TWO BIRDS WITH ONE STONE: APPLICATION OF FUNDAMENTAL COGNITIVE THEORY OF VISUAL PERCEPTION SUPPORTING FRATRICIDE PREVENTION AND SENSOR MODELING EXPERIMENTATION

John D. O'Connor
US Army Night Vision and Electronic Sensors Directorate
Fort Belvoir, Virginia

Abstract

Effective prevention of direct fire fratricide and sensor modeling experimentation are both reliant on the understanding of cognitive processes associated with visual perception. Research conducted at the US Army Night Vision and Electronic Sensors Directorate (NVESD) from the mid 1990's to the present has applied the recognition-by-components cognitive theory to develop combat identification tools to prevent fratricide, misidentification and collateral damage, and to aid in the development of sensor models through the improvement of human perception testing.

1. Introduction

In the case of direct fire target acquisition and the prevention of fratricide, visual confirmation of the identity of the acquired target is required before a target may be engaged. The individual or individuals responsible for the engagement decision must therefore be trained to accurately determine the identity of potential targets through visual or electro-optical verification. Deficiencies in combat identification training are believed to have been a significant contributor to direct fire fratricides during Operation Desert Storm.

The informational, situational and phenomenological cognitive requirements associated with accurate target acquisition decisions must necessarily be well-characterized for the development of training tools capable of reducing fratricide and collateral damage. The theory of recognition-by-components developed by Dr. Irving Biederman presented a cognitive and phenomenological frame work for the successful development of the Recognition of Combat Vehicles (ROC-V) computer based trainer.[1] The ROC-V program has been credited with reducing direct fire fratricides involving vehicles during Operation Iraqi Freedom. ROC-V is currently required training for Soldiers and Marines.

Human perception experiments involving visual target identification are the foundation of NVESD's sensor performance models. NVESD's models are relied upon by acquisition authorities, sensor developers, and sensor simulators worldwide. In recent years, the majority of NVESD experiments contributing to sensor model advancements have employed the application of recognition by components theory to improve the sensitivity, accuracy and precision of human perception testing and experimentation. NVESD researchers in visual complex object perception have answered task-related questions that support improvement of man-machine task performance and selection of sensors to meet warfighter requirements.

The application of Biederman's recognition by components theory has supported significant advancements in training, phenomenology, man-machine task performance and sensor modeling. The success of the integration of Cognitive Engineering into NVESD engineering research indicates a potential for the incorporation of cognitive theory in other disciplines where the human decision process plays a significant role.

2. Visual Perception and Cognitive Modeling for Military Applications

The cognitive processes associated with visual complex object perception [2,3,4,5,6,7] are well documented for many military applications. The performance, for example of visual perception tasks associated with complex objects for military targeting including detection, recognition,

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **DEC 2008**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE **Two Birds With One Stone: Application Of Fundamental Cognitive Theory Of Visual Perception Supporting Fratricide Prevention And Sensor Modeling Experimentation**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **US Army Night Vision and Electronic Sensors Directorate Fort Belvoir, Virginia**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)

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

13. SUPPLEMENTARY NOTES
See also ADM002187. Proceedings of the Army Science Conference (26th) Held in Orlando, Florida on 1-4 December 2008

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**	**UU**	**8**	

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

classification, orientation, and identification. These tasks have been supported by research leading to such varied tools as Paul Garber's World War II target kites and modern computer-based controls, interfaces, trainers and simulators. Such tools are credited, within the military frame of reference, with saving hundreds of lives and increasing unit combat power and effectiveness.

I. A. Ryback et al. propose a notional model describing visual perception and recognition and the underlying processes termed the Behavioral Model of Visual Perception and Recognition, or BMV.[8] The primary units of object perception are features extracted from an attention window corresponding to the set of perceived image information from a single foveation, associated with a single saccade. The extracted features are processed in sensory memory; working memory aids in the development of a hypothesis regarding the object characteristics or identity. This hypothesis is rejected or accepted based on a priori knowledge, deduction, or possibly reflex or conditioning. If the hypothesis is rejected, multiple foveal fixations may be required to extract enough features from the image to produce the desired level of recognition (assuming the image information is sufficient for the task); thus the process is iterative. The assumption of validity of the BMV model is convenient for the purposes of this literature review, as it relates to further discussion, but the author does not discount the validity of other visual perception models.

