Modeling and integration of situational awareness and soldier target search

Paul F Evangelista, Christian J Darken and Patrick Jungkunz

The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology published online 22 September 2011

DOI: 10.1177/1548512911415726

Original article

JDMS

Journal of Defense Modeling and Simulation: Applications, Methodology, Technology
1–19

DOI: 10.1177/1548512911415726
dms.sagepub.com
SAGE

Modeling and integration of situational awareness and soldier target search

Paul F Evangelista[1], Christian J Darken[2] and Patrick Jungkunz[3]

Abstract

Representation of search and target acquisition (STA) in military models and simulations arguably abstracts the most critical aspects of combat. This research focuses on the search aspect of STA for the unaided human eye. It is intuitive that an individual's environmental characteristics and interpretation of the environment in the context of all comprehended information, commonly summarized as their situational awareness (SA), influences attention and search. Current simulation models use a primitive sweeping search method that devotes an unbiased amount of time to every area in an entity's field of regard and neglects the effects of SA. The goal of this research is to provide empirical results and recommend modeling approaches that improve the representation of unaided search in military models and simulations. The major contributions towards this goal include novel empirical results from two incremental eye-tracking experiments, analysis and modeling of the eye-tracking data to illustrate the effect of the environment and SA on search, and a recommended model for unaided search for high-fidelity combat simulation models. The results of this work support soldier search models driven by metrics that summarize the threat based on environmental characteristics and contextual information.

Keywords

combat models, search and target acquisition, situational awareness

1 Introduction

This study seeks to improve understanding of the effects of situational awareness (SA) on unaided search and target acquisition (STA). The ultimate goal of this research is to improve representation of STA in combat simulation models. Future soldier systems focus on creating information dominance to enhance SA with an expectation that enhanced SA improves operational effectiveness. However, existing models and simulations are not capable of evaluating the effects of even the most primitive SA. Specifically, STA in current combat simulation models follow preset patterns, an inaccurate representation of reality that neglects SA and potentially corrupts the outcome of the simulation.

Measuring the behavior and performance of the unaided human eye remains an open area of research.[1,2] Unaided search differs significantly from aided search and detection. Devices such as scopes and other technical sensors immediately bound search behavior characteristics. Aided search yields parameters such as resolution and restricted fields of view. Human behavior becomes more predictable under these circumstances, especially if training and doctrine advocate particular techniques. Device parameters, training, and doctrine create structure that defines aided search and supports modeling of this behavior. STA of the unaided human eye does not offer the same structure. Unaided human vision does not have easily bounded or predictable parameters.

Certain components of human vision are well understood; however, other components remain a mystery. Scientists understand that humans perceive a fraction of the electromagnetic spectrum, extract meaning from color and shape, and

[1]TRADOC Analysis Center – Monterey, USA
[2]MOVES Institute and Department of Computer Science, Naval Postgraduate School, USA
[3]German Naval Office, Germany

Corresponding author:
Paul F Evangelista, TRADOC Analysis Center – Monterey, Naval Postgraduate School, Monterey, CA 93943, USA.
Email: paul.evangelista@us.army.mil

see clearly at a two-degree foveal point, but clarity degrades in the periphery. Other human vision phenomena are not as clearly understood. Consider a person's propensity to fixate at certain areas in a scene but ignore others. Soldiers almost instinctively anticipate threat from certain areas within a scene but quickly surmise other areas as safe. Psychological theories detail the likely cognitive processes that lead to these types of behaviors, but predicting human visual behavior given complex and natural visual stimuli remains an open problem.[3,4] Bruce et al.[5] summarize visual search as the product of two factors: the properties of the surrounding environment and the goals of the observer.

A major flaw that currently exists in combat models involves the application of unbiased search. The underlying acquisition algorithm accepted and widely used in many combat models can be traced to an experiment performed by Johnson[6] that essentially correlated and summarized the difference between target contrast and detection by the observer. This experiment, supported largely by the US Army Night Vision and Electronic Sensors Directorate (NVESD), was developed to support modeling of aided eye detection with a restricted field of view, such as looking through a scope. The algorithm used in most US military combat models to represent soldier STA uses Johnson's results to support target acquisition; however the algorithm typically employs search patterns based upon aided search. Although there are recommended search strategies for the unaided eye, research supports that pre-attentive processing of environmental surroundings supported by human learning, specifically our ability to aggregate and associate to build context, trumps any attempts to train or control human search behavior.[1,2]

1.1 Assumptions, related work, and organization of this paper

The critical assumptions that surround this work are as follows:

- eye fixations represent areas of search;
- target scenes presented in tier I and tier II experiments represented enough of the visual field to exercise realistic target search (the target scenes presented in tier I and tier II covered visual fields of 40 degrees and 71 degrees, respectively);
- eye velocity less than 12.5 degrees per second indicates a fixation;
- visual determination sufficed to create levels of saliency to support the tier I single-target experiment;
- in the tier II experiment, the median frame of a fixation adequately indicates the center of the fixation (see Section 4);
- the urban scenes presented to subjects in tier I and tier II represent realistic combat target scenes;
- features in this study, specifically the coefficient of variance, generalize to mixed urban and other environments (e.g. wooded, desert).

A substantial body of literature relates to this work, and many of these references have been cited throughout this paper. Vaughn's survey[7] of STA for military models systematically summarizes and compares several models. Doll and Home[3] emphasize the role of pre-attentive processing and clutter, and they also reinforce the false nature of random search. Entin and Entin[8] systematically varied the context of a target scene and found that false positives increased as the reports of the density of enemy presence increased. Zelinsky[9] provides a comprehensive survey of eye-movement modeling and presents a model that trains on an image and searches a subsequent scene. Mazz et al.[10] present the affect of target movement and contrast in target detection.

The major contributions of this work include empirical results that support fundamentally changing the representation of soldier target search in combat simulation models and a recommended search model that can incorporate these empirical results. Experimental results from this work support the inclusion of pre-attentive processing in target search, essentially enabling entities to (1) guide search based on nearly instantaneous processing of environmental characteristics and (2) detect high-contrast targets anywhere within their field of regard. Results also indicate that processing of basic contextual information does affect target search. The modeling approach promotes inclusion of environmental characteristics as a critical component to drive soldier target search.

This paper is organized as follows. Section 2 details the overall experimental methodology; Section 3 details the tier I experiment to include a single-target experiment and a static experiment; Section 4 covers the tier II static experiment; Section 5 proposes a modeling approach capable of incorporating the empirical results from the tier I and tier II experiments.

1.2 Terms and definitions

Concepts in STA research for military models and simulations often relate directly to concepts and theories in psychology, and this can create conflicts in terminology. Human vision covers nearly 180 degrees, but acuity in this field varies. The foveal point is approximately 2 degrees and enables our most detailed vision. Acuity degrades rapidly in the periphery; however, the periphery is sensitive to movement and variations in light. Military models typically refer to the entire vision field as the field of regard, and this paper will use this term similarly. The field of view in military models is generally 30–45 degrees, representing the primary area of attention. The models usually restrict acquisitions to objects only within the field of view.

2 Experimental methodology

The two incremental experiments that supported this research relied heavily on eye-tracking technology. As mentioned previously, the behavior of the human eye is not entirely understood. Eye tracking provides a unique perspective towards improving this understanding. The first eye-tracking experiment took place at the Naval Postgraduate School. Subjects were military officers from various branches of service, and eye tracking occurred with a desk mounted eye-tracking system. Eye-tracking data quality in the first experiment largely hinged on screen calibration and time synchronization between the computer presenting stimuli and the eye-tracking system. The controlled nature of the first experiment created an opportunity to collect data on fundamental target search and eye-movement behavior, such as the effect of eccentricity and saliency. The first experiment also provided an opportunity to rehearse eye-tracking experiment protocol and refine the design of experiments (DOE) that also supported the second experiment (see Figure 1). This experiment provided the study team with exploratory data to gain insights on the significant factors; however, the conditions were far from the combat environment that the experiment needed to replicate. The second experiment occurred in Fort Benning, Georgia, and the subjects were infantry soldiers ranging in rank from private to captain. Target scenes were projected on a 10 foot screen, and the soldiers stood in a virtual environment with a calibrated weapon, helmet, and mobile eye-tracking system. The complexity of the second experiment was clearly greater than the first. Eye-tracking data quality in the second experiment also required calibration and synchronization, and the mobile eye-tracking system created realistic freedom of movement, which introduced significant post-processing challenges that will be discussed in detail later in this paper. The second experiment created experimental conditions with the potential to produce convincing empirical results.

