Visual Search in Operational Environments

Balancing Operational Constraints with Experimental Control

Ann Speed[(✉)]

Sandia National Laboratories, Albuquerque, NM 87185, USA
`aespeed@sandia.gov`

Abstract. Visual search has been an active area of research – empirically and theoretically – for a number of decades, however much of that work is based on novice searchers performing basic tasks in a laboratory. This paper summarizes some of the issues associated with quantifying expert, domain-specific visual search behavior in operationally realistic environments.

Keywords: Visual search · Expertise · Domain specific tasks · Ecological validity · Operational environments

1 Introduction

Visual search has been a very active area of research for a number of decades, and many theories of different aspects of visual search and visual cognition exist to explain any number of behavioral and neural phenomena. This particular paper, and associated conference session, is not intended to re-tread this ground. Rather, the intent is to highlight a program of research at Sandia National Laboratories in domain specific visual search by experts and novices in a variety of high-consequence, real-world, national security problems. This program of research is relatively new at the Laboratory, with human subjects research going back to ∼2009. However visual search spans a large number of problems within the mission space of the Laboratories, thus the area has rapidly grown to comprise 15–20 researchers who explore the human cognition aspect of the problem (as opposed to focusing on algorithm or visualization development) using both qualitative and quantitative empirical methods.

This paper discusses what we believe are important task differences between domain-specific search in national security environments and the domain–general tasks typically used to develop the extant theoretical literature. Then, summaries of several key themes in each of the papers appearing in this session are presented.

1.1   Comparing Visual Search Tasks in the Laboratory and the Field

Visual search in the typical laboratory setting involves stimuli that the vast majority of subjects have experience seeing, such as letters or natural scenes. Of course, the actual construction of the stimuli depends on the specific question being asked (e.g., Figures 1 and 2).
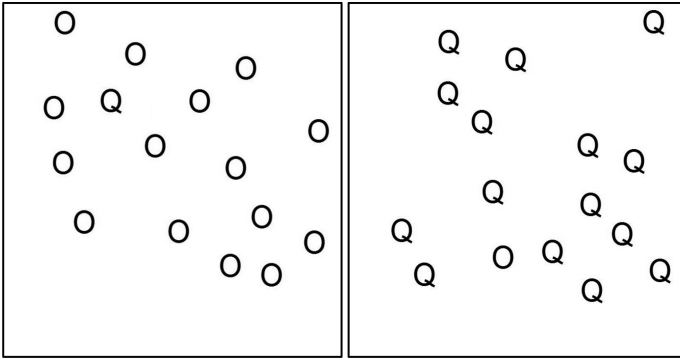

Fig. 1. Stimuli to investigate parallel (left side) versus serial (right side) visual search. Because of the differences in ratios of Os to Qs in each of these stimuli, they appear to elicit different search strategies, revealing something about how the visual system processes information.
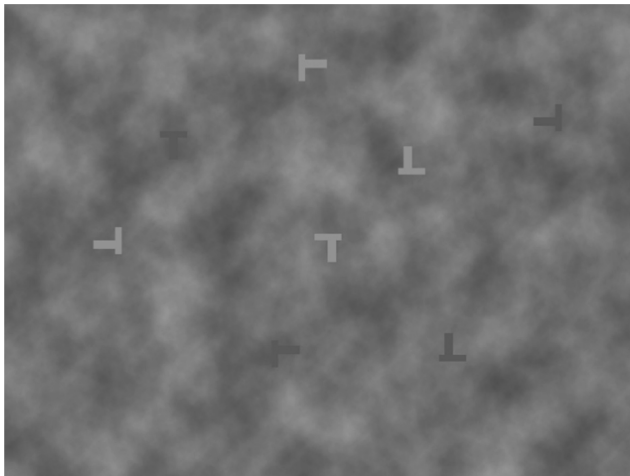

Fig. 2. A sparsely populated visual search task involving the search for perfect Ts amidst a field of Ls.

