Designing, Developing, and Validating an Adaptive Visual Search Training Platform

Kelly S. Hale[1]($\boxtimes$), Katie Del Giudice[1], Jesse Flint[1],
Darren P. Wilson[2], Katherine Muse[3], and Bonnie Kudrick[3]

[1] Design Interactive, Inc., Orlando, FL, USA
{Kelly,Katie,Jesse.Flint}@designinteractive.net
[2] Department of Homeland Security, Houston, USA
Darren.Wilson@HQ.DHS.GOV
[3] Transportation Security Administration, Arlington, USA
{Katherine.Muse,Bonnie.Kudrick}@TSA.DHS.gov

Abstract. Transportation Security Officers (TSOs) are at the forefront of our nation's security, and are tasked with screening every bag boarding a commercial aircraft within the United States. TSOs undergo extensive classroom and simulation-based visual search training to learn how to identify threats within X-ray imagery. Integration of eye tracking technology into simulation-based training could further enhance training by providing in-process measures of traditionally "unobservable" visual search performance. This paper outlines the research and development approach taken to create an innovative training solution for X-ray image threat detection and resolution utilizing advances in eye tracking measurement and training science that provides individualized performance feedback to optimize training effectiveness and efficiency.

1 Introduction

The process of visual search involves finding a target in a field of non-targets or distractors. Visual search is used in a variety of professions, including radiography, surveillance, security, and inspections. It is one of the primary responsibilities of Transportation Security Officers (TSOs) working at airports throughout the United States. Considering that the visual search TSO process alone involves searching for, identifying, and removing potential threat items within passenger carry-on bags, the challenges associated with this task are numerous. First, images presented to TSOs are two-dimensional X-ray interpretations of three-dimensional bags, which differ from photographic visual representations to which most people are accustomed [1]. X-ray images provide information about the inner structure of the object and depict density, whereas photographs created by light reflection provide information about an object's surface. TSOs must thus learn to recognize threats based on an unnatural visual representation, and may utilize different feature components than those that would be used in a real-world visual search task. In addition, the extensive list of prohibited items [2] requires high-cognitive load and attentiveness to evaluate features of a presented image against an extensive internal mental 'library' of features indicative of threat items. Given the low likelihood of encountering a threat during operations compared to the

number of bags screened during a given shift, challenges with sustained attention [3] and vigilance can also impact performance. Furthermore, TSOs operate in a high-stress environment—both physically (e.g., noise, lighting) and psychologically stressful (e.g., passenger throughput, knowing the potentially catastrophic consequences of a missed threat, seeing line lengths increasing).

Because of the complex nature of the task, training TSOs to effectively and efficiently search each carry-on item passing through airport checkpoints requires extensive time and resources. Current training practices include traditional classroom instruction such as lectures, videos, and simulation-based training and mentor-supervised, on-the-job training. Current evaluation methods are often limited to observable behavioral metrics (e.g., detections, false alarms), which are challenging when trying to identify root cause (s) of performance errors (e.g., scan vs. recognition error) and associated influencing factors (e.g., threat type, location, orientation, clutter in bag, etc.). Current evaluation methods need to quantitatively measure details of cognitive states (e.g., inattention) that could negatively affect training outcomes. Due to the challenges with current data collection techniques, diagnosis, and feedback based on *process-level* measures of trainee visual search performance, evaluations and feedback provided by instructors may not address the underlying sources of the trainee's performance level.

Implementing new training methods from visual search research and leveraging emerging technologies can assist in improving the training process and maximize training efficiency and effectiveness. There are multiple reasons this opportunity for improvement exists, including (1) the challenge for an instructor to detect all trainee visual search errors due to the high workload associated with monitoring a complex scenario; (2) the challenge of the instructor to monitor subtle physical behaviors such as scanning patterns or cognitive processes; and (3) the challenge of the instructor to accurately and reliably identify key patterns into which visual search errors fall. As a result, traditional instructor-led training systems may not be capable of identifying the root cause of visual search errors within the baggage screening process.