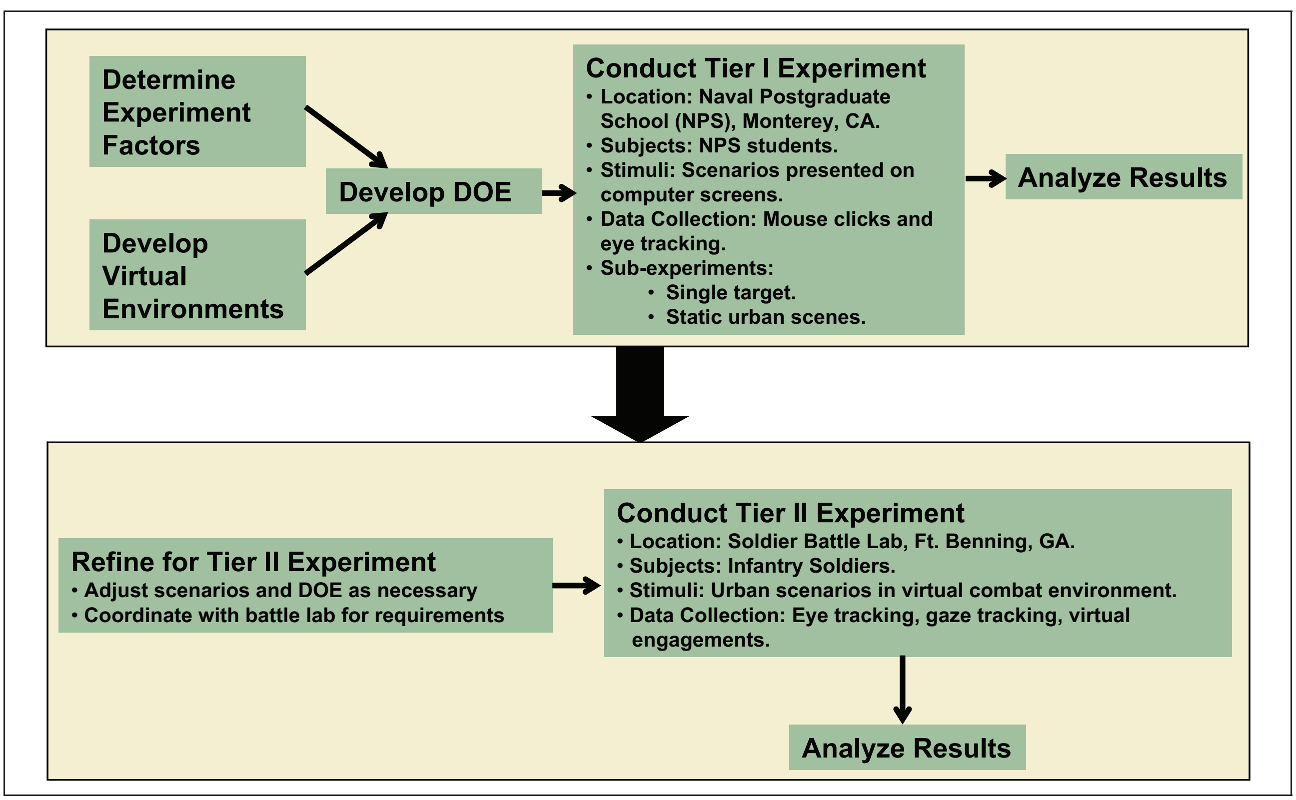

Figure 1. Experiment methodology. DOE: design of experiments.

3 The tier I experiment

The tier I experiment included two sub-experiments. The first sub-experiment will be referred to as the single-target experiment. The single-target experiment focused on fundamental search behavior. This included measuring a subject's ability to detect a single target while systematically varying target eccentricity, saliency, and distracters. This sub-experiment supports the notion of including pre-attentive processing as a component in STA models. Although the outset of this paper clearly defined our focus as target search, pre-attentive processing merges search and acquisition nearly instantaneously. As described by Treisman,[11] 'pre-attentive processing of visual information is performed automatically on the entire visual field detecting basic features of objects in the display. Such basic features include colors, closure, line ends, contrast, tilt, curvature and size. These simple features are extracted from the visual display in the pre-attentive system and later joined in the focused attention system into coherent objects. Pre-attentive processing is done quickly, effortlessly and in parallel without any attention being focused on the display.' There is an argument that salient targets that enter the visual field of regard receive immediate attention and classification; however, targets of high contrast within a short range do not require search. Reece and Wirthlin[12] introduce this concept with several proposed algorithms that consider acquisitions beyond the primary field of view. Pre-attentive processing supports the identification of the presence of an immediate and obvious target; however, this vision function also supports the identification of what is not present. It is uncommon for a person to search an area that does not present contrast, unless prior information indicates the presence of a target or there is belief that a target could emerge. Pre-attentive processing essentially prevents humans from searching an empty sky or a region where lack of contrast deems target detection as unlikely. This is an important concept that extends from the first single-target trials conducted in the tier I experiment to the complex urban search conducted in the tier II experiment. Pre-attentive processing enables rapid detection of immediate targets; however, it also suppresses needless searching. This paper proposes that pre-attentive processing suppresses attention as much as it directs attention, and this is an important consideration for target search models. The target-only trials in the tier I experiment create a foundation to support this concept. The second tier I sub-experiment will be referred to as the tier I static experiment. The sole purpose of this experiment was to rehearse and refine the experiment protocol and DOE for 16 urban search scenes that would remain consistent for the tier I static experiment and the tier II static experiment.

3.1 Tier I single-target experiment

Nineteen students and faculty of the Naval Postgraduate School in Monterey participated in the experiment after providing informed consent. All participants were members of the US Armed Forces across the four services: Army, Marine Corps, Air Force, and Navy. The participants volunteered and did not receive any compensation. All participants were naive with respect to the hypotheses of the experiment. The stimuli consisted of 32 scenes containing one target, one distracter, and one location with semantic influence on the search task (see Figure 3). All stimuli were developed with the Delta3D game engine. The simulated targets used within this experiment depicted uniformed infantry soldiers. The distracter was an unfolded piece of newspaper seemingly attached to a wall behind the target. The location with semantically relevant content was a doorway that could be used for cover, ingress, or egress, and this will be referred to as the hiding location.

3.1.1 Tier I single-target experiment conditions and execution. The size of the target, distracter, and hiding location remained constant over the course of the experiment. Within the scenes the target and distracter varied in eccentricity and saliency, and only the eccentricity of the hiding location varied. The eccentricity of these entities assumed three possible levels (0, 1, and 2) corresponding to values of 5, 13, or 17 degrees of visual angle respectively. These variations occurred along the horizontal axis located at the center of the screen. See Figure 2 for an illustration.

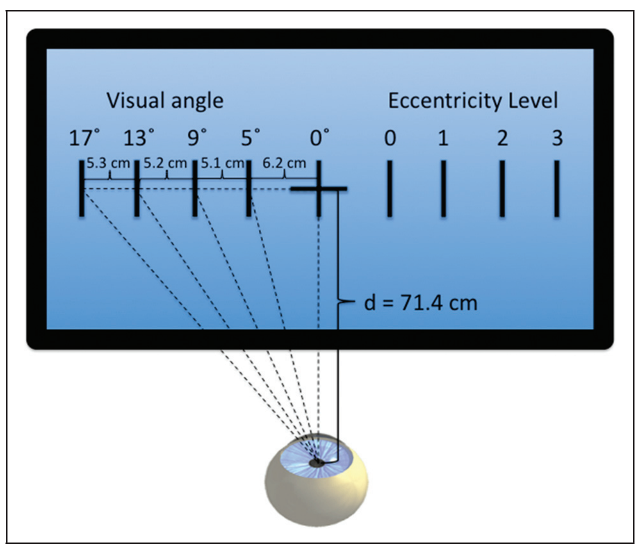

Figure 2. An illustration of the eccentricity levels depicted in terms of degrees of visual angle on the left and eccentricity levels on the right. The crosshairs in the center indicate the location participants were asked to fixate upon before stimulus onset.

The distracter and the hiding location could assume eccentricity levels 0 and 1; the target could assume levels 1 and 2. The salience of the target and distracter varied on a scale of the following three possible levels (0, 1, and 2): one low level, one medium-low level, and one high level. The target could assume levels 0 and 1, whereas the distracter could assume salience levels 1 and 2. The low saliency level was picked in a way such that the object was almost invisible for the lowest level, but could be discriminated from the background when fixated directly. The next higher saliency level, medium low, was set a little higher, so that the object was still hard to see but not as hard as in the lowest setting. At the high saliency level the object was very clearly visible. All three levels had been visually judged by a human observer. Then, the parameters determining the saliency within the stimulus display software were fixed to ensure equal saliency values for all objects with the same levels.

The visual judgment of the contrast level was performed at the same screen, in the same location, and with the same lighting conditions as the actual experiment. The experiment took place in a completely darkened laboratory. The stimuli were presented on a 24 inch thin-film transistor (TFT) monitor set to 60 Hz at a resolution of 1920×1200 pixels measuring 52 cm $\times$ 32.5 cm. Eye tracking was performed with the Seeing Machines FaceLab4 eye tracker. Eye-tracking sampling occurred at 60 Hz and the experiment was only conducted for participants that achieved a screen calibration with a mean error of 1 degree of visual angle or better. Participants sat 71 cm from the monitor resulting in the screen covering a visual angle of 40 degrees. The viewing distance was maintained with a chest rest, against which participants leaned during the experiment. Head movements were not restricted.

Before taking part in the experiment every participant provided an informed consent. Visual acuity and color vision of participants were tested using a modified Snellen chart and the Ishihara color test, respectively. Only participants with an uncorrected vision of 20/30 or better took part in the experiment. With respect to color vision, participants were required to correctly read charts 1–14 of the Ishihara color test in order to be eligible for participation. In order to increase eye-tracking accuracy, participants were not allowed to wear glasses or contact lenses during the experiment.

Participants fulfilling the stated criteria proceeded with the experiment. After calibration of the eye tracker, participants received instructions on the experiment. They were told that their task was to spot enemy targets in camouflage uniform in an urban environment as quickly as possible. The participants were asked to find the targets as quickly as possible, fixate on the targets with their eyes, and then press the spacebar to indicate a successful search. If they could not find the target they were asked to say 'next' and the scene would be advanced for them. They were also informed that in addition to the search target, additional objects could appear as would be expected in an urban environment. No further information about the nature of the objects and their meaning for the experiment was provided in order to avoid any biasing or priming with respect to the distracter or the hiding location. Before the start of the experiment the target was introduced to participants in the high-contrast setting. Neither the distracter nor the hiding location was shown prior to the experiment.

In order to control for the eccentricity of the entities, a fixation cue was displayed before each scene. This fixation cue, black crosshairs in a white circle on a black background, was located at the center of the screen. Before the experiment the participants were told to look at the crosshairs and continue to do so until the search scene was displayed. Scenes at which the initial fixation was not located within 2 degrees of the scene center were considered errors and excluded from the analysis.

Before the start of the experiment, each of the participants demonstrated an understanding of the experiment protocol by answering several questions with regard to the tasks that they would perform. In addition, each participant conducted two practice trials to increase familiarity with the flow of fixation cues, scenes, and the expected input. All 64 stimuli were presented in one session without any interruption. The scene presentation was random.

3.1.2 Tier 1 single-target experiment results. The eye-movement velocity threshold used to discriminate fixations from saccades was 12.5 degrees per second. Unfortunately, this did not allow for the detection of extremely short fixations. These occurred typically when the initial saccade occurred in the direction of the distracter. Due to the important influence of these saccades on the response variables, an additional fixation criteria, direction change, was introduced. Eye-movement vectors were determined by looking at consecutive gaze locations and computing the vector from one location to the next. Whenever the angle between two consecutive eye-movement vectors during a saccade was larger than 60 degrees it was defined as the end of a saccade and beginning of a fixation. If the following gaze recordings dropped below the speed thresholds, they were included in the fixation; otherwise the fixation ended. Visual inspection of scene overlays showed that this method effectively separated saccades from fixations and managed to capture very short fixations that were apparent through a sharp direction change only.

