, IR image, and fused image.

Test No. 2 was ranking three types of images: visual, IR and fused. The test subjects were shown 32 sets of three images: visual, IR and fused (the total number of images was 96). The RPG was present in every picture. They were asked the question: "Which image is better in terms of detecting the RPG?" They were asked to rank images in the set as # 1, 2 and 3. The noise, brightness and contrast levels in some images were adjusted to make it harder to detect a threat. Figure 7 shows three identical images of men with the RPG; the noise level in these images was adjusted. The image on the left side is the visual image, the image in the middle is the infra-red (IR) image, and the image on the right is the fused image.

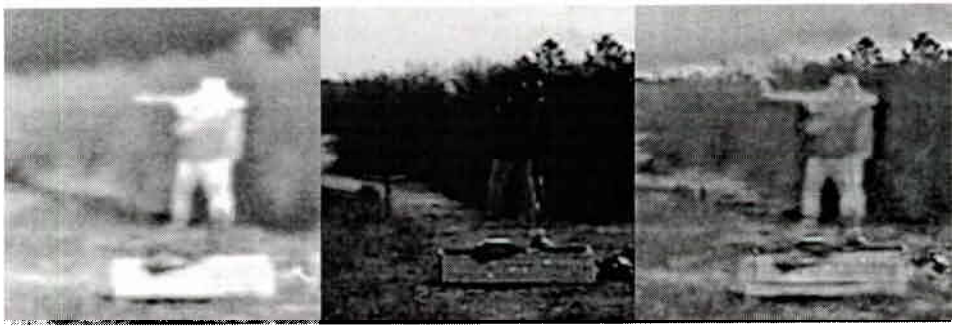

Figure 7. Images of men with the RPG with the noise level adjusted: visual image, IR image, and fused .

Figure 8 shows three identical images of men with the RPG; the brightness level in these images was adjusted. The response time was also measured.

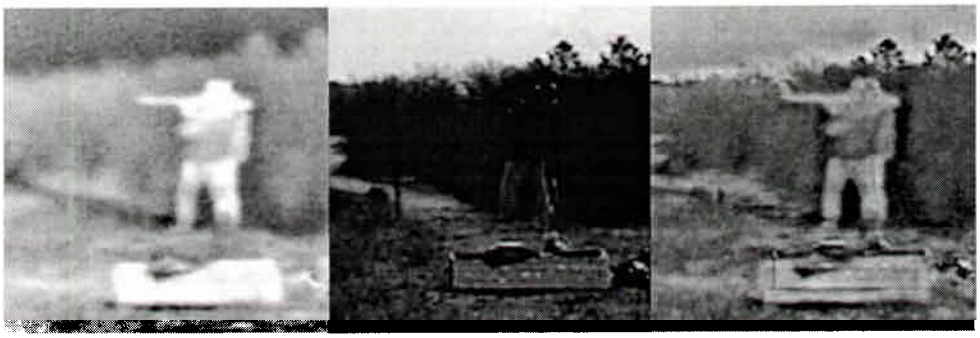

Figure 8. Images of men with the RPG with the brightness level adjusted: visual image, IR image, and fused image.

Figure 9 shows three identical images of men with the RPG; the contrast level in these images was adjusted. The objective was to see the relationship between the probability of detection and the entropy metric.

2. Summary of Test Data.

The statistical results of the RPG test were analyzed. Figure 10 shows the detection rate of the RPG in the image versus the sensor type. As we can see from Figure 11, the highest detection rate was achieved using the IR sensor, and the lowest detection rate was with the visual image. The test subjects also commented on the fact that they preferred the fused image, because it had more details.