The training system discussed in this paper was designed to address this gap in performance evaluation by providing instructors with the capability of diagnosing the root cause of performance errors using real-time measures of visual search. The goal of this innovative research and development (R&D) effort was to identify advances in visual search measurement and training science that could be incorporated into a simulation-based training platform to substantially enhance training effectiveness and efficiency of X-ray baggage screening. One advancement was to integrate visual search process measures that 'peer into the mind of the users' to capture perceptual and cognitive processes not otherwise accessible and that are capable of providing quantitative metrics to evaluate trainee cognitive state throughout a training session. Such measures can provide a more comprehensive assessment of progress as a trainee advances through visual search training, and can identify the root cause(s) of training deficiencies/inefficiencies. By understanding specific error patterns in the visual search process, specialized training strategies and training content can be implemented that are tailored to the individual trainee's needs. The resultant training system uses eye tracking measures, which capture where a trainee focused, in conjunction with meta-data available within the image being reviewed, to provide instructors and trainees with

individualized feedback such as visual search patterns overlaid on X-ray images and a summary of accuracy across various threat types. This improvement in data collection, diagnosis, and feedback of process-level measures of an individual trainee's real-time visual search performance will address the underlying sources of poor performance in the training process and improve performance.

2 Background

A first primary step in visual search tasks, as well as many other cognitive processing tasks, is attention [4]. Within complex visual scenes, attention is used to select and modulate information based on behavioral relevance to appropriately deal with the problem of too much information. The process of focusing attention involves both parallel and serial mechanisms [5], and this concept forms the basis of many theories of attention. For example, Treisman and Gelade [6] proposed a feature integration theory of attention that follows the idea of guiding attention to a specific location based on a number of underlying processes and factors. Their theory includes a two-step process: an initial, preattentive parallel search where multiple features are registered automatically and a slower, serial search where focal attention is processed to determine what is visible at that location. The Guided Search Theory [7] maintains the two-stage process with the distinction that the preattentive stage guides attention to select appropriate objects for the second stage.

Follow-on work and theoretical development has led to a general consensus that attention is primarily driven by one of two processes: stimulus-directed or goal-directed [8–10], where attention is guided either by the stimulus itself or by the goals inherent in the observer completing the visual search. Building on this notion of two distinct control mechanisms of attention, Chun, Golomb, Turk-Browne [11] created a taxonomy of Internal and External attention. External attention refers to a stimulus-driven mode for selection, where—within a visual search task—attention can be directed to spatial locations or time points (assuming some dynamic aspect to the task) alone, or to features or objects that can be selected across space and time. Within this stimulus-driven process, there are two distinct visual search strategies that have been proposed in the literature. The first strategy is Exogenous Search, where specific aspects of the visual scene are captured based on "hard wiring" of humans [10]. In other words, there are certain features that draw attention naturally, such as color, spatial location, and orientation. A second stimulus-based strategy is Endogenous Feature-Based Search, which has been termed habitual [10]. Here, features stand out "automatically" based on specific task knowledge or experience. In the context of baggage screening, there are certain features that are indicative of potential threats (e.g., sharp edge, etc.), and through experience and training, these features naturally 'pop-out' of the visual scene. Internal attention is driven by internal cognitive processes and pulls from representations in working memory and long-term memory based on task rules, decisions, and responses [11]. Here, an Endogenous Goal-Directed Search Strategy is employed where areas within a visual scene are evaluated against specific attentional/perceptual sets to assess relevance to the overarching task goal. In the

context of baggage screening, specific areas of the image are compared against known threat 'maps' stored in long-term memory to assess for similarity.