In order to assess the contributions and interactions of the top-down, bottom-up, and semantic factors on attention allocation and eye movements, six response variables were analyzed. The first two variables, namely *the number*

Figure 3. An image exemplifying a stimulus that contains the target, the distracter, and the hiding location. The target is located near the left edge.

of fixations until target fixation and the *time until target fixation*, are closely related. Both are indicators not only of the search performance, but also on the capture of overt attention. Longer times or higher numbers show reduced performance and thus attention capture by the distracter. The time until target fixation was measured from scene onset until the first fixation lands on the target area. The target area is a rectangle around the target extending 2 degrees or 96 pixels out to either side from the minimum and maximum target-coordinate values in both the x and y directions. The number of fixations until target fixation is counted starting with the first fixation, leaving a circle with a radius of 2 degrees around the screen center up until and including the first fixation on the target area. This means that if the first saccade ends in the target area, the *number of fixations until target fixation* is 1.

The reaction time was measured from scene onset until the participants pressed the spacebar to indicate a successful search. Since participants were instructed to first fixate the target and then press the spacebar, this time cannot be compared to other search experiments where the reaction time is usually the only response variable and does not require a concurrent fixation. The incentive for this instruction was to discourage participants from making guesses. The measured reaction time is still different from the time until target fixation, since participants frequently pressed the spacebar during the saccade onto the target. Therefore, the time at which participants pressed the spacebar is still a valid measure of reaction time.

The next response variable, *initial saccade latency*, is the time that expires from scene onset until the end of the last fixation within the 2 degree circle around the screen center. The length of the first on-target saccade measures the perceptual span. It indicates how far a target can be from a fixation location and still be directly recognized and fixated with one saccade. Lastly, *the initial saccade direction* tells whether the target, distracter, or hiding location captured the overtly deployed attention.

The *number of fixations until target fixation* had a mean of 2.35 fixations with a standard deviation of 2.01 fixations. A total of 39.3% (160 out of 405) trials resulted in target fixation after the first saccade, and a total of 33.3% (135) resulted in target fixation after the second saccade.

There was a main effect of target saliency ($p = 0.0002$) and hiding location eccentricity ($p = 0.0002$) on the number of fixations until target fixation. Higher target saliency resulted in a lower number of fixations. This means that targets are spotted more quickly once their saliency or contrast reaches a certain level. Increasing hiding location eccentricity, on the other hand, increased the number of fixations until target fixation (see Figure 4). In addition to the main effects, there was also an interaction between

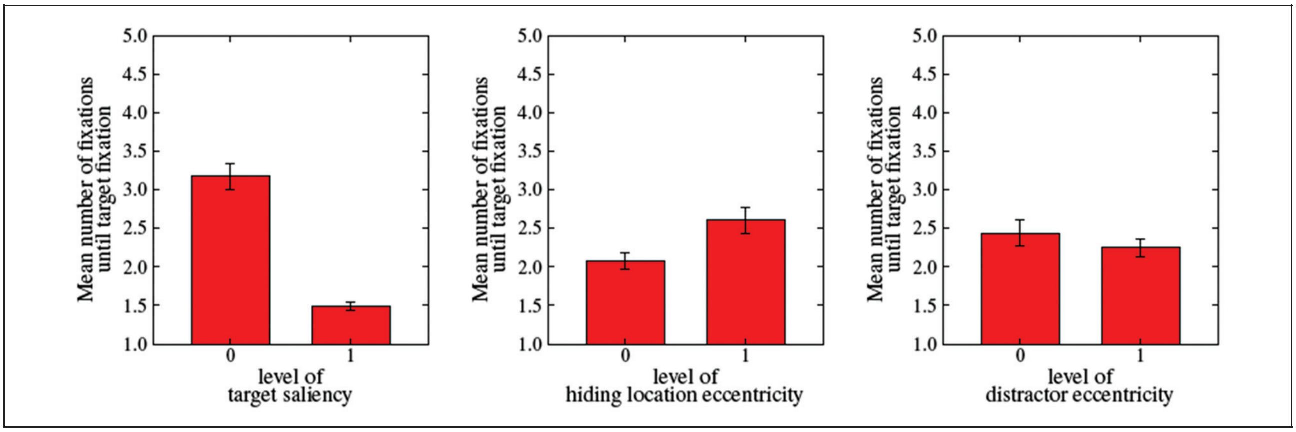

Figure 4. Effects of target saliency, hiding location eccentricity, and distracter eccentricity on the number of fixations until target fixation.

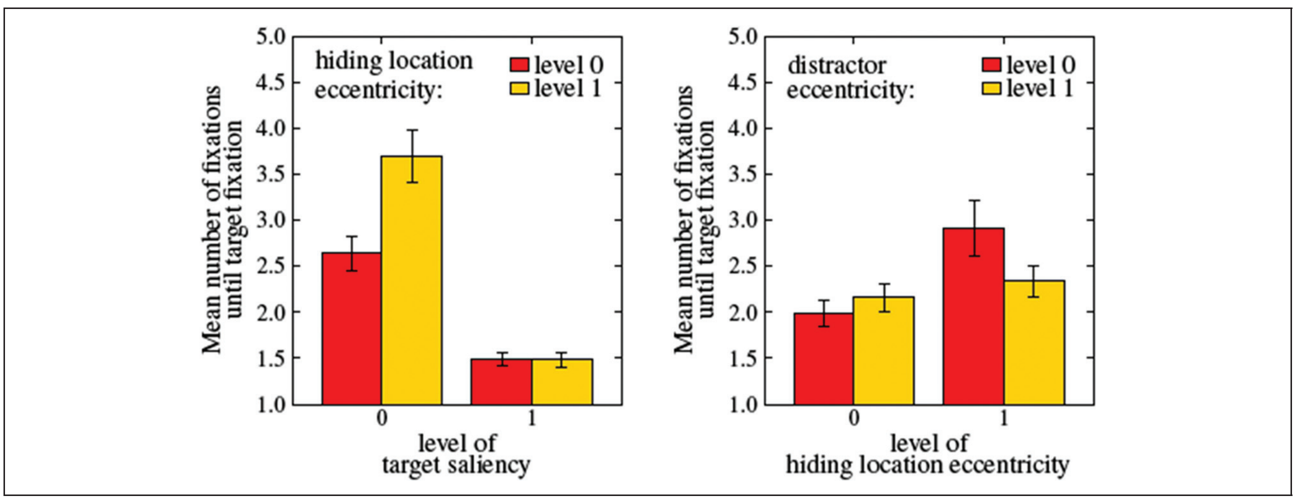

Figure 5. Interaction effects of target saliency with hiding location eccentricity and hiding location eccentricity with distracter eccentricity on the number of fixations until target fixation.

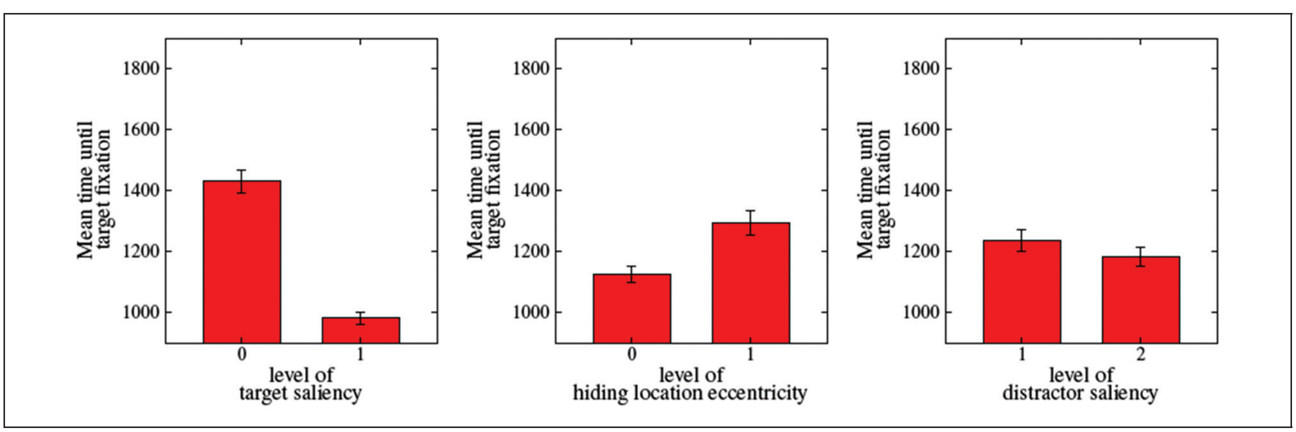

Figure 6. Effects of target saliency, hiding location eccentricity, and distracter saliency on the time until target fixation.

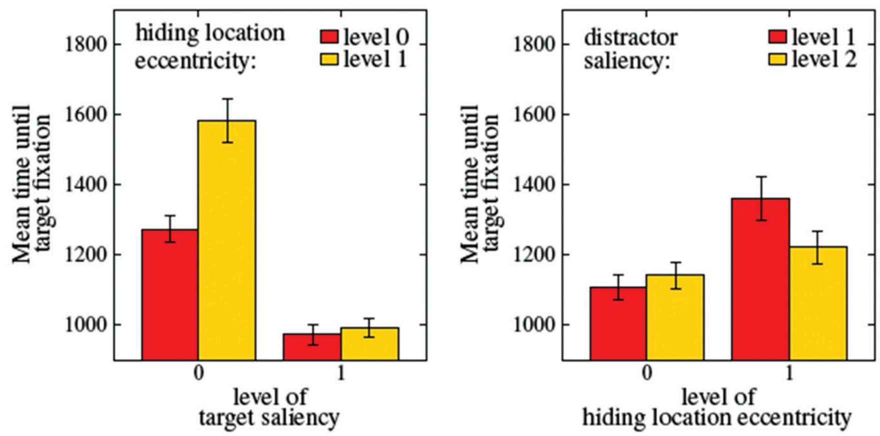

Figure 7. Interaction effects of target saliency with hiding location eccentricity and distracter saliency with hiding location eccentricity on the time until target fixation.