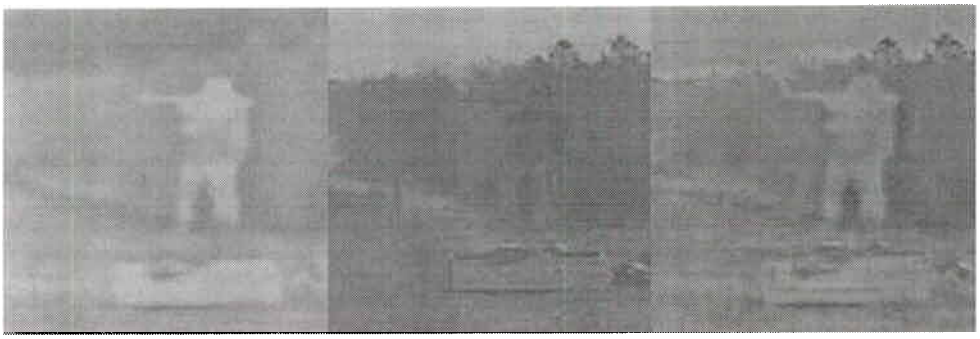

Figure 9. Images of men with the RPG with the contrast level adjusted: visual image, IR image, and fused.

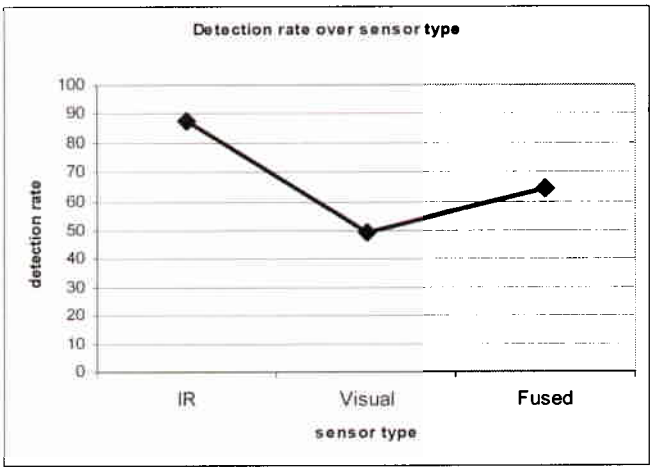

Figure 10. The detection rate of the RPG in the image versus the sensor type.

The results of test No.2 are shown in Figure 11.

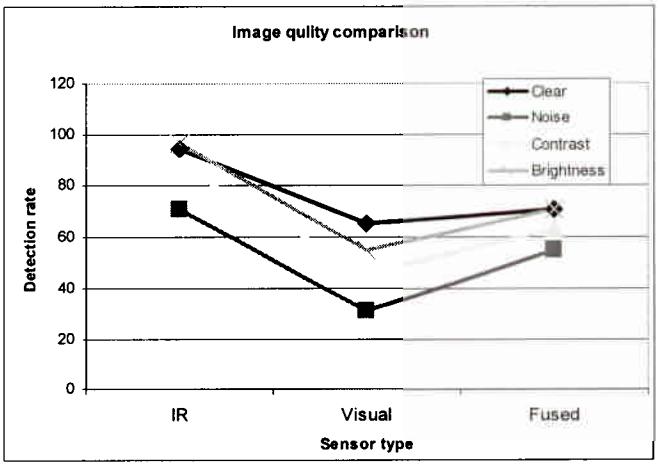

Figure 11. Image quality comparison: detection rate versus the sensor type.

Test No. 2 was ranking three types of images: visual, IR and fused. The test subjects were sets of three images: visual, IR and fused. They were asked the question: "Which image is better in terms of detecting the

RPG?" They were asked to rank images in the set as # 1, 2 and 3. The noise, brightness and contrast levels in some images were adjusted to make it harder to detect a threat.