The theoretical framework presented in Fig. 1 builds from Treisman and Gelade's [6] Feature Integration Theory model, and incorporates Chun's [11] taxonomy of Internal and External attention to frame two key search strategies relevant for carry-on baggage screening: exogenous search and endogenous search, with the latter being further subdivided into stimulus driven endogenous search types—feature-based, position-based, and scene-based—and goal-directed endogenous search. Although the literature shows differing support for which strategy comes into play under what circumstances, Neskovic & Cooper [12] promotes that fixations are initially driven by stimulus features, while subsequent fixations are constrained/focused by cognitive expectations during the recognition process. More recent summaries have proposed a 'guiding representation' that guides attention, but is not itself part of the perceptual pathway [13]. Within the guiding representation, a number of attributes have been identified that guide attention without reference to specific pathways (serial or parallel, preattentive). This theoretical model served as the foundation for the training system outlined in this paper, guiding training objectives and system requirements to optimize visual search and detection skills training.
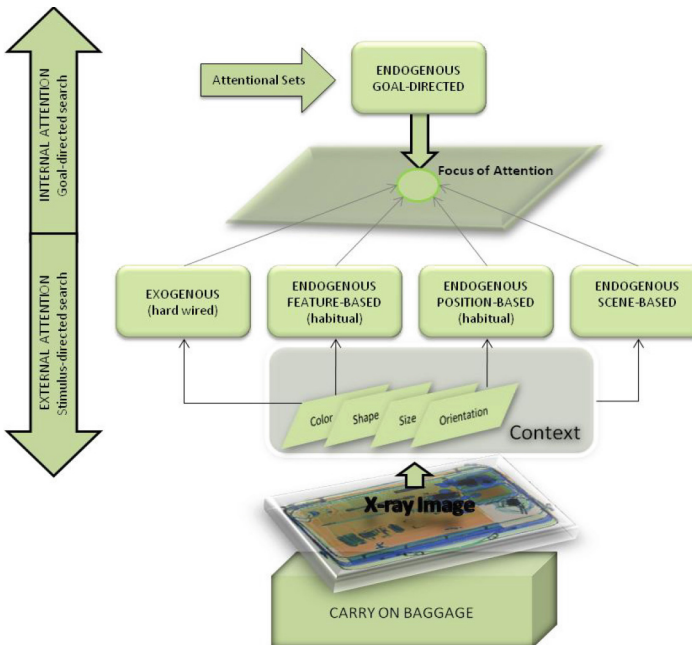
Fig. 1. Conceptual model of visual search strategies for carry-on baggage screening evaluation

3 Development Process

A user-centered training design and development approach [14] was utilized to develop an adaptive visual search training system. This process included: (1) identification of training needs and objectives for the system, (2) system requirements identification, (3) system specifications and architecture development, and (4) graphical user interface and database design. Once the system design was complete, an agile software development process was used to translate the design into functional training software.

3.1 Identification of Training Needs and Objectives

As a first step in the development process, subject matter expert (SME) interviews were completed with representatives from Transportation Security Administration (TSA) to fully understand their current operational and training methods, including the types of visual search challenges included in carry-on baggage screening. The data collected during the SME interviews was used to develop a Concept of Operations for initial system conceptualization and identification of high-level system requirements. Table 1 provides the high-level system requirements developed from SME interview data. This data was also used to identify measures of training effectiveness for both initial training and refresher training that could be used in future system validation.

Table 1. High-level system requirements

A closed-loop, adaptive training system that adapts in real time to user needs based on training strategy, content, and difficulty level
After action review sessions that allow trainees to identify and target areas for improvement
Content control options that allow instructors to maintain up to date images
A flexible image presentation system that reduces image repetition and encourages image novelty
Incremental difficulty levels that allow trainees to advance from novice to expert
Training sessions that allow trainees to increase their exposure to threat images
Training sessions that allow trainees to focus on fine details that distinguish threats from non-threats
System sensors that are portable, lightweight, non-invasive, in an easy-to-use form-factor, and deployable at a reasonable cost

3.2 Developing Detailed Design

Traditional observable training measures do not provide the granularity necessary to diagnose the root cause of visual search performance issues in order to effectively adapt training to address an individual trainee's needs. During this stage of the design process, a combination of observable measures and measures of perceptual and cognitive processes were identified. These measures included both outcomes, such as response accuracy, and processes, such as cognitive activity. An initial analysis of the cognitive

activity measurement domain indicated that real-time sensing of task engagement, target awareness, visual attention, and alertness levels could be used to trigger adaptive training interventions.