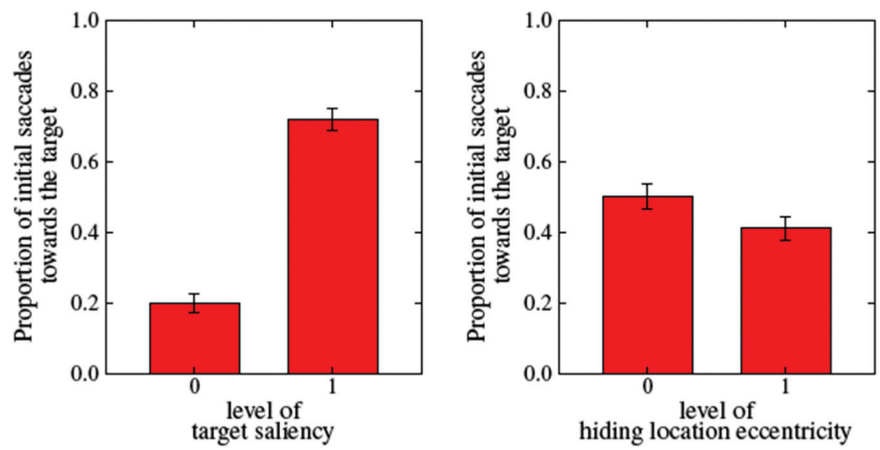

Figure 8. Effects of target saliency and hiding location eccentricity on initial saccade direction. The graphs show the ratio of initial saccade being directed towards the target.

target saliency and eccentricity of the hiding location. Inspection of a plot of the number of fixations until target fixation against hiding location eccentricity grouped by target saliency showed that increasing hiding location eccentricity increased the number of fixations only in the case of low target saliency. There was also an interaction between distracter eccentricity and hiding location eccentricity. The increasing effect of hiding location eccentricity on the number of fixations until target fixation was modulated by the distracter eccentricity. Higher eccentricity of the distracter reduced the effect of hiding location eccentricity (see Figure 5).

The mean time until target fixation was 1209 ms with a standard deviation of 491 ms. The time until target detection showed the main effects of target saliency (p = 0.0001) and hiding location eccentricity ($p < 0.0001$), as well as an interaction of these two factors ($p = 0.0004$), as can be seen in Figures 6 and 7. As target saliency increased, the *time until target fixation* decreased, and as eccentricity increased the *time until target fixation* increased. However, the effect of increased time until target detection with increased hiding location eccentricity vanished for target saliency level 1. This means that at target saliency level 1, the target was easy to spot and the hiding location had no importance for the search task. In addition to that, there was an interaction of distracter saliency and hiding location eccentricity ($p = 0.0372$). The increase of *time until target fixation* caused by the increase of hiding location eccentricity was less as distracter saliency decreased.

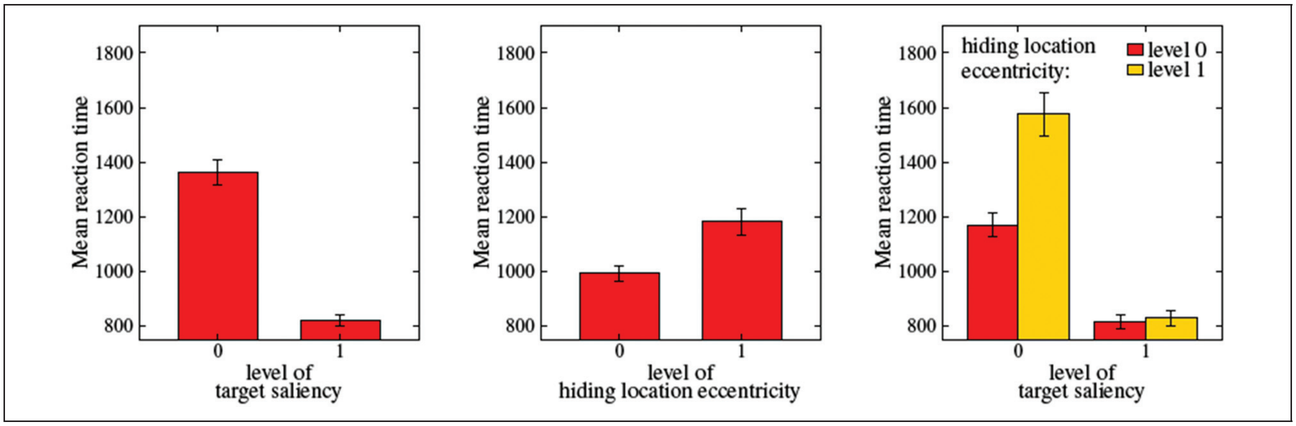

Figure 9. Main effects of target saliency, hiding location eccentricity, and interaction effect of target saliency and hiding location eccentricity on reaction time.

The mean initial saccade latency was 640 ms with a standard deviation of 290 ms. No factor effect on the initial saccade latency was observed. The length of the first on-target saccade showed a mean of 760 pixels with a standard deviation of 362 pixels. This amounts to 15.8 degrees of visual angle and 7.5 degrees of visual angle, respectively. No significant factor effect was observed for the length of the first on-target saccade. The initial saccade was directed towards the target hemifield in 45.7% of the trials. A chi-squared test of initial saccade direction revealed a main effect of target saliency ($p < 0.0001$) and a main effect of hiding location eccentricity ($p = 0.0319$). Increasing target saliency increased the number of initial saccades directed towards the target. Conversely, the number of initial saccades towards the target decreased with increasing hiding location eccentricity (see Figure 7).

The mean reaction time was 1085 ms and the standard deviation 555 ms. Similar to the time until first target fixation, the reaction time showed main effects of target saliency ($p = 0.002$) and hiding location eccentricity ($p < 0.0001$), as well as an interaction of these two factors ($p < 0.0001$). Again, increasing target saliency decreased the reaction time and increasing hiding location eccentricity increased reaction time. This effect was almost nonexistent for target saliency level 1 (Figure 9).

The results of the experiment clearly show that search performance depends on target salience. High target salience reduces search times and numbers of fixations. This means that there is hardly any non-task directed behavior observable, if the target is easy to spot. At the same time, it can be seen that increasing hiding location eccentricity yields worse outcomes in all of the response variables. This shows that search performance declines with increasing hiding location eccentricity.

The opposite is true in the case of increasing distracter eccentricity where search performance increases. This is not very easy to detect and can only be shown indirectly through an interaction between distracter eccentricity and the hiding location eccentricity for the number of fixations until the first target fixation. The *number of fixations until target fixation* increases with increasing hiding location eccentricity. This increase differs depending on the distracter eccentricity. The increase of the *number of fixations until target fixation* with increasing hiding location eccentricity is much stronger when the distracter eccentricity is lower (see Figure 7). This means that the effect of hiding location on search performance is modulated by distracter eccentricity. The higher the distracter eccentricity, the lesser the worsening influence of distracter eccentricity on search performance. This indicates that a higher distracter eccentricity improves the detection performance of a human observer. Looking at the effects of hiding location eccentricity and distracter eccentricity, one can clearly see that hiding location and the visually salient distracter affect search performance in different ways.

In addition, the interaction of target salience and hiding location eccentricity show that the hiding location draws the eyes only if the target is hard to spot. The search performance is completely unaffected by the hiding location eccentricity if the target salience exceeds a certain threshold. Below that threshold, search performance is essentially determined by the hiding location eccentricity.

3.1.3 Tier 1 single-target experiment conclusions. The different effects of the distracter and hiding location on the response variables show that a visually salient distracter and a semantically relevant scene location have different effects on the eye-movement behavior during a target search. Whereas the influence of a distracter was reduced with higher eccentricity, the influence of the hiding location strengthened. This is an indication for a different level of processing of the two. The visually salient distracter is capturing reflexively controlled attention, and this effect apparently wears off at higher eccentricities. The hiding location, on the other hand, did not show this reduction. On the contrary, it showed an increase, and therefore it can be concluded that it does not capture the attention based on reflexive control but based on higher level cognitive processes. Furthermore, the interaction effects of target and hiding location clearly show that both entities affect human search behavior.

However, the effect of hiding location eccentricity is not significant if the target has a high salience. This indicates that the eye-movement behavior of a human is clearly goal directed and not necessarily affected by distracters. If the target cannot be spotted easily, attention is guided to the hiding location due to its meaning for the search task. The hiding location has a higher likelihood of containing the target than any other scene location and therefore it is fixated and closely examined, which can be seen by the increasing number of fixations and time until first target fixation with increasing hiding location eccentricity.

The experimental design used here is an indirect approach of showing the influence of semantically relevant locations on the attention guidance in a search task. It was not examined whether the search performance actually improved when the target was in a semantically relevant location. However, this indirect approach makes the significance of the semantically relevant information visible. It is important to note that participants were not provided with any information telling them that the target should be expected at the hiding location, nor did they receive any training from which they could have learned this association. The association must have been natural to the participants based on either past experience or a general association of doorways and target presence. This means that semantically relevant information can be accessed for attention guidance during any kind of search, given that such semantically relevant information pertaining to the search task exists.

This result is in contrast to the findings of Kunar et al.[13] They claim that there is no improvement on search performance based on repeated presentations of search arrays with the same arrangement, as has been shown by Chun and Jiang.[14] Kunar et al. claim the improvement is a function of response selection. For the presented experiment this possibility can be ruled out, since search benefits due to contextual guidance were not measured. For the same reason, it cannot be concluded that the presence of a semantically relevant cue will actually speed up the search process, but it can be concluded that examining locations associated with target presence is a human search strategy if the target location is not apparent.

It is not surprising that repeated presentations of search arrays do not improve search efficiency. It is hardly conceivable to interpret search array layouts as a meaningful cue. Apparently, there needs to be semantically relevant information content in order to provide effective contextual cueing. The hiding location in this experiment showed this property and therefore it could serve as a semantically relevant distracter.