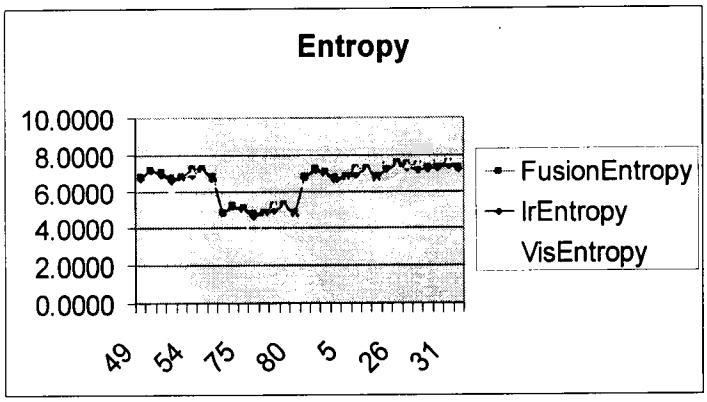

Figure 12. The entropy of the image versus the image type.

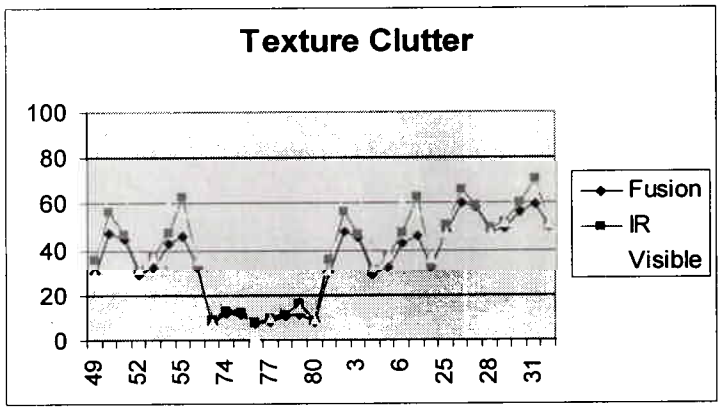

Figure 13. The texture clutter versus the image type.

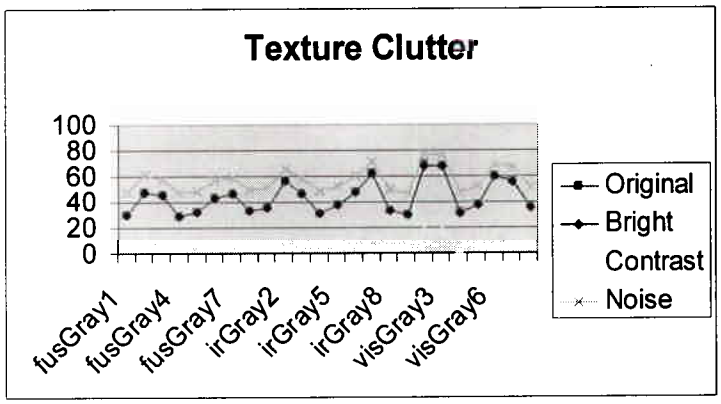

Figure 14. The texture clutter versus the image type.

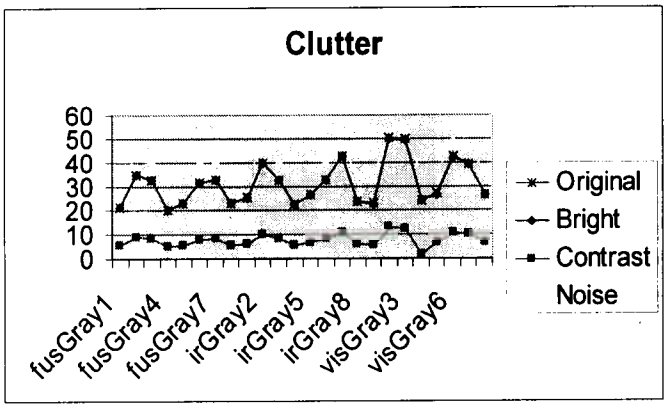

Figure 15. The clutter versus the image type.

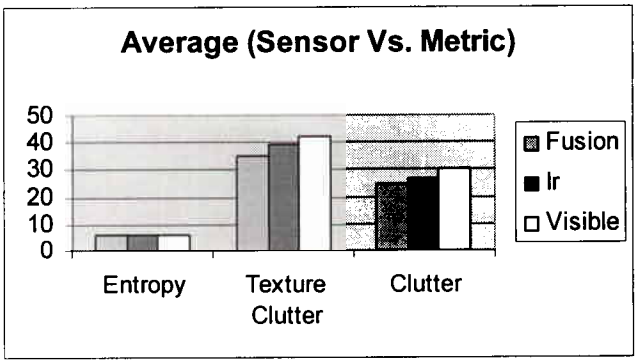

Figure 16. The average sensor versus metric.