3. Complex Object Perception and Infrared Battle Sights

Initially, American development of night vision technologies focused on the Image Intensification (I^2) devices because of limits of the original lead sulfate infrared detector design. In the late 1960's, development of Mercury Cadmium Telluride devices combined with advances in cryogenics created a more viable and effective platform-based thermal infrared night vision system.

There is, however, a human factors problem associated with thermal infrared battle sights that is not relevant to I^2 systems. Because I^2 systems "see" reflected wavelengths, overlapping with or close to the reflected wavelengths seen by the human eye, the cognitive processes and paradigms associated with interpreting information from I^2 displays are nearly identical to those associated with normal vision. Conversely, thermal infrared battle sights convert emitted heat to an electronically displayed image. The images created by these passive infrared systems are not intuitively decipherable. The sight user "sees" emitted radiation rather than reflected and, other than general shape information, the features of an image do not conform to the cognitive visual paradigms of the user. Because sight users were looking for targets that tended to be significantly different from the temperature of background objects, detection was much easier than with reflective wavelength sensors and normal vision. Target features necessary for identification, however, were much harder to discern based on the differing distributions of energy across the target and the higher optical and atmospheric blurs associated with longer wavelengths of radiation. [9]

During the Gulf War of 1991, a significant proportion of Coalition casualties were believed to have been fratricides. Some of these fratricides were determined to be the result of misidentification.[10]

The first generation of thermal infrared battle sights (known as Forward Looking Infrared systems or FLIRs) could detect objects at long ranges, but did not have significant resolution and sensitivity to support long-range combat identification. The US Army approved production of improved thermal electro-optical sensors (known today as 2nd generation FLIR) capable of improving situational awareness and supporting long range combat identification. Such a capability could reduce fratricides from misidentification while increasing combat power.

The benefit associated with improved FLIR performance could not be exploited unless sensor operators could extract accurate information from their battle sights. The US Army and Marine Corps therefore funded NVESD to develop thermal combat identification training to support 2nd generation FLIR's improved operational capabilities. NVESD researchers needed to determine the critical informational components in thermal images and integrate those components into an effective combat ID trainer.

The key to thermal combat identification was discovered in an unusual place: chick sexing. Biederman and Shiffrar [11] conducted object identification experiments focused on chick-sexing as a model for complex visual object recognition. Their results characterized cognitive processes directly related to infrared battle sight target identification. Their conclusions indicated that the performance of identification tasks related to complex objects could be improved through the use of graphic aids and the development of an identification paradigm.

Chick sexing is the process of separating male from female chicks by the visual inspection of their external genitalia. Chick sexing is especially important in the commercial egg production industry; the cost of supporting unproductive and disruptive males is significant, so removal of males from as early as possible from the egg laying population can reap substantial financial benefits.

The chick-sexing task has historically been performed by trained experts, or professional sexers, and their accuracy rate on the job is typically 95% or greater. The training of the experts involved sexing of live chicks, with no evidence of the use of diagrams or other training aids. The standard training period required approximately 2-3 months, with a requirement of 95% or better chick sexing accuracy at a rate of 900-1000 chicks per hour. After several years of on-the-job experience, higher chick-sexing accuracy was reported, with some sexers achieving rates above 99%.

When professional chick-sexers perform their visual task, they face many of the same challenges as infrared gunners. Their attention window interrogates a spatially small target that is commonly within the area of a single foveation. The features to be extracted for and transformed for sensory memory are often small or subtle. Those features are often collocated among other similar features, creating a perceptual signal-to-noise issue.[12] Several saccades may be required for accurate identification, but correct identification usually occurs in less than .5 seconds. Resolution, blur, apparent ΔT, target motion, clutter, and target feature variance are common dimensions affecting the visual identification of the target, whether chick or military vehicle.