Similar to the findings presented in this work, Brockmodel et al.[15] show that search in naturalistic scenes is facilitated through recurring global and local context,

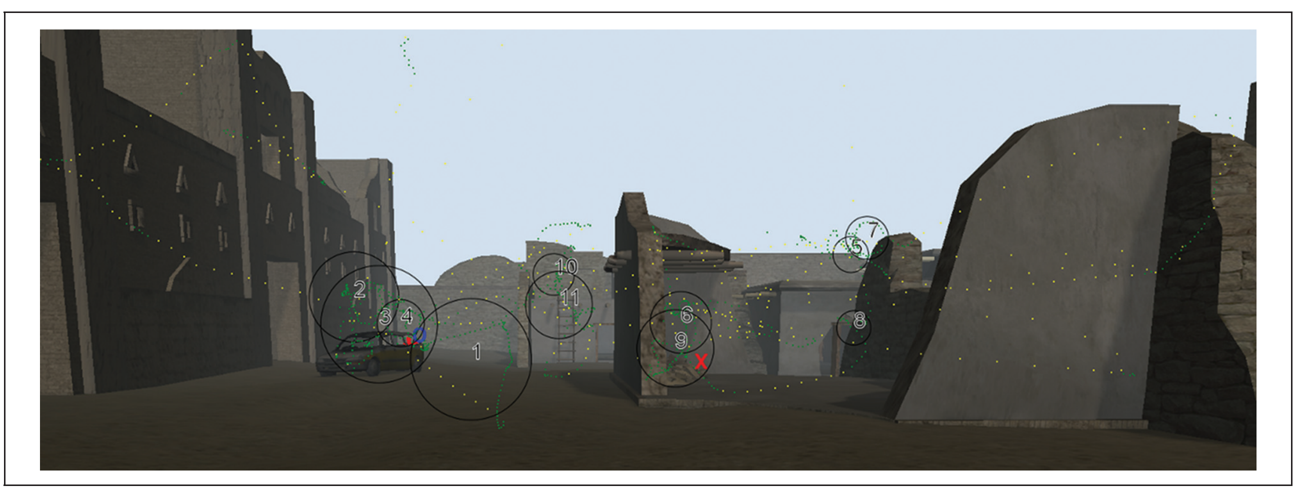

Figure 10. Tier I static experiment eye-tracking results overlay this cropped scene. This scene contained an audible report referencing the car and a moving target behind the car. The small bold circle near the car represents a true positive detection; the X indicates a false positive. Dots represent eye movement. Numbered circles represent long fixations (areas of sustained low velocity eye movement for 20 μs or longer). Long fixations are numbered in chronological order.

with global context being more influential. However, their findings are based on contexts learned for specific scenes only. The results presented here show that eye-movement guidance through semantically relevant information is not constrained to specific pre-learned scene arrangements, but rather relies on stored associations that provide contextual cueing. In other words, the guidance of semantically relevant locations is different in nature as compared to the contextual guidance. Contextual guidance seems to apply to learned co-occurrences of objects only, whereas the guidance of semantically relevant scene locations is based on the meaning of these locations associated with the current task.

3.2 Tier I static experiment

This experiment consisted of 16 scenes presented to 19 participants with a DOE that included four binary factors: a moving target, an audible report, a written situation report (SITREP), and a minimap. Including these four binary factors in a full factorial design equates to 16 design points, with each design point associated with a particular scene. Each scene consisted of an urban environment with a unique field of view and array of structures. The moving target varies from 0.2 to 0.4 degrees in size, and it moved into the scene for approximately 1 second and then moved back out of the scene. The audible report consisted of a brief (5–10 seconds) report indicating a target location with a semantic reference from the scene. Correct interpretation of the audible report directed attention to an area of the scene that was always within the immediate vicinity of a target. In other words, there were no false audible reports. The written SITREP was provided on a sheet of paper that provided several bulleted remarks about the physical nature of the environment and likely enemy courses of action. The SITREP also provided an overhead schematic of the scene. Each scene lasted 20 seconds, and the participants indicated detection of a target with a mouse click on the target location. This experiment occurred immediately following the tier I single-target experiment, and therefore the eye-tracking, hardware setup, and participant conditions were identical to that experiment. An example scene along with eye-tracking results from a single participant is shown in Figure 10.

Studying the eye behavior in Figure 10 enables postulation and qualitative assessment of the behavior, which can be useful in forming hypotheses worthy of quantitative exploration. Notice the strong center bias, possibly a result of the audible reference and movement, two basic aspects of SA included as factors in this experiment. It is also interesting to note the lack of interest in the outer portions of the scene. Minimal ambiguity and close range contribute to a person's ability to pre-attentively clear portions of scenes, resulting in the appearance of ignoring or paying minimal attention to a portion of the scene. Unambiguous, uncluttered, close-range components of scenes received very little attention during this experiment. Assuming time fixated represents areas of perceived threat, the perceived threat resides in ambiguous, cluttered areas and areas where external information indicated immediate threat. In addition to providing a means to refine the experimental protocol and DOE, the tier I static experiment provided an

(a) (b)

Figure 11. (a) A soldier engages targets in the virtual environment. The weapon is an M4 carbine mounted with an interactive laser, and the soldier wears the Mobile Eye system. (b) The video image captured by the eye-tracking system.

introduction to target search behavior and planted seeds for hypotheses that would be more deeply explored in the data from the tier II static experiment.

4 Tier II static experiment

The tier II static experiment took place at the Maneuver Battle Lab (MBL) in Fort Benning, Georgia in April of 2009. A total of 27 infantry soldiers participated over the nine-day experiment with three individual participants per day. Subjects ranged in age from 20 to 42 (mean of 26), and they ranged in years of service from 1 to 23 (mean of 6.12). Nineteen out of 27 subjects had deployed to either Iraq or Afghanistan. The soldiers stood 7 feet from a rear-projected screen that measured 7.5 feet tall and 10 feet wide. The MBL uses the virtual environment software SVS produced by Advanced Interactive Systems Incorporated. This virtual environment interacts with a calibrated weapon to provide feedback and battle damage assessment to the participant. The DOE and protocol for the tier II static experiment were identical to those for the tier I static experiment. A description of this DOE and protocol can be found in the section discussing the tier I static experiment.

The Mobile Eye tracking system, produced by Applied Science Laboratories, recorded eye behavior. The weight of the system is negligible (76 grams), and soldiers stated that the system felt much like the safety glasses worn during combat operations. Soldier view was completely unaided. Contact lenses were permitted, but eye glasses were not. Similar to the tier I experiment, visual acuity and color vision of participants were tested using a modified Snellen chart and the Ishihara color test, respectively. Only participants with an uncorrected vision of 20/30 or better took part in the experiment.

The *Mobile Eye* system calibrates the participant's foveal point with a video frame. The foveal point, represented as a red crosshair, overlays the recorded scene in a video file. Figure 11(b) shows a video frame captured during one of the trials. Notice in Figure 11 the shadow of the weapon muzzle at the bottom of the screen. The dot near the crosshairs is the weapon optic simulation. The red crosshairs represent the foveal point, and the canted image was a common effect of aiming the weapon. Convergence of the weapon optic and the eye-tracking crosshairs during an engagement indicated well-calibrated eye-tracking data. The video files captured one of two data outputs produced by the eye-tracking system. The other data output is a flat file that indicates the position of the red crosshair relative to the video frame. The important measure from the eye-tracking data was the foveal position over time relative to the scene. The flat file provides the foveal position relative to the video frame; the video frame constantly moves as the subject moves. This created a significant post-processing challenge.

To determine the foveal point relative to the scene required an innovative approach to correlate the video file to the flat file output from the software. This non-trivial task included the frame-by-frame parsing of each video file. Each 20 second video file consisting of 30 frames per second (fps), rendering 600 unique images. The total quantity of images for the experiment (16 scenes and 27 participants) was approximately 259,200. The flat file that included the foveal point relative to the video frame provided a means to calculate fixations. Similar to the tier I experiments, periods of eye movement with a velocity less than 12.5 degrees per second were classified as fixations. The frame numbers of the fixation periods could be extracted from the flat file, and the median chronological frame for each set of frames associated with a fixation was used to approximate the fixation location. Manual interpretation of the crosshair location for this median frame was then used to approximate the location in the scene of the fixation. This proved to be a tedious but accurate process, supported to a large degree with some useful mouse and keyboard emulation macros and validated visually by inspecting video files and overlaying scene data with fixation location results (see Figure 12).

Analysis of images marked by fixations provided a method to validate fixation and eye-tracking results. It also created a tool to formulate hypotheses. Data visualization can spark ideas and inform intuition in a unique way. The visualization of results from tier I indicated that semantic meaning affects search. Analysis of images also revealed attraction to edges, corners, and clutter. Ambiguous areas that provide cover and hiding locations attracted the attention of soldiers. This observation became important during the exploration of the data.

4.1 Analysis of tier II static experiment data

The premise of this work is that it is possible to build a model that predicts search behavior, given the characteristics of a target scene. In order to explore this premise, data formulation included independent variables that represented target scene characteristics and a dependent variable that represented eye fixations. Each target scene can be shown as an 1114×598 pixel image. A uniform grid was laid over the top of each scene, creating 784 (28×28) unique areas (each area contained approximately 780 pixels). Let us define point (i,j) as the discrete area in the ith row and jth column of the scene, where $i = (1,\ldots,28)$ and $j = (1,\ldots,28)$. The independent variables for each of these unique areas was measured. The five independent variables are shown in Table 1.