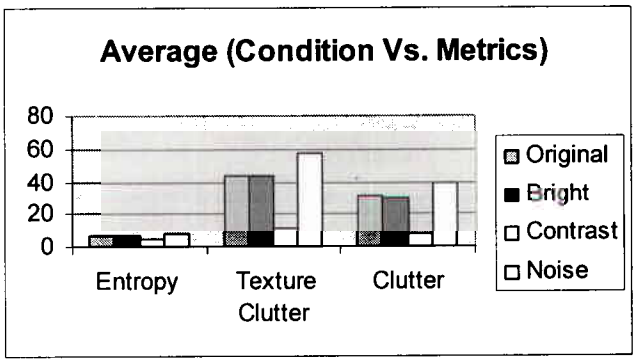

Figure 17. The average condition versus metric.

Table 3 below shows a good correlation of entropy of image with detection, as well as good correlation of texture clutter and clutter with detection. As we can see from Column 4 at the bottom of Table 3, the correlation coefficients are close to 1 or to 100%.

Table 3. Correlation of entropy with detection; correlation of texture clutter and clutter with detection.

Shooter	Entropy	Texture Clutter	Clutter	% DET
Shooter 1	6.37	31.61	22.15	49.9
Shooter 2	6.68	46.60	33.03	80.8
Shooter 3	6.66	46.80	32.75	91.5
Shooter 4	6.25	30.44	21.40	46.5

	Column 1	*Column 2*	*Column 3*	*Column 4*
Column 1	1.000			
Column 2	0.983	1.000		
Column 3	0.983	1.000	1.000	
Column 4	0.962	0.982	0.977	1.000

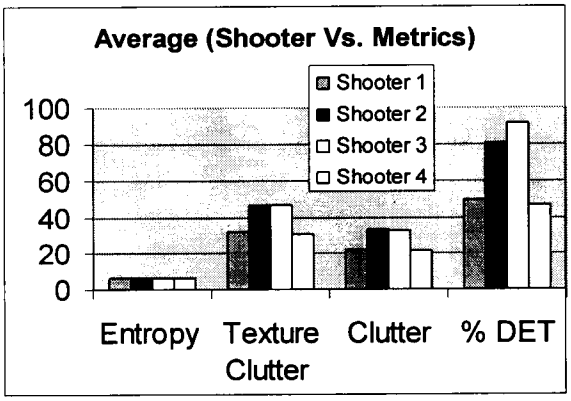

Figure 18. Correlation of entropy with detection; correlation of texture clutter and clutter with detection.

3. Data Analysis

We wish to investigate whether there are differences in sensor type effect the ability to detect threats. Two factors that will be considered will be sensor type shooter profile. An experiment was designed using three levels of sensor and four levels of shooter. The level of effect was held at four, time at one, and threat at one. The response variable is whether the subject correctly identified a shooter with an RPG. A rank transformation was applied to the response variable. This will be our dependent variable. Once a subject is chosen, the order in which the twelve treatment combinations are randomly determined. Furthermore, subjects differ in their skill and ability to detect threats. Thus, we shall use subjects as blocks. Now, we have a 3 X 4 factorial experiment in a randomized complete block.

The linear model for this experiment is:

$$y(i,j,k) = \mu + \gamma(i) + \beta(j) + (\tau,\beta)(i,j) + \delta(k) + \varepsilon(i,j,k), \quad \text{here } i = 1,2,3;\ j = 1,2,3,4;\ k = 1,2,\ldots,30$$

Where $\tau(i)$ represents the sensor effect, $\beta(j)$ represents the shooter effect, $(\tau,\beta)(ij)$ is the interaction, $\delta(k)$ is the block effect, and $\varepsilon(ijk)$are independent and identically distributed normal error term.