Chick sexers and infrared sight users also face similar challenges regarding training for their respective tasks. The working memory and long term memory must have completed many repetitions of the task before identification accuracy is reliable. Further, identification paradigms are commonly employed to increase accuracy and efficiency. One significant cognitive difference between the tasks is the number of possible outcomes; a chick sexer can only identify as male or female, while the infrared sight user must discriminate between dozens or hundreds of potential identities.[13]

The chick sexing experiment was conducted as an object identification experiment comparing the results of three groups of subjects. The first group of 18 observers, (university students and faculty), took a pretest containing 18 images of differing chick genitalia, and a post-test of the same images re-randomized. No feedback or training was provided to the first group. The second group of 18 observers, (also university students and faculty), took a pretest containing the same 18 images of differing chick genitalia, and were given approximately one minute of training on the use of instructional sheet with drawings of critical features, before taking the same post-test of the same re-randomized images. The instructional sheet was developed primarily using input from experienced chick sexers. The third group comprised five current or retired professional chick sexers. The professional sexers were asked to identify the sex of the chicks in each of the 18 images presented to the first and second groups. It is important to note that the 18 images chosen depicted rare and difficult types
.

The naïve subjects of the first group averaged 59.0 % correct on the pretest and 54.1% on the post-test. Naïve subjects who were trained (second group) averaged 60.5% on the pretest and 84% on the post-test. The professional sexers averaged 72% correct. The Pearson product moment correlation between the naïve subjects and professional sexers was .21, while that between trained subjects and professional sexers was .82. Biederman and Shiffrar conclude that "…after instruction the performance of the naïve subjects more closely resemble that of the professional sexers than their own uninstructed performance one minute earlier." They further conclude that there is "a considerable advantage achievable from a simple set of pictorial instructions that specify the location of diagnostic binary or trilevel contrasts of contour".

The experiments of O'Connor and O'Kane translated the chick sexing paradigm into a infrared signature identification paradigm: "What's Hot and What's Not".[14] This general principle for infrared image understanding for combat vehicle identification specifies the location of robust emissive source locations, such as engines and exhausts, as primary identifiers, with additional shape cues used for vehicles with similar emissive source configurations. The US Army and Marine Corps tasked O'Kane and O'Connor with transforming this principle into the Recognition of Vehicles, (ROC-V), combat identification trainer. The trainer contains visible and infrared signatures of nearly 200 vehicles and aircraft and is required training for all gunners and sight users, infrared or otherwise. Officials from the US Joint Combat Identification and Evaluation Team have gone on record stating that the ROC-V trainer helped significantly lower rates of fratricide during Operation Iraqi Freedom, and is thus credited with saving numerous coalition lives.

The infrared image identification problem was ultimately solved by first developing a strategy for infrared combat identification associated with an effective training tool capable of creating a population of infrared gunners who could reliably identify the infrared signatures of friend and foe alike. The strategy for infrared combat identification was inspired by the recognition-by-components theory[15]

The development of a strategy for thermal combat identification, combined with a viable thermal combat ID trainer (ROC-V), made possible the development of methods to enhance the infrared gunners' vehicle identification skills. Prior to the ROC-V trainer, few individuals were capable of accurately assessing thermal vehicle images, and thus the critical information content associated with thermal sensor tasks was difficult to characterize.

4. Engineering Applications for Cognitive Models of Perception

A significant peripheral benefit of the development of the ROC-V thermal ID trainer improved the accuracy of sensor performance experiments by reducing the human error and variability associated with any given experiment requiring combat identification. All observers could be trained to a higher and more equivalent skill level, such that the variability in observer responses was predominantly due to the experimental sensor treatments or renderings rather than observer variation.

The use of trained observers improved the sensitivity of sensor modeling experiments, contributing directly to the development of greatly improved sensor models, such as NVTherm and SSCam. The development of the V50 task difficulty metric used in current sensor models resulted directly from these more sensitive modeling experiments.