Calculating cv_{ij} first requires line of sight (LOS) calculations to each discrete area in the scene. The LOS is the

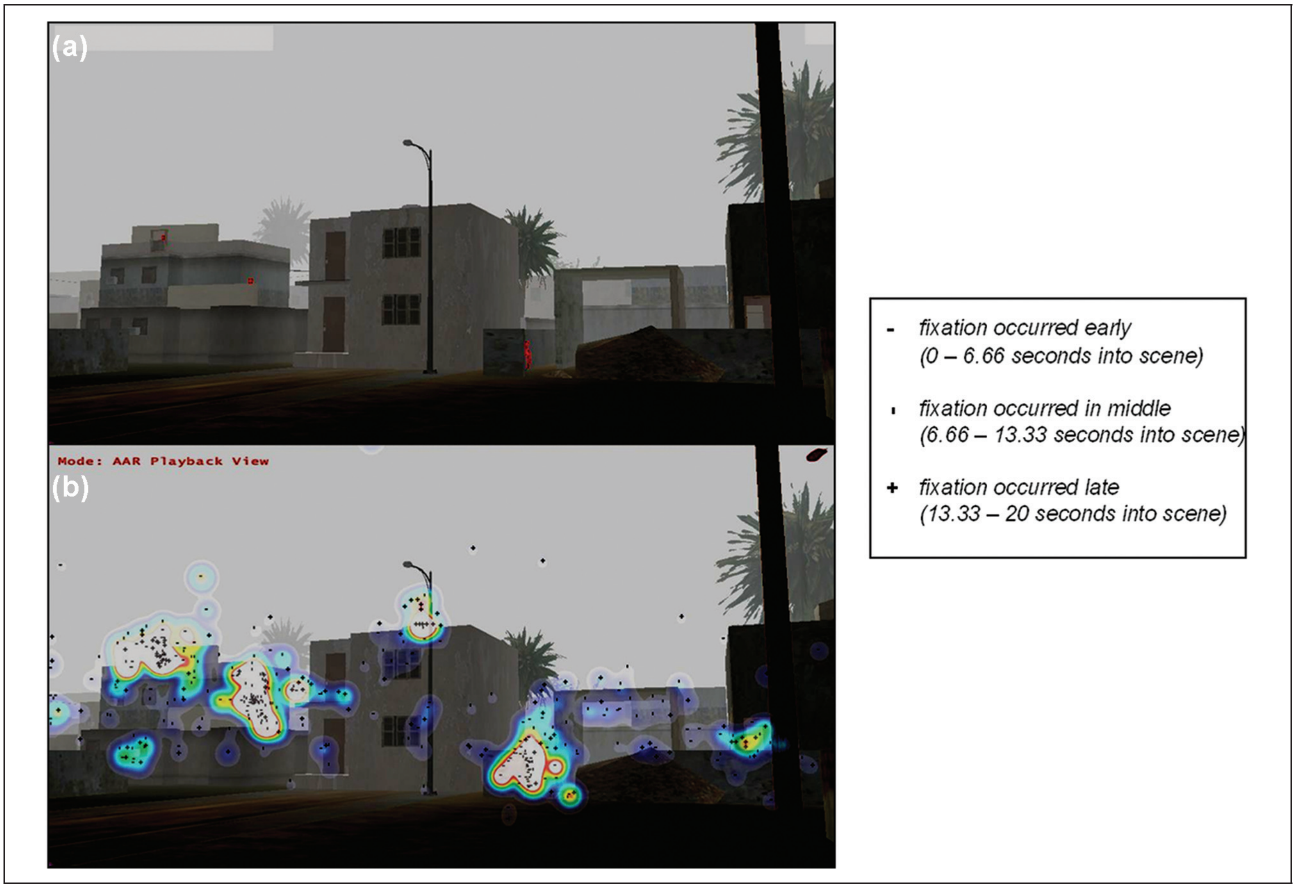

Figure 12. (a) A target scene with targets highlighted in red (targets were not highlighted for participants). (b) Aggregated fixation locations. The underlying heat map reflects aggregate time spent during fixations. (Color online only.)

length in meters of a ray shot from the soldier to the discrete area (LOS to the center point of each area was used to approximate LOS to the area). If the discrete area lands on a building that is 32 meters away in the virtual environment, LOS is 32. If the discrete area is in the sky, the LOS typically measured slightly over 6000 meters. Let LOS_{ij} represent the LOS to area (i,j). cv_{ij} is calculated as follows:

$$cv_{ij} = \sigma_{ij} / \mu_{ij}$$

σ_{ij} is the sample standard deviation of the LOS of the set of areas including and adjacent to (i,j). This includes $LOS_{(i-1)(j-1)}$, $LOS_{(i-1)j}$, $LOS_{(i-1)(j+1)}$, $LOS_{i(j-1)}$, LOS_{ij}, $LOS_{i(j+1)}$, $LOS_{(i+1)(j-1)}$, $LOS_{(i+1)j}$, and $LOS_{(i+1)(j+1)}$. μ_{ij} is the sample mean of this same set. The coefficient of variance is a measure of relative dispersion and a dimensionless measure. Large values for cv_{ij} indicate areas where depth changes significantly. Corners and edges often create an abrupt change in depth or LOS, and the corners and

Table 1. Definition of variables.

Independent variable	Definition
w_{ij}	Distance in pixels from the center of discrete area (i,j) to nearest window.
d_{ij}	Distance in pixels from discrete area (i,j) to nearest door.
m_{ij}	Distance in pixels from discrete area (i,j) to nearest moving target.
a_{ij}	Distance in pixels from discrete area (i,j) to nearest audible cue.
cv_{ij}	Coefficient of variance of the LOS to (i,j).
Dependent (response) variable	**Definition**
f_{ij}	1 if a fixation of any participant occurred within 20 pixels of point (i,j), 0 otherwise.

LOS: line of sight

Table 2. Predictive power by feature.

Instances ranked by w_{ij} and all instances from a scene with a window:	$U = 0.70$
Instances ranked by d_{ij} and all instances from a scene with a door:	$U = 0.65$
Instances ranked by m_{ij} and all instances from a scene with a moving target:	$U = 0.59$
Instances ranked by a_{ij} and all instances from a scene with an audible cue:	$U = 0.63$
Instances ranked by cv_{ij}:	$U = 0.69$

edges that attract the eye in a target scene actually do have semantic meaning. These corners and edges represent intervisibility lines where cover and concealment could be afforded to an enemy soldier. These areas include rooftops, building edges, windows, and doors. Windows and doors are not entirely detected with the coefficient of variance because these portals could be shut or have minimal depth that does not create a significant coefficient of variance signature. The semantic meaning and obvious threat that windows and doors present prompted their inclusion as separate independent variables.

Let us slightly modify the data formulation for the sake of further statistical discussion. Given a total of 16 scenes, each represented by 784 discrete areas, 12,544 instances are created. Each instance is defined by a five-dimensional vector, $\mathbf{x}_i$, $i = (1,\ldots,12{,}544)$, and the five dimensions are the five independent variables defined in Table 1. Each instance has a response, y_i, where $y_i = 1$ or 0 based on the fixation criteria described for f_{ij}. It is possible to feed data formatted this way into prediction models; however, the statistical analysis performed for this research focuses on transparency with regard to the contribution of features and simplicity for the sake of discussion and potential implementation. Advanced pattern recognition models and other numerical methods could be used to explore this dataset; however, that will be left for future work.

Formatting the data in this manner also enables the use of the Mann–Whitney statistic, referred to as U, to calculate contributions of the features and explore performance of various feature combinations. A detailed discussion of this statistic and its uses can be found in Lehmann.[16] Given the matrix $\mathbf{X}\,|\,y$, (12,544 rows and 6 columns), it is possible to rank the instances based on any feature. For example, sorting on the window distance feature now ranks the instances based on their proximity to a window. The U-statistic is then calculated as follows: Let p represent the number of positive instances (fixation occurred), b represent the number of benign or negative instances (no fixation), and R_i represent the rank of the ith instance. For this dataset, $p = 3264$ and $b = 9280$:

$$W_r = \sum\nolimits_{i=1}^{b} R_i$$

$$W_{YX} = W_r - \frac{1}{2}b(b+1)$$

$$U = \frac{W_{YX}}{pb}$$

This method uses the sum of the ranks of the benign instances (it is also possible to use the sum of the positive instances) to calculate a statistic that measures the degree of response segregation in relation to the rank order. Perfect segregation results in $U = 1$. A U of 0.5 would indicate a random ordering of instances. There are two interesting properties of this statistic. The first property is that the statistic equates exactly to the area under the receiver operating characteristic curve, a curve that shows the relationship between the true positive and false positive rates.[17] The second interesting property is actually the meaning of the statistic. The Mann–Whitney U-statistic is the probability that a random positive instance is ranked higher than a random negative instance. This statistic is commonly used to measure the predictive power of binary classification techniques and is very widely used in medical practice and research.[18]

Every scene except one included a window, and the same was true for doors. However, only half of the scenes included moving targets and half of the scenes included an audible cue. Therefore, it is not reasonable to rank all of the instances using one of these variables. Table 2 shows the performance of each feature for the instances when the feature was present in the scene. For example, only one scene did not contain a window. Rank ordering all instances from all other scenes by the window feature, $U = 0.70$. Table 2 summarizes results from all the features.

All of these features were significant ($p \ll 0.01$). The only feature that applied to every scene was the coefficient of variance; however, it was desirable to apply every feature to all of the scenes in an attempt to improve overall performance. Ensembles of classifiers are a common technique to improve classifier performance, and aggregating the features in this data presented an opportunity to build an ensemble.[19] One of the most important considerations when building pattern recognition models or ensembles of models is the data scaling method. Data scaling allows comparison and combinations of data that have different units of measure.

Rankings are a data scaling method, and that was the technique used in this work. It is empirically shown by Evangelista[18] that aggregation of classifiers using ranks yields promising results. Also shown in that work are the merits of various types of aggregators, specifically for unbalanced classification. Three common and simple aggregators include the minimum (min), maximum (max), and average (avg). The field of fuzzy logic explores various aggregators that essentially span the spectrum between min and max. For the data considered in this research, using the min to aggregate the feature rankings created improved results. Aggregating the ranks of all five features achieved a $U = 0.78$. Removing two features, the audible cue and moving targets, $U = 0.79$ (see Figure 13).

4.2 Discussion of tier II static experiment

The tier I and tier II experiments supported many of the initial expectations established at the outset of this project. Humans quickly surmise the semantic meaning of objects in their environment. Human vision aggregates and summarizes objects as a function of contrast; edges, corners, and curves become familiar shapes, and the human brain learns to classify specific combinations of these features that have meaning in our life.[1] The startle that you feel at the encounter of an unexpected human face or large dog is a function of your brain's ability to almost instantaneously aggregate the visual features of these objects. Attraction to doors and windows is a simple example of the effect of this process. Exploring attraction to 'hiding locations', as described by Jungkunz,[4] presents an even more fundamental decomposition of target search. Doors and windows are a subset of hiding locations. These 'hiding locations' are areas that provide cover and concealment. The challenge to the modeling community involves creating a quantity that detects the existence of these areas. In this paper, the coefficient of variance of LOS serves as a proxy to detect these hiding locations. This feature is the only generic predictive variable explored in this paper, and it emerged as a powerful predictor. Audible cues and moving targets should also be included in search methods. Unsurprisingly, both of these features created a significant effect on target search, and both of these variables have recently been accepted and implemented to affect search in combat models.