A literature review was completed that focused on innovative behavioral sensors (e.g., eye tracking), physiological sensors, and brain-based technologies for each process measure. Each sensor category was evaluated to determine the feasibility or ease of near-term (less than three years) deployment within a TSO training environment. Remote eye tracking and electroencephalography (EEG) were considered as sensor inputs for providing real-time, process-level measures of performance. EEG is still considered the preferred method for ambulatory cognitive state sensing due to its relative ease of deployment and high-temporal resolution. However, although commercial EEG systems are available, the technology is not currently suited to the TSA environment due to the high procurement and maintenance costs. Remote eye tracking is a non-invasive method to assess multiple states such as loci of visual attention (gaze position), and level of alertness (blinking behavior). Eye tracking sensors meet the high-level requirements of being portable, lightweight, non-invasive, in an easy-to-use form-factor, and deployable at a reasonable cost, and provide the following measures related to visual perceptual processes [15]:

- Number of overall fixations – inversely correlated with search efficiency
- Gaze percent on each area of interest (AOI) – longer gazes equated with importance or difficulty of information extraction
- Mean Fixation Duration – longer fixations equated with difficulty of extracting information
- Number of fixations on each AOI – reflects importance of that area

The above eye tracking metrics were integrated with trainee responses (threat detection outcomes) to classify each visual image search using an adapted signal detection theory category. Traditional signal detection theory [16] separates visual search responses into four distinct categories in which a searcher can correctly identify a threat (hit), fail to identify a threat (miss), mistakenly identify a safe item as a threat (false alarm), or correctly clear a bag with no threat (correct rejection). These categories are effective at determining the sensitivity of the observer (how good they are at detecting threats). However, these categories are only intended to classify errors at a level needed to determine sensitivity, and are not appropriate for determining the root cause of errors. Including eye tracking metrics in the assessment of performance allows for a more granular breakdown of the miss category to distinguish searches in which the observer fixated on the target and failed to recognize that it was a targeted search item (recognition error), and searches in which the observer did not fixate on the location where the targeted search item was located (scanning error) (Table 2). Recognition and scanning errors can be used to diagnose root cause of errors to develop more focused training.

A diagnostic and adaptation framework was created to monitor eye tracking and behavioral performance in real time, and to determine (1) when, where, and why performance inefficiencies/deficiencies occurred for a given individual; (2) when to continue practice opportunities to increase performance efficiency; and (3) when to advance training to the next stage based on the trainee achievement at a defined task

Table 2. Adapted signal detection theory categories

Threat Present	Behavioral Response	Eye fixation on area of interest	Signal Detection Theory Category
No	No threat	__	Correct Rejection
No	Threat	__	False Alarm
Yes	No threat	No fixation	Miss – scan
Yes	No threat	Fixation	Miss – recognition
Yes	Threat	No fixation	Hit
Yes	Threat	Fixation	Hit

difficulty level. Building on the developer's experience with mitigation strategies [17], training system design [18], and discrimination training theory [19], the current effort conceptualized targeted After Action Review strategies for optimizing and individualizing training to advance training effectiveness and efficiency. At the current stage of development, two types of training have been developed to address the underlying needs of each type of visual search performance error. If a pattern of scan misses are detected, the system provides *exposure training*, which provides the opportunity to view threats and learn what threat items look like when X-rayed. In contrast, if a pattern of recognition misses are detected, the system provides *discrimination training*, which involves pairs of targets with or without salient differences presented in two separate side-by-side bag images. Discrimination training allows trainees to focus on the details of threat items that will enable them to distinguish between threats and non-threats in X-ray images.