By way of comparison, stimuli in real world visual search tasks include stimuli such as X-ray or MRI imagery in medical or quality inspection environments. While the consequences of incorrect interpretations of these images are certainly high, there are a couple of critical difference between these real-world tasks and the national security domains in which this group of researchers has been working. First is the fact that visual search in radiography (or fuselage inspection, or quality control of manufactured items) is constrained by the anatomy of the patient, airplane, or widget. Thus, searchers in these domains have a frame of reference for what is and is not "normal" – even in spite of individual variability. Second, and possibly more important, is the fact that searchers in these contexts are unlikely to face a situation in which the target of the search (e.g., cancer) is intentionally being concealed by some human adversary who is being driven by their own goals with equally significant consequences.

Thus, while there are numerous industries in which humans play a critical role in quality and safety control through visual search and inspection, visual search in the national security arena has been studied less frequently than these other real-world problems – most likely because of the sensitivity of the domains and the overall lack of access to domain expert visual searchers. In a number of the problems under the national security umbrella, searchers are not simply looking at raw images (x-ray or otherwise), they are actually looking at products of images. That is, the data from the sensors is subjected to post-processing that is intended to highlight aspects of the image that might be particularly useful to the image analyst. One might think of the results of this post-processing to be automatic methods for creating a cued visual search environment [1]. For example, Fig. 3 displays an X-ray image of a carryon bag with a gun in it. While the image is a veridical representation of the contents of the bag, the dual manipulation of the image being an X-ray (versus a visual light photograph) and falsely colored potentially has implications for how Transportation Security Officers (TSOs) search for target items in these images.

Figure 4 presents synthetic aperture radar (SAR) images of two locations on Kirtland Air Force Base in Albuquerque, New Mexico. As with the TSA X-ray image, these are not visible light photographs, thus the dark and light spots carry different information than they would if they were normal black and white photographs. Additional post-processing is often done on SAR images to further highlight potentially interesting information (see Matzen et al. [2] for additional information). However, the nature of these images and the fact that manipulations are performed specifically to try and enhance imagery analyst search of them is likely to have implications for how we understand human visual search and the neural machinery enabling it.

1.2   Task Differences and How They Might Impact Visual Search Behavior

In addition to stimulus differences between domain general and domain specific tasks, operationally-oriented research also has procedures that are constrained by the operational environment in which they occur. For example, one effect that creates a stir in operational environments is Wolfe's prevalence effect [3, 4] in which subjects are more likely to miss targets when they occur infrequently. Interestingly, this effect seems to be
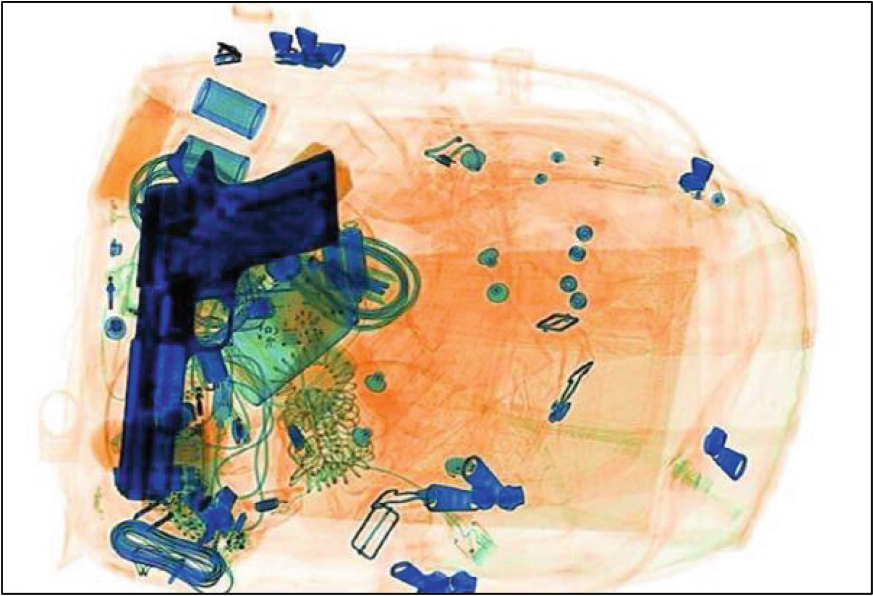
partially mediated by the trial-by-trial feedback provided to subjects in the lab [5, 6] – a luxury rarely afforded to real-world analysts in their everyday jobs. Thus, research attempting to understand performance in that everyday world will often include procedures that mimic standard operating procedures (SOPs) that are in use on the job, rather than using procedures directly out of the peer-reviewed literature. Of course, the comparison between performance in a domain-specific task under SOPs versus lab-based procedures can help to highlight differences in search behavior that are not due solely to the stimuli or to the neural machinery of the visual system.