The results from this paper recommend using the min aggregator, essentially combining these features in a manner that sensitizes the response to the most extreme signal presented by any one of the features. Due to the sparse nature of audible cues and moving targets, run-time integration of these features is not an issue. However, run-time integration of the coefficient of variance of LOS and proximity to windows and doors would not work. LOS calculations already consume a considerable portion of processing required by combat simulation models. Preprocessing environmental characteristics, such as LOS, is one method that could overcome the run-time processing challenges. Preprocessing environmental characteristics could become an important aspect of improving the fidelity of soldier target search.

5 A modeling approach for searching with the unaided eye

The experiments above provide evidence that observers looking for a human target do not spend equal time on all areas in a scene, but instead are biased towards areas that might contain a threat, such as near a door, window, or other edge that can be hidden behind. One approach to representing this phenomenon in current simulations would be to tag all such features in the geometry of the terrain. Such an approach might be both labor intensive and error prone, however. In this section, we describe an alternative approach that requires substantial computer preprocessing, but less human labor. In addition, it uses decision theoretic principles to go beyond modeling just the distribution of gaze points to get at their order as well.

The model proposed in this section represents human visual search (gaze direction) based on decision theoretic principles. This model can include metrics that summarize threat based on environmental characteristics, such as those proposed in the previous section. There is a long history of using decision theory (utility optimization) to produce models of human behavior in areas where our knowledge of fundamental principles is lacking. This work is in that tradition. Specifically, assume that there is a function that describes the observer's subjective likelihood that a target is present at each possible position. The observer desires to maximize the speed with which they find a target. This model is a simple, computationally tractable approach to using the available information to achieve the desired end. The algorithm presented is close kin to general decision theoretic search algorithms, such as the Myopic Search with Discrete Looks algorithm from Washburn.[20]

5.1 Assumptions

Let us assume that we are given T, a finite set of representative target positions. In our work, we produce T by selecting an approximately uniformly spaced set of positions everywhere that a human target could stand, using a previously published automated technique.[21] Let us assume that given that view point of the searcher, the cumulative distribution of detection times, $P_d(t)$, is known for each possible target position. This information requirement matches what is available from existing target acquisition models in the ACQUIRE family. In the case of simulations containing textured three-dimensional (3D) models, such as that used for this study, the detection time model can take into account the detailed appearance of the target and its background from the perspective of the observer as they actually appear in the simulation.[22] For constructive models, such as Combat XXI, the subjective size of the visible portion of the target can be obtained from the simulation, while factors affecting contrast to background are coarsely represented by looking up typical values in a table.

Let us assume that the prior probability (prior to observation) that each position contains a target is known, p_i, for all $i \in T$. This distribution allows us to model the knowledge that the target may prefer rubble piles and rooftops to flat open areas, for example. In addition, it allows us to model knowledge of where a target may be acquired by other means, for example, verbal communication or viewing a Common Operational Picture displayed on a screen. We have described elsewhere how systems we have implemented maintain a threat location distribution that is fully dynamic, accounting both for locations that have been fully or partially searched and for possible movement of the threats.[23,24] If no particular information is available, an initialization of all values to an arbitrary constant can be used.

Standard ACQUIRE-like target acquisition models assume that either a target is in view or it is not (see, however, the two-region version in Reece and Wirthlin[12]). This contradicts the known facts of human vision, specifically that the eye is more capable (at least for motionless targets) towards the center of its field of view. If we consider that all targets on the screen are in view, then there is, from the point of view of ACQUIRE, no reason for shifting gaze at all. We reconcile the facts of vision and the requirements of ACQUIRE by assuming that there is a radius around the center of the field of view such that any target that intersects this radius is subject to detection, while all others are not.

Another issue having to do with applying ACQUIRE-like models has to do with potential targets that are brought into view more than once, having been lost to view due to eye movement in the interim. How should such targets be treated? The two obvious possibilities are either to treat each fresh viewing of the target as independent of the previous ones or to consider the second and later viewings as extensions of the original one. The problem with the former approach is that since target detection is a statistical process, treating each viewing as independent increases the overall chance of detection. Therefore we use the latter approach, which treats two consecutive viewings of duration T_1 and T_2 with an interruption between them the same as a single uninterrupted viewing of duration $T_1 + T_2$.

Finally, this approach considers the effect of the motion of the point of gaze on the time required to find the targets. Once the eye starts to move, it is ineffective at detecting targets until it stabilizes in its new position. The number of milliseconds required is given by $d(\theta) = 37 + 2.7\theta$,[25] where θ is the amplitude in degrees of the change in eye position. Note that at least 37 milliseconds is required to stabilize the eye after even very small movements. Obviously, $d(0) = 0$.

5.2 Algorithm description

Given these assumptions, algorithm development follows. Let the gaze point of the model be initialized in any convenient and appropriate manner. At any time, and for each position, the distribution of detection times for a target in that position is known. This distribution contains all the information we need to determine how much the likelihood that a threat is present at a given position is reduced, given that we observe that position for any given amount of time. We assume that the best the observer can do to optimize his performance is to compare a limited number of gaze trajectories over a finite window of time and select the best. Thus for each possible view point at some update rate, we compute the likelihood of detecting a target if the current gaze point immediately shifts to the view point and remains there for the remainder of the time window. The view point that provides the greatest detection probability is chosen to be the next gaze point. Given that any motion of the gaze point at all results in considerable dead time in which no detection is possible, it will often be the case that the current gaze point scores best and is chosen to be the next gaze point as well. Note that the update rate for evaluation of the view points may be (and usually is) chosen

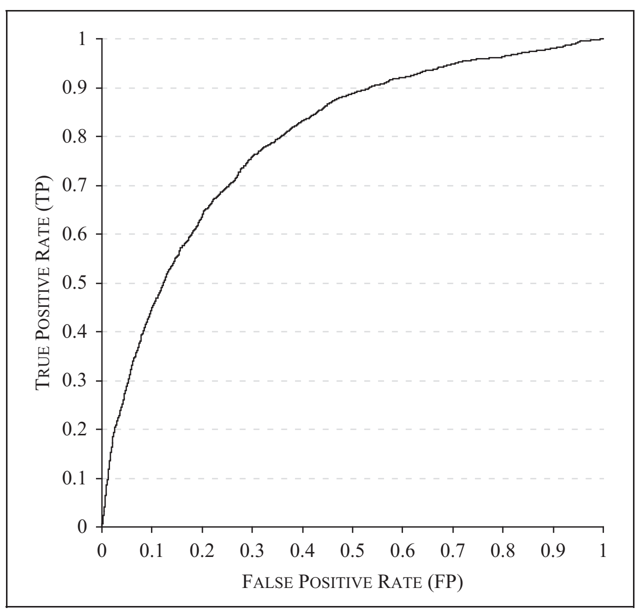

Figure 13. Receiver operating characteristic (ROC) curve for an ensemble of ranks of w_{ij}, d_{ij}, cv_{ij}. Area under the curve (AUC) = U = 0.79.

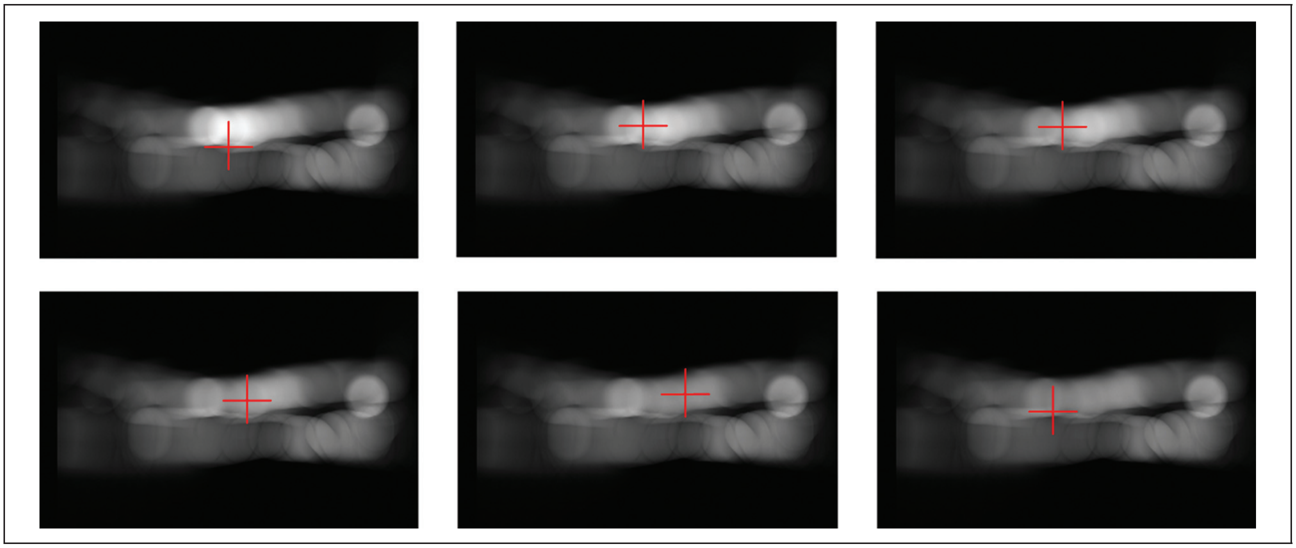

Figure 14. Example run of the algorithm. The crosshair represents the gaze point. The first six gaze points are shown.

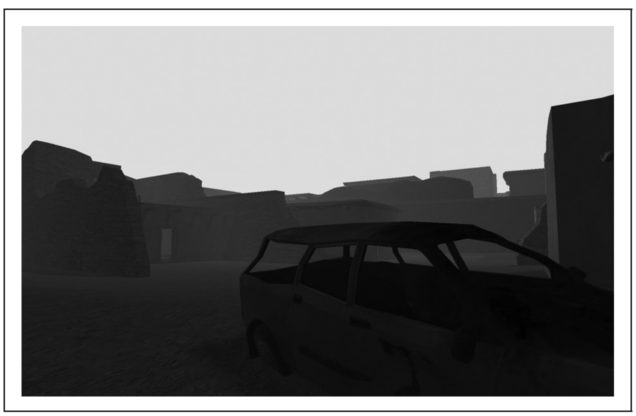

Figure 15. Example scene.