The complete analysis of variance for this experiment is summarized in Table (4). Both sensor and shooter are significant. The interaction of these two factors is also significant. When the interaction is significant comparisons between the means of the factor sensor may be obscured by the interaction. One approach to this situation is to fix the factor shooter at a specific level and apply a multiple comparison test to the mean of the factor sensor at that level. The results of this analysis are summarized in Table (5). The starred values indicate pairs of means that are significantly different.

The normal probability plot of the residuals is shown in Figure (19). There is no severe indication of non-normality. The plots of the residuals versus each factor and block respectively give no indication of inequality of variance. These plots are not included in the paper. Thus, we have checked the adequacy of the assumed model.

Table 4. Tests of Between-Subjects Effects

Dependent Variable: RANK of RESPONSE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power
Corrected Model	1791720.000	40	44793.000	14.181	.000	567.225	1.000
Intercept	11728890.000	1	11728890.000	3713.147	.000	3713.147	1.000
SUBJECT	180360.000	29	6219.310	1.969	.003	57.099	.998
SENSOR	470340.000	2	235170.000	74.450	.000	148.901	1.000
SHOOTER	807120.000	3	269040.000	85.173	.000	255.519	1.000
SENSOR * SHOOTER	333900.000	6	55650.000	17.618	.000	105.707	1.000
Error	1007640.000	319	3158.746				
Total	14528250.000	360					
Corrected Total	2799360.000	359					

a Computed using alpha = .05

b R Squared = .640 (Adjusted R Squared = .595)

Table 5. Pairwise Comparisons

Dependent Variable: RANK of RESPONSE

SHOOTER	(I) SENSOR	(J) SENSOR	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference Lower Bound	Upper Bound
1	1	2	-90.000	14.511	.000	-118.550	-61.450
		3	-30.000	14.511	.040	-58.550	-1.450
	2	1	90.000	14.511	.000	61.450	118.550
		3	60.000	14.511	.000	31.450	88.550
	3	1	30.000	14.511	.040	1.450	58.550
		2	-60.000	14.511	.000	-88.550	-31.450
2	1	2	-138.000	14.511	.000	-166.550	-109.450
		3	-138.000	14.511	.000	-166.550	-109.450
	2	1	138.000	14.511	.000	109.450	166.550
		3	-7.105E-14	14.511	1.000	-28.550	28.550
	3	1	138.000	14.511	.000	109.450	166.550
		2	7.105E-14	14.511	1.000	-28.550	28.550
3	1	2	6.000	14.511	.680	-22.550	34.550
		3	6.000	14.511	.680	-22.550	34.550
	2	1	-6.000	14.511	.680	-34.550	22.550
		3	-9.948E-14	14.511	1.000	-28.550	28.550
	3	1	-6.000	14.511	.680	-34.550	22.550
		2	9.948E-14	14.511	1.000	-28.550	28.550
4	1	2	-132.000	14.511	.000	-160.550	-103.450
		3	-24.000	14.511	.099	-52.550	4.550
	2	1	132.000	14.511	.000	103.450	160.550
		3	108.000	14.511	.000	79.450	136.550
	3	1	24.000	14.511	.099	-4.550	52.550
		2	-108.000	14.511	.000	-136.550	-79.450

Based on estimated marginal means

* The mean difference is significant at the .05 level.

a Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

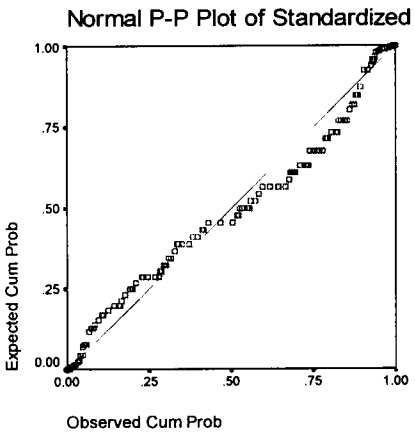

Figure 19. The normal probability plot of the residuals.