In one of the earliest experiments employing observers trained using the ROC-V program, NVESD scientists investigated the display rendering of digital thermal information comparing automated and manual image enhancements for the combat identification task.[16] The advent of digital output detectors for infrared sensing created a potential for providing more and better information for infrared sight users. When first implemented, however, the control of displayed information was relegated to the old analog controls. Analog controls limited information to a linear and continuous transformation of output greyshades from 4096 (12 bit) greyshades or more to 256 (8 bit) or less. The loss of available contrast information to down-sampling could thus be greater than an order of magnitude. In short, the full potential of digital sensor systems could not be realized.

Early automated display algorithms employed a variety of display output density mapping schemes to convert the 12-16 bit digital sensor data to 8 bits for display. These included histogram equalization, local area processing, and region of interest processing. Manual Display Mapping (MDM), developed by the author, differs from these algorithms because it allows the user to manipulate the display output density of different regions of the sensor output by linear or non-linear mapping to pixel values. The user can thus allocate, or "tune", pixel intensities (greyshades) to output regions expected to contain targets or relevant information (Figure 1).

Two experiments compared the identification (ID) performance of observers viewing low-resolution images with a uniform atmospheric degradation and images blurred to simulate multiple ranges. In each experiment, images were processed with an automatic complex algorithm held in high regard by the infrared sensor community, and MDM. Results of these experiments indicated a highly significant MDM range performance improvement vis-à-vis the automated algorithm under a variety of typical infrared imaging conditions.

All test subjects in both experiments were US Army Soldiers or Marines. All test subjects in both experiments were required to train to a 96% medium range P(id) standard (less than 2000 pixels on target) for a set of twelve vehicles that included all vehicles in both experiments before participating. The US Army's *Recognition of Combat Vehicles*[17] (ROC-V) infrared signature training software was used to achieve this level of proficiency.

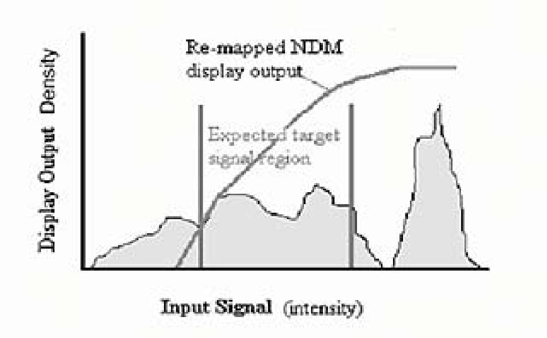

Figure 1. The MDM concept (an earlier acronym was NDM)

The first experiment tested target ID performance of 36 observers presented images of 9 tracked vehicles (Figure 2) viewed from 8 aspect orientations and processed with MDM and a second generation FLIR B-kit automatic algorithm emulator. For the MDM processing, the author viewed representative images and their histograms for each of the five collection environments from which images were drawn. The author then determined regions to be compressed or enhanced with MDM and applied this to each image to create scene dependent and target independent optimization. Observers were presented with 50 target images processed with each method (Figure 4) in a within-subjects design. The images were uniform in range to target and resolved target pixels. Images had a low count of resolved target pixels (less than 400) and were significantly degraded by atmospheric blur, so the mean expected probability of ID was low.

In this one-range forced-choice experiment, the mean B-kit ID score was 19.2% when corrected for chance; the mean MDM ID score was 28.2% when corrected for chance. A 1-tailed *t*-test (35)=8.22, (p< .01) of the mean MDM improvement over B-Kit was significant. These results suggested that non-linear mapping could improve ID performance by displaying more of the relevant information in the sensor output in conditions where blur, magnification, and noise were limiting human performance.