1.3   Summary

Because of stimulus and procedural differences, the generalizability of the peer-reviewed theoretical and empirical research to operational environments is unknown. Certainly work in real-world visual search and inspection tasks, such as radiology inspection, aircraft fuselage inspections, inspections of machined parts exists [e.g., 7–14], but it is unclear that even these results generalize well to high-consequence national security domains in which there is an adversary attempting to hide target items and for which there has been a large amount of work done on post-processing raw images in an attempt to help the analyst better search the space. Thus, Sandia, along with several other government agencies tasked with national security missions, has embarked on a program of human subjects studies and experiments to better understand where high-consequence expert visual search departs from what is already known and described in the literature.

To that end, the remainder of this paper summarizes some of the recent work done at Sandia on expert, domain-specific visual search. Due to the sensitivity of much of this research, the primary focus is on methods for collecting both qualitative and quantitative data. However, where possible, results are presented.

2   Summary of Session Papers

The papers in this session describe a number of different methods – both qualitative and quantitative – aimed at better understanding the nature and complexity of visual search in an adversarial environment. To set a larger context in which these methods have been developed, each of the projects involved has collected (or is currently collecting) data on the relevant experts' domain-specific visual search task (e.g., SAR analysis, X-ray analysis) and each has collected data on a common battery of domain-general visual cognition tasks (described in more detail in Matzen et al. [2]) including:

- Parallel versus serial visual search (the O/Q task in Fig. 1)
- A visual inspection task (the T/L task in Fig. 2)
- Spatial working memory, mental rotation, attention beam, and Raven's-like matrix reasoning problems

As of the date of the writing of this paper, an insufficient amount of data on these tasks had been collected to be presented (with the exception of Matzen et al. [2] and Trumbo et al. [15]). However, we anticipate future publications covering these results.

Methods used in the following papers include qualitative approaches from cultural anthropology to perform workflow and cognitive task analysis and more quantitative methods of eliciting knowledge from these experts. Additional work describes domain experts performing laboratory-based tasks (e.g., a rapid serial visual presentation (RSVP) paradigm) using real-world stimuli and methods for collecting data on experts performing their domain-specific visual search task in a near-real operational environment. Finally, some exploratory data analysis methods are presented for dealing with data that has high temporal and spatial fidelity, which characterizes the data that many of these projects will generate.

2.1 Understanding the Nature of Visual Search Work: Knowledge Elicitation and Workflow Analysis

One of the issues with experimentally studying domain experts performing their domain task in a national security environment is that disruption of the analysts' workflow to instrument their workstation in a way that allows for quantifying their behavior can be very disruptive to the mission. Furthermore, there can be a very strong push against any modifications of their systems when the stakes are what they tend to be in these situations and because such modifications can be quite costly. Thus, insight into how these analysts perform their jobs is often limited to observation, interviews of various sorts, and examination of work documents like standard operating procedures.

McNamara and colleagues [16] describe a nice combination of a number of methods from cultural anthropology and psychology for exploring the way analysts conduct their work including ethnographic approaches, work analysis and hierarchical cognitive task analysis. They demonstrate that this "hybrid" approach yields understanding of analysts' methods that would not have been identified otherwise.

Haass and colleagues [17] take this approach a step further, incorporating eye tracking into their study of analysts performing abductive reasoning on data analogous to spectroscopic waveforms. Fortunately, despite their small sample of analysts, Haass et al., were able to collect data from highly experience ($\sim$15 years), "practitioners" who had about 5.5 years of experience and novices who had no experience with the task, but who were technically qualified and cleared to perform the task. As with McNamara, et al., Haass and colleagues demonstrated the ability to detect differences in analyst behaviors and their narrative about how they were making decisions.

2.2 Stimuli – Creation and Validation

ough they can be collected. In the case of the TSA, ground truth about the bag contents is not known. Similarly, for the SAR tasks, ground truth is often not known – specifically for target events that are not detected by analysts. In addition, if a stimulus-specific independent variable is of interest (e.g., threat prevalence rates) using real stimuli often prevents this because of the lack of control of other variables that could function as confounds. Thus, stimuli need to be created that mimic the operational environment as closely as possible. Several of the papers in this session include methods for creating realistic stimuli, but the most detailed description of this process is in Speed et al. [18].