Future Research

The majority of computer programs and metrics presently used by the U.S. Army to predict the probability of detection of ground vehicles employ one of the two following strategies; 1. wavelet-like spatial frequency decomposition and color analysis of imagery based on algorithms from the literature on vision research, 2.) image metrics based on, or extrapolated from, the Johnson criteria that have been adjusted over many years to obtain closer correlation to experimental photosimulaton tests.[1,2,3] None of the present paradigms for modeling the visual detection likelihood of a particular vehicle, with or without, a stealth treatment involve a system level understanding of neural systems beyond the circuitry in the visual cortex. A problem has been how to get inside and perform experiments on the brain, and at the same time, use stimuli that are relevant to the U.S. Army. Most vision research is still done using fundamental geometric shapes that have little similarity to ground vehicles. Psychophysical experiments from subjects undergoing functional Magnetic Resonance Imaging (fMRI) [4] of the brain offer literally a new window in to how the vision system is connected to other higher processes in the brain such as, decision making, identification, and search. This new tool, it is hoped, will provide an auxiliary and enhancing technique to existing psychophysical techniques.

The image sequence below in Fig. 20 shows a sample of images used to recently to confirm that recognition can be measured using and fMRI scan. The scan in Fig. 21 shows the activity that occurred in the subjects brain during the viewing of the first and third image in Fig. 20, and Fig. 22 shows the scan that resulted when the subject was viewing the center image in Fig. 20. The striking point about the scan in Fig.'s 22, is that it shows that Broca's area of the brain and the lateral geniculate nucleus, the areas of the brain responsible for language, recognition and processing of visual information respectively, are active while the subject is viewing the center image in Fig.20. Additional functional images of the brain are shown in Fig.'s 23 and 24. The goal of further research is to distinguish "active search" from "object recognition".

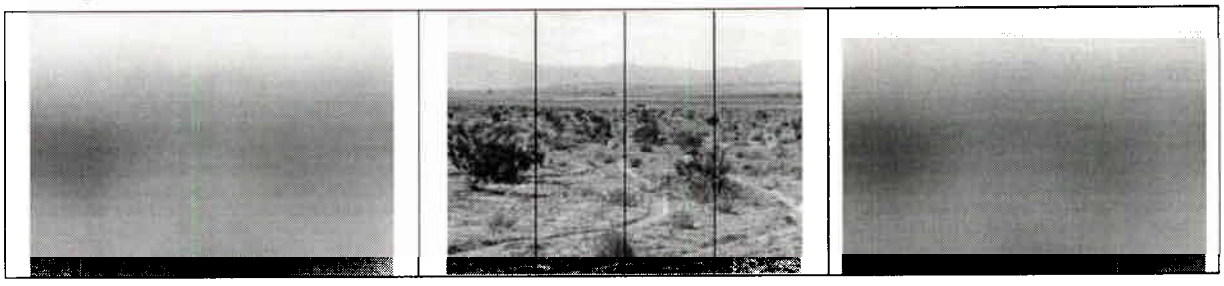

Fig. 20: Image sequence to test for detection of search and identification in an fMRI scan