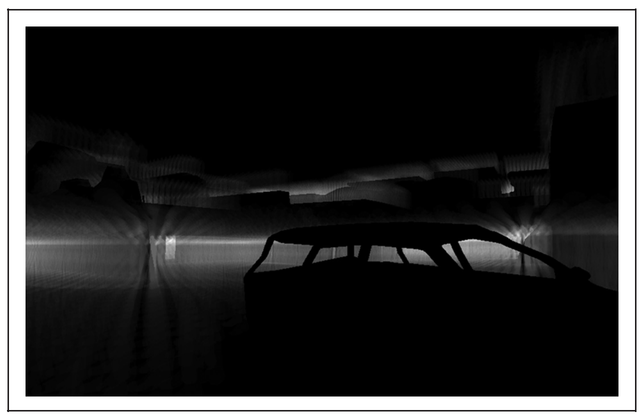

Figure 16. Ghost images show all target positions considered for one particular scene. The silhouette of a vehicle is visible in the foreground.

so that there are multiple evaluations in the time window used to evaluate the view points, so just because a view point is chosen as being the best performer over the next time window does not rule out another view point being chosen as superior part way through the window.

Three important parameters of this model deserve attention: the angle around the view point used to determine which positions are in view (φ), the length of the time window used for scoring alternative view points (L), and the update frequency of the algorithm (f).

5.3 Algorithm pseudocode

1. Initialize the prior probability that a target is at each position (p_i) as appropriate.
2. Initialize the total viewed time for all target positions to zero ($t_i = 0$ for all $i \in T$).
3. Initialize the gaze position to the currently viewed point (x, y).
4. For all possible gaze positions, compute the probability of target detection over the next L seconds if that position is chosen. Let θ be the angle between the current gaze position and the one under consideration. Let S be the subset of target positions T such that any point of a target at that position would be within φ of the gaze position. Taking saccade time into account, the time that would be spent observing that gaze point if it is selected would be $\Delta t = L - d(\theta)$. The probability

of detecting a target at this gaze point in this amount of time is $\sum_{i \in S} p_i(P_d(\Delta t + t_i) - P_d(t_i))$.

5. Saccade to the gaze point with the best detection probability.
6. After time $1/f$ has passed, go to step 4.

5.4 Implementation details and example run

In order to make the algorithm run faster, we pre-compute and store the parameters of the detection time distribution for each target position. We use a simple exponential model for target detection rather than a standard model, such as ACQUIRE, in order to keep the model as simple as possible and avoid unnecessary distribution restriction. Figure 14 shows the first few saccades predicted by running the algorithm on one of the scenes used in the experiments previously described (the same as was used for Figure 15). The grayscale image codes the total probability of detecting a target if the gaze is fixed at that pixel forever. The crosshair shows the gaze position. Regions that are selected for view become progressively dimmer as the remaining probability of detecting a target is reduced.

Human performance data exists to validate this model. Our plan is to compare, at minimum, the fixation duration and saccade length distributions generated by the model to human performance data. The model could be recast as a discrete event model, removing the need for regular algorithm updates, some of which may be unnecessary. Note for example that the second and third gaze points in Figure 16 are identical. The cost for this is inverting the detection time probability distribution function so as to be able to calculate when a new view point will become superior to the current gaze point. It is also possible that a coarser algorithm, but one that could run much faster, could be generated by only considering a small number of view points and pre-caching data about the targets that are visible from those few view points. There is no reason why some kind of graded response function could not be imposed on top of an ACQUIRE-like target acquisition model to make acquisition less likely the further a target is from the center of a field of view, as is believed to be the case for the human eye.

6 Future work and conclusion

Opportunities for future work exist throughout this body of research. The data presented in this paper and other eye-tracking data deserves closer examination, detailing characteristics such as fixation duration, saccade directions, and fixation ordering. Driving searches based on 'hiding locations' is intuitively attractive and supportable; however, quantifying hiding locations in a manner that is accessible to simulation models needs refinement. The technique proposed in this paper, analysis of the coefficient of variance, represents one possible metric. LOS processing currently creates a significant burden on combat simulation models, and efforts to improve LOS determination would be a worthy effort. Efficient LOS processing or preprocessing LOS from multiple locations also presents an opportunity to support automated planning and decision making in combat models and simulation, another active research area.

In conclusion, this paper sought to present compelling experimental results that support the inclusion of pre-attentive processing, environmental characteristics, and contextual information in soldier search models. Secondly, the modeling approach proposed in this paper advocates a simplistic yet theoretically supported construct that has been prototyped. Although a thorough understanding of unaided search remains incomplete, sufficient evidence exists that the current unbiased representation is not adequate. The evidence and results presented propose a solution that could advance the state of soldier modeling and improve the validity of combat simulation models.

Acknowledgement

The intellectual contributions of the following individuals significantly contributed to the thoughts and results expressed in this paper: Nita Miller, Larry Shattuck, Patrick Carrillo, Kyle Tuttle, Jeffrey Thomas, Matthew Hasting, Tim Chung, Dave Hudak, Jon Alt, Eric Tollefson, Jack Jackson, Chen Lai, Andrew Regan, John Mazz, and Jim Starling.

Funding

This work was supported by the Army Modeling and Simulation Office.

Conflict of interest statement

The views expressed in this paper are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

References

1. Hoffman DD. *Visual intelligence*. New York: W.W. Norton and Company, 1998.
2. Medina J. *Brain rules*. Seattle: Pear Press, 2008.
3. Doll T and Home R. Lessons learned in developing and validating models of visual search and target acquisition. In: *Proceedings of the RTO SCI Workshop on Search and Target Acquisition*, Utrecht, The Netherlands, 21–23 June 1999.
4. Jungkunz P. *Modeling human visual perception for target detection in military simulations*. PhD. dissertation, Naval Postgraduate School, 2009.
5. Bruce NDB, Loach DP and Tsotsos JK. Visual correlates of fixation selection: a look at the spatial frequency domain. In: *Proceedings of the IEEE International Conference on Image Processing*, San Antonio, 16–19 September 2007.

6. Johnson J. Analysis of image forming systems. In: *Proceedings of the Image Intensifier Symposium*, US Army Engineer Research and Development Lab, Ft. Belvoir, VA, DTIC report AD 220 160, 1958.
7. Vaughn BD. *Soldier-in-the-loop target acquisition performance through 2001: integration of perceptual and cognitive models*. US Army Research Laboratory Technical Report ARL-TR-3833, 2006.
8. Entin EB and Entin EE. Situation awareness and human performance in search and target acquisition. In: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 1995: Intelligent Systems for the 21st Century*, Vol. 4, 1995, pp. 3833–3837.
9. Zelinsky GJ. A theory of eye movements during target acquisition. *Psychol Rev* 2008; 115: 787–835.
10. Mazz JP, Kistner RW and Pibil WT. Detection of low-contrast moving targets. In: *Proceedings of the RTO SCI Workshop on Search and Target Acquisition*, Utrecht, The Netherlands, 21–23 June 1999.
11. Treisman A. Features and objects in visual processing. *Sci Am* 1986; 255: 114–125.
12. Reece DA and Wirthlin R. Detection models for computer generated individual combatants. In: *Proceedings of the 6th Conference on Computer Generated Forces and Behavioral Representation*, University of Central Florida, 1996.
13. Kunar MA, Flusberg S, Horowitz TS and Wolfe JM. Does contextual cuing guide the deployment of attention? *J Exp Psychol Hum Percept Perform* 2007; 33: 816–828.
14. Chun MM and Jiang Y. Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognit Psychol* 1998; 36: 28–71.
15. Brockmodel JR, Castelhano MS and Henderson JM. Contextual cueing in naturalistic scenes: global and local contexts. *J Exp Psychol Learn Mem Cognit* 2006; 32: 699–706.
16. Lehmann EL. *Nonparametrics: statistical methods based on ranks*. San Francisco, CA: Holden-Day, Inc., 1975.
17. Hanley JA and McNeil B. The meaning and use of the area under the receiver operating characteristic curve. *Radiology* 1982; 143: 29–36.
18. Evangelista PF. *The unbalanced classification problem: detecting breaches in security*. PhD dissertation, Rensselaer Polytechnic Institute, 2006.
19. Kuncheva LI. *Combining pattern classifiers: methods and algorithms*. Hoboken, NJ: John Wiley and Sons, Inc., 2004.
20. Washburn A. *Search and detection*. 4th ed. Institute for Operations Research and the Management Sciences, 2002, Linthicum, MD, USA.
21. Darken C. Level annotation and test by autonomous exploration. In: *Proceedings of Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, 2007.
22. Darken C. Computer graphics-based models of target detection. *The J Defense Model Simulat* 2007; 4: 262–276.
23. Alt J and Darken C. A reference model of soldier attention and behavior. In: *Proceedings of Behavior Representation in Modeling and Simulation (BRIMS)*, 2008.
24. Darken C and Anderegg B. Particle filters and simulacra for more realistic opponent tracking. In: Rabin S (ed.) *Game AI programming Wisdom 4*. Boston, MA: Course Technology, 2008, pp. 419–429.
25. Baloh RW, Konrad HR, Sills AW and Hornrubia V. The saccade velocity test. *Neurology* 1975; 25: 1071.

Author biographies

Paul Evangelista is a major in the US Army and an Operations Research Analyst at the Training & Doctrine Command (TRADOC) Analysis Center, Naval Postgraduate School, Monterey, CA. He received his PhD from Rensselaer Polytechnic Institute in 2006.

Christian Darken is an associate professor of computer science at the Naval Postgraduate School, and an academic faculty member of the MOVES Institute. His research focus is on cognitive and behavioral modeling for simulations. Previously he was project manager at Siemens Corporate Research. He was also the artificial intelligence (AI) programmer on the team that built Meridian 59, the first 3D massively multiplayer game. He received his PhD from Yale University in 1993.

Patrick Jungkunz is an officer in the communications and information systems section of the German Naval Office. In his free time he is conducting research on cognitive and behavioral modeling for simulations. He received his PhD from the Naval Postgraduate School in 2009.