For the second experiment, 28 subjects completed a 100-image object identification perception test upon completion of the required ROC-V training. The observers were presented 50 images treated with each method, (100 total) of 10 tracked vehicles (Figure 3) viewed from 8 aspects in a Latin Squares within-subjects design. The 50 base images were divided into five sets of 10 images, balanced for comparative target discrimination difficulty. The range of pixels on target was 440 to 1120 pixels. Each set was processed with Gaussian blur with radii of 7,8,9,10, and 11 pixels to generate notional "ranges" (Figure 5). In the second experiment, the corrected mean scores for B-kit automatic and MDM images were respectively 23.9% and 47.8%. A 1-tailed *t*-test (27)=13.79 (p< .01) of the mean MDM improvement over B-Kit was significant. [18]

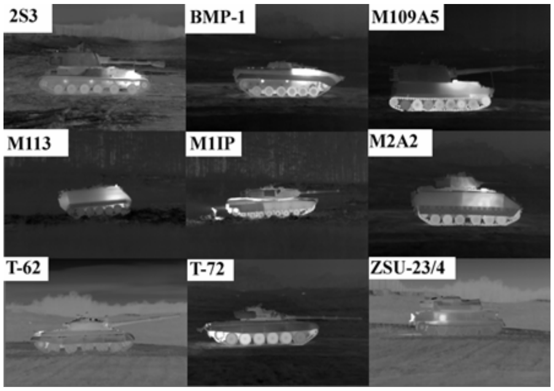

Figure 2. Experiment 1 targets

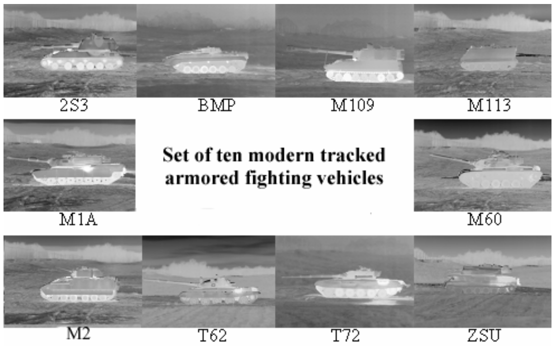

Figure 3. Experiment 2 targets

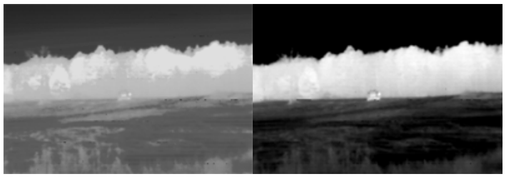

Figure 4. Experiment 1 image examples: M109A5 w/ B-kit emulator (left), MDM (right)

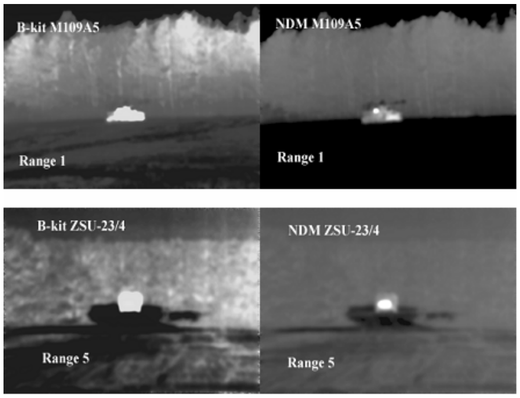

Figure 5. Experiment 2 image examples

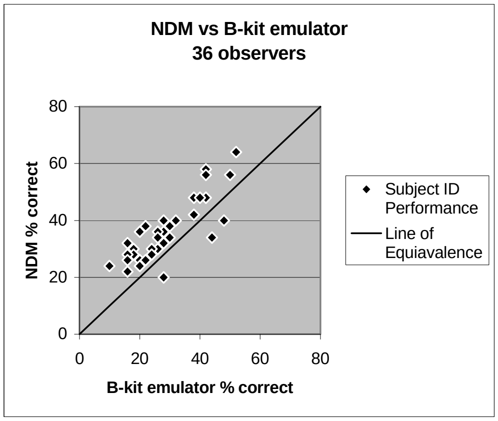

Figure 6. Experiment 1 Subject ID performance (NDM later changed to MDM)