These feedback strategies were designed to summarize performance process and outcome measures relative to targeted training goal(s), and the strategies provide suggested next training steps (e.g., proceed to higher difficultly, train specific need [orientation of threat or type of threat that is routinely missed], etc.). These displays may aid instructors in determining readiness for operations, while also allowing individual screeners to train independent of instructors. In addition, adaptive content features were integrated into the training system. Instructors also have the capability to upload updated training images. This provides the user with the capability to dynamically respond to specific operational needs and keep training current and relevant to the threat and security environment.

3.3   Iterative Development Using Agile Process

An agile development process was used to develop the training software. Agile development allows system development to occur in sprints, with development priorities being set at the beginning of the sprint cycle and a working software build being released at the conclusion of each sprint cycle. Two-week sprints were used throughout the development lifecycle, as this allowed adequate time to address the development priorities set at the beginning of the sprint and time for quality assurance testing of each released build. Some of the many benefits of the agile development process as opposed

to traditional waterfall or spiral development models include: (1) flexible prioritization of requirements throughout the development lifecycle, (2) a working version of software maintained throughout development that can be used for user testing and feedback, and (3) additional time for quality assurance testing and resolution of identified bugs during sprints instead of at the conclusion of development.

3.4   Empirical Evaluation of Training System Effectiveness

To empirically evaluate training effectiveness, lab-based and field-based studies were completed. Lab-based studies focused on examining how the addition of eye tracking impacted the adaptive training paradigm, and were used to help develop the training platform and content [20, 21]. After initial system development was complete, a training effectiveness evaluation was conducted in the field. Working with the customer, an experimental design was developed that could be executed within operational constraints of space, time, and resources. The effectiveness evaluation was conducted with approximately 128 TSOs across three (3) airports using pre/post-test evaluation of visual search performance compared to control group training sessions. After engaging

Table 3.   Critical system development components

Based on theoretical foundations of visual search
Identification of the method by which eye tracking metrics could provide meaningful feedback
Customer input was solicited early and often to: -   validate high-level requirements -   inform operational constraints that will limit end-user adoption (in this case, EEG) -   evaluate feedback screens
Having an internal champion to: -   identify key stakeholders within the organization -   provide SMEs -   coordinate field testing experimentation
Use of agile process to: -   provide transparency to customer -   decrease time to integrate new functionality -   consistently refine and reprioritize based on fluctuating needs
Empirical evaluation of the training effectiveness of the system

in a pretest consisting of 100 X-ray bags to determine baseline performance at the image analysis task, each TSO was then exposed to 4.5 h of training across five consecutive days on the newly developed software or control training software. Results indicated that the training session that used the newly developed software resulted in significantly lower false alarm rates and time to identify threats when compared to the control group.

4   Summary

The R&D process outlined in this paper led to successful development of an innovative simulation-based training system designed to enhance visual search training for carry-on baggage at airport checkpoints. Table 3 highlights the components of the development process that were critical to the successful creation of the system.

5   Conclusion

This paper outlined the R&D process utilized to design and develop an innovative, adaptive training system for visual search. While the focus in this effort was on carry-on baggage screening, the training system framework is applicable to other domains such as radiology, law enforcement, and the military. A user-centered design process was essential to the success of the system, as key stakeholders' and end-users' feedback was captured throughout the effort to adapt and refine system design to meet their needs within operational constraints. Implementing an agile development process allowed for early and often stakeholder review, ensuring design elements discussed were integrated effectively into the end system. The research and development of the resulting system was sponsored by the Department of Homeland Security Science and Technology Directorate's Homeland Security Advanced Research Projects Agency and the TSA Office of Security Capabilities. The system was put through a training effectiveness evaluation in collaboration with the TSA Office of Training and Workforce Engagement, and has been positively received by TSOs.