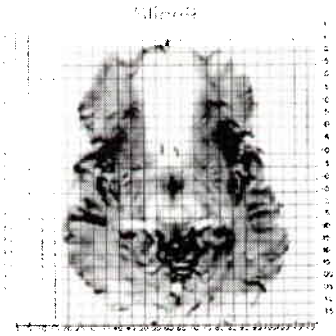

Fig. 21 Scan during presentation of null

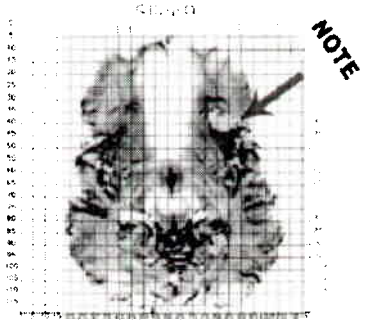

Fig. 22: Scan during recognition of target

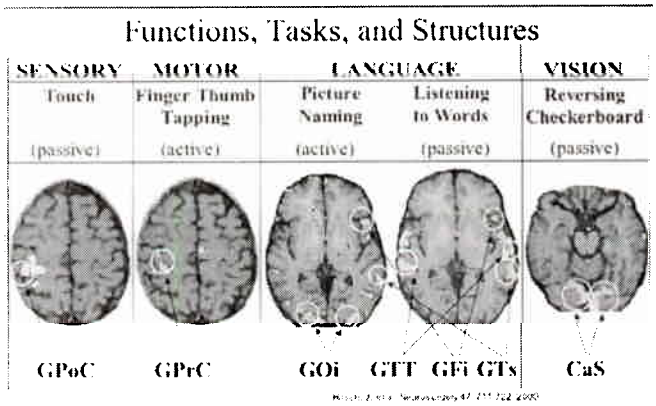

Fig. 23: Functional mapping using fMRI [5]

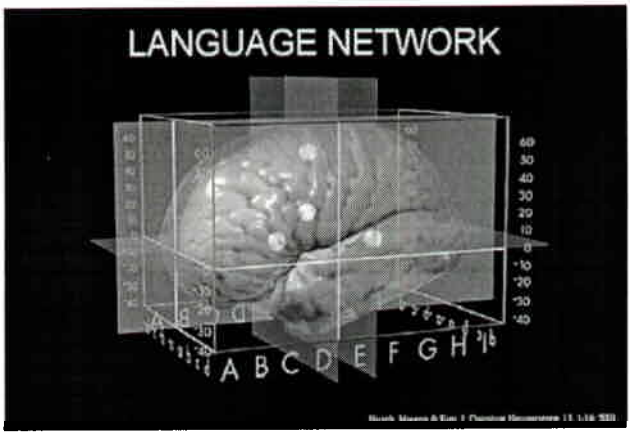

Fig. 24: Broca's area from fMRI scan [6]

IV. CONCLUSION.

Entropy correlates to Probability of detection (Pd) for the LAV data set. We want to explore other data sets to see if the entropy metric correlates well for them. The statistical results of the RPG test were analyzed. The detection rate of the RPG in the image was analyzed versus the sensor type. People had the highest detection rate of the RPG using the IR sensor, and the lowest detection rate was with the visual image. People taking the RPG test also commented on the fact that they preferred the fused image, because it had more details. Future experiments will involve using functional magnetic resonance imaging (FMRI) as a window into the brain to develop system models to distinguish active search from object recognition.

References:

1. Vollmerhausen, R.H., Driggers, R.G., Ronald, G., and Tomkinson, M., "Improved image quality metric for predicting tactical vehicle identification," Proc. SPIE Vol. 4030, p. 60-69, Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XI, Holst, G., Ed., July, 2000.
2. Mazz, J., "Search and Target Acquisition Model Comparisons, Unaided Eye-Analysis", AMSAA Division Note, DN-CI-11, January 1997.
3. Meitzler, T., Sohn, E., Singh, H., and Gerhart, G., "Detection Probability using relative clutter in infrared images, IEEE Transactions on Aerospace and Electronic Systems, Vol. 34, (3), p. 955, July, 1998.
4. Hirsch, J., www.fmri.org.
5. Hirsch, J., et al., "An Integrated Functional Magnetic Resonance Imaging Procedure for Preoperative Mapping of Cortical Areas Associated with Tactile, Motor, Language, and Visual Functions," Neurosurgery, Vol. 47, No. 3, Sept. 2000, pp. 711-722.
6. Hirsch, J., Moreno, D.R., Kim, H.S., " Interconnected Large-Scale Systems for Three Fundamental Cognitive Tasks Revealed by Functional MRI," J. of Cognitive Neuroscience, Vol. 13-3, pp. 389-405