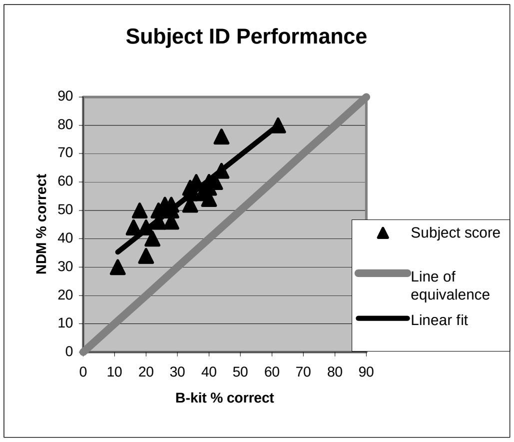

Figure 7. Experiment 2 Subject ID Performance

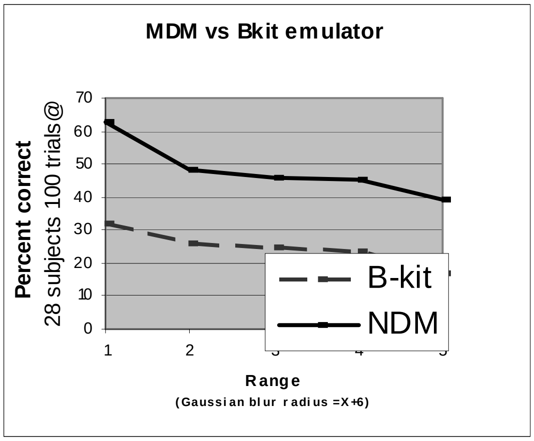

Figure 9. Experiment 2 ID performance, MDM versus B-kit emulator

5. Conclusions and Discussion

Object identification experimentation methods have made significant contributions in the area of visual complex object perception. For example, the US Army Night Vision and Electronic Sensors Directorate, (NVESD) has published the results of over 100 experiments since 1996 using a single object identification experimental design developed by Dr. Barbara O'Kane, Richard Vollmerhausen, and John O'Connor. The design is powerful because it accounts for chance and does not yield ambiguous data. In combination with rigorous observer task training, (such as in the aforementioned method of O'Kane, Vollmerhausen, and O'Connor), variance caused by differences in observer task skill level can be mitigated, resulting in a lower residual error, better measurement of effect size, and greater power.

With respect to Cognitive Engineering, both O'Kane and O'Connor's and O'Connor and Olson's experiments inspired improvements in human-machine performance. The first improved training methods, the second improved design and modeling of electro-optical systems.

Biederman and Shiffrar's conclusions and ideas were seminal for development of O'Connor and O'Kane's validated infrared combat identification method, "what's hot and what's not". That method inspired multi-service funding and support for research and development of what is now ROC-V, a highly successful institutionalized computer-based fratricide prevention and combat identification training tool credited with making the battlefield safer for coalition forces.

The ROC-V trainer development, and its associated implications for the cognitive theory of infrared image understanding, made possible the MDM experiments which affect automated image processing and algorithm development strategies, and development of human-in-the-loop digital processing controls that allow the user to "tune" the display to maximize extraction of the desired information. The MDM experiments support the author's contention of the mathematical impossibility of any single algorithm, or reasonable and discrete set of algorithms, achieving optimal or near optimal visual discrimination performance, given the infinite variability of infrared scenes, targets, users, and objects of interest.

Automation, by its very nature, cannot address the strengths, weaknesses, informational requirements, and situational challenges of the individual users. Resources previously allocated to the development of "magic button" automated algorithms are being better spent in the development of user controls that allow the human to fully exploit the task performance advantages of digital, as compared to analog, systems designs. Digital sensor systems, inspired by these experiments, now have controls that allow the user to select from, manipulate, and create a wide variety of signal intensity transfer functions, (SITFs).

The author believes, based on extensive experience with users and sensors, allowing a well-trained user to manipulate the SITF will lead to the user's better understanding the characteristics, meaning, and value of scene and object information. Further, with continued application, the development of robust, effective, and reliable "tuning" strategies may improve task performance. Future research in this area has the potential to impact display utility for military target acquisition tasks, and perhaps other digital signal processing tasks as well.