For that project, the goal was to have Transportation Security Officers (TSOs) interrogate X-ray images of mock passenger bags for two hours in order to determine if there are significant decrements in threat detection performance over that timescale. Because the task was self-paced, based on prior research with TSOs, it was estimated that in order to ensure every TSO performed the task for two hours, there would need to be 1,000 different passenger items for each TSO to interrogate. In order to replicate the image manipulation capabilities TSOs have access to at the checkpoint X-ray, more than 83,000 unique images were loaded into a custom-built software X-ray emulator. Thus, validating that there were, indeed, unique images for each requested image

product, and that there were the right number of image products for each passenger item, became a very important task.

2.3   Expert Performance on Basic Visual Search Tasks

As mentioned previously, many of the projects presented include collection of data from expert visual searchers on a common battery of basic visual search tasks. Trumbo et al. [15] and Matzen et al. [2] describe in more detail the performance of domain experts on an RSVP task that uses chips of X-ray images and on that general visual search battery, respectively.

Trumbo et al. [15] describe a variation of an approach initially developed by DARPA for the Neurotechnology for Intelligence Analysts (NIA) program. Specifically, that program utilized event-related potentials (ERPs) in electroencephalography (EEG) to enable satellite imagery analysts to triage large numbers of images. Specifically, researchers presented "chips" of large satellite images to analysts using a rapid serial visual presentation (RSVP) paradigm and utilized the presence of specific ERPs to determine if those image chips, and their associated whole images, needed to be looked at more closely. Trumbo et al., apply this method to TSOs, using chips of actual false-color X-ray images. The researchers demonstrated that TSOs were able to identify threats in the chips despite the speed at which they were presented (100 ms per image chip) and that there was a positive waveform appearing approximately 300 ms after chip onset for stereotypically presented threat items, thus demonstrating the applicability of the EEG-based triage approach to X-ray baggage screening.

Matzen et al. [2] present data for expert and novice visual searchers on both their domain-specific task (SAR imagery) and on the aforementioned domain-general cognitive battery. They find important differences between experts and novices on both tasks, thus demonstrating that expertise in visual search does correlate with changes in performance in other visual cognition domains. This finding replicates other similar research (e.g., Biggs et al. [19]).

Silva et al. [20] describe measuring visual search in a slightly different domain: cyber incident responders (IRs). Interestingly, IRs are sometimes faced with a difficult visual search task in searching log files for malware. This task is driven much less by the characteristics of the stimulus and more by the individual IR's knowledge of malware code structure.

2.4   Data Analysis

Several of the projects described in this session involve collection not only of behavioral data (e.g., classifying a stimulus as either "normal" or "abnormal"), they also include eye tracking data and, in some cases, very temporally detailed information about how the analysts interact with the stimulus presentation system to make their decisions. Thus, analyses of the resulting data necessarily go beyond traditional parametric statistical tests and branch into machine learning and, in some cases, text analysis-based methods. Stracuzzi et al. [21] describe a framework for analyzing data

that have both temporal and spatial aspects to them. This framework will likely be applied to many of the datasets being generated by the projects described in this session.

3   Conclusion

As Matzen et al. [2] point out, there is a lack of research on domain expertise in visual search and how that expertise both impacts performance on domain-general tasks and how performance on both domain-specific and domain-general tasks differs between novices and experts. While much of the data being collected for the projects described in this session is yet to be analyzed as of the writing of this paper, hopefully many of the methodological issues surrounding collecting such data can inform others' efforts such that this gap in the literature can be more quickly closed. Understanding the nature of expertise in various real-world, high-consequence visual search domains, how that differs from the behavior of domain novices, and how that might impact (or be predicted by) performance on domain-general tasks potentially has significant implications for theories of visual search, including understanding the neural machinery underlying visual search behavior.

For additional examples of this kind of research, the reader is referred to another session in these same Proceedings entitled "Applying Science to Complex Operational Environments: Methodological Case Studies from the Transportation Security Administration's Human Factors Group."

References

1. Eckstein, M.