References

1. Hilscher, M.: Performance Implications of Alternative Color-Codes in Airport X-ray Baggage Screening. University of Central Florida Student Dissertation (2005)
2. TSA.: Prohibited Items List (2011). http://www.tsa.gov/travelers/airtravel/prohibited/permitted-prohibited-items.shtm. Accessed 20 February 2011
3. Green, M.: Inattention blindness and conspicuity (2004). http://www.visualexpert.com/Resources/inattentionalblindness.html. Accessed 8 April 2011
4. Chun, M.M., Golomb, J.D., Turk-Browne, N.B.: A taxonomy of external and internal attention. Annu. Rev. Psychol. **62**, 73–101 (2011)
5. Williams, E.: Visual Search: A novel psychophysics for preattentive vision. http://web.mit.edu/rsi/www/pdfs/papers/99/ekwillia.pdf (1999)

6. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. Cogn. Psychol. **12**(1), 97–136 (1980)
7. Wolfe, J.M.: Guided search 2.0 a revised model of visual search. Psychon. Bull. Rev. **1**(2), 202–238 (1994)
8. Theeuwes, J.: Endogenous and exogenous control of visual selection. Perception **23**(4), 429–440 (1994)
9. Theeuwes, J.: Top-down and bottom-up control of visual selection. Acta. Psychol. **135**, 77–99 (2010)
10. Trick, L.M., Enns, J.T., Mills, J., Vavrik, J.: Paying attention behind the wheel: a framework for studying the role of attention in driving. Theor. Issues Ergon. Sci. **5**(5), 385–424 (2004)
11. Chun, M.M., Golomb, J.D., Turk-Browne, N.B.: A taxonomy of external and internal attention. Annu. Rev. Psychol. **62**, 73–101 (2011)
12. Neskovic, P., Cooper, L.N.: Visual search for object features. In: Wang, L., Chen, K., S. Ong, Y. (eds.) ICNC 2005. LNCS, vol. 3610, pp. 877–887. Springer, Heidelberg (2005)
13. Wolfe, J.M., Horowitz, T.S.: What attributes guide the deployment of visual attention and how do they do it? Nat. Rev. Neurosci. **5**(6), 495–501 (2004)
14. Goransson, B., Gulliksen, J., Boivie, I.: The usability design process - integrating user-centered systems design in the software development process. Softw. Process Improv. Pract. **8**, 111–131 (2003)
15. Radach, R., Hyona, J., Deubel, H. (eds.): The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research. Elsevier, Amsterdam (2003)
16. Macmillan, N.A., Creelman, C.D.: Detection Theory: A User's Guide. Lawrence Erlbaum Associates, Mahwah (2005)
17. Fuchs, S., Hale, K.S., Stanney, K.M., Juhnke, J., Schmorrow, D.D.: Enhancing mitigation in augmented cognition. J. Cogn. Eng. Decis. Making **1**(3), 309–326 (2007)
18. Milham, L.M., Carroll, M.B., Stanney, K.M., Becker, W.: Training requirements analysis. In: Schmorrow, D., Cohn, J., Nicholson, D. (eds.) The Handbook of Virtual Environment Training: Understanding, Predicting and Implementing Effective Training Solutions for Accelerated and Experiential Learning. Ashgate Publishing, Aldershot (2008)
19. Sellers, B., Rivera, J., Fiore, S.M., Schuster, D., Jentsch, F.: Assessing X-ray Security Screening Detection following Training with and without Threat-item Overlap. Proc. Hum. Factors Ergon. Soc. Annu. Meet. **54**(19), 1645–1649 (2010)
20. Hale, K.S., Carpenter, A., Johnston, M., Costello, J., Flint, J.D., Fiore, S.M.: Adaptive Training for Visual Search. In: Proceedings of the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC 2012). Orlando. Paper # 12144 (2012)
21. Winslow, B., Carpenter, A., Flint, J.D., Wang, X., Tomasetti, D., Johnston, M., Hale, K.: Combining EEG and eye tracking: using fixation-locked potentials in visual search. J. Eye Mov. Res. **1**(1), 1–12 (2013)