In combination, these two experiments suggest that the adaptability of the human eye-brain system for the extraction of information through the attention window, combined with a manual control of digital contrast adjustment and sufficient task training, are more effective than machine-based solutions for sensor-based object identification tasks. It can be argued that many discoveries are yet to be made in the fledgling field of automated signal processing for visual display. It can also, however, be argued that manual digital signal processing is also in its infancy. The author believes that, while machines have a limited potential to achieve higher performance on complex visual object perception in the remote future, significant benefits of further developing training tools and human-in-the loop digital signal processing tools are realizable in the near future.

6. References

[1]. Biederman, Irving. *On processing information from a glance at a scene*, "User-Oriented Design of Interactive Graphics Systems". (ACM, New York, 1977)

[2]. Bieberman, Lucien and Tsou, Brian. *Image Display Technology and Problems with Emphasis on Airborne Systems.* (Institute for Defense Analysis, Alexandria VA, 1991) pp 2-2, 2-5

[3]. J.J. Gibson & L.E. Crooks, (1938) *A theoretical field-analysis of automobile driving*, American Journal of Psychology vol. 51 pp 453-471

[4]. J.J. Gibson, P. Olum et al. (1955) *Parallax and perspective during aircraft landings*, American Journal of Psychology vol. 68 pp 372-385

[5]. J.J. Gibson & B. Bridgeman, (1987) *The visual perception of surface texture in*

photographs, Psychological Review vol. 49(1) pp 1-5

[6].John Johnson, (1958) *Analysis of Image Forming Systems*, Proc. of Image Intensifier Symposium, pp. 249-273.

[7]. J. O'Connor & B. O'Kane. (1998) *Recognition of human activities using hand-held thermal systems*, Proc. of SPIE, vol. 3394-08 pp 51-61

[8]. Rybak, I. A., et al. (2005) *Attention-guided recognition based on "what" and "where" representations: A behavioral model*, *Neurobiology of Attention* (Eds. Itti, L., Rees, G. and Tsotsos, J.). Elsevier Acad. Press, pp. 663-670

[9]. Norman S. Kopeika. "A System Approach to Imaging", (SPIE Optical Engineering Press, Bellingham WA, 1998) pp 127-128

[10]. Geoffrey Regan. "Blue on Blue: a history of friendly fire" (Avon, New York 1995)

[11]. Biederman, I. & Shiffrar, M. (1987) *Sexing day-old chicks: a case study and expert systems analysis of a difficult perceptual learning task*, Journal of Experimental Psychology: Learning, Memory, and Cognition, vol. 13 #4 pp 640-645

[12]. Vicente, Kim. "The Human Factor" (Routledge, New York & London, 2006) pp 138-139

[13]. Biederman. I. and Cooper, E.(1999) "Subordinate-level object classification re-examined". Psychological Research vol. 62 pp. 132-153

[14]. O'Connor, J. & O'Kane, B. (2000) *What's hot and what's not: general principle for image understanding applied to combat vehicle identification*, Proc. of SPIE, vol. 4052-15 pp 111-117

[15]. Biederman, Irving. (1987) *Recognition-by-components: a theory of human image understanding*, Psychology Review vol. 94 pp 115-147

[16]. John D. O'Connor, Jeffrey T. Olson. (2002) *Non-Linear Mapping for Improvement of Display Comprehension of Low Resolution Images*, Proceedings of the Ground Target Modeling and Validation Conference pp 296-303

[17]. Recognition of Combat Vehicles (ROC-V) version 5E vehicle thermal signature trainer, Developed by NVESD, distributed by Project Manager Ground Combat Tactical Trainer (PM GCTT), sponsored by Joint Forces Identification Evaluation Team (JCIET) for US Department of Defense.

[18]. O'Connor, J. et al. (2004). *Performance evaluations of a manual display mapping method*, Journal of Electronic Imaging vol. 13(4) pp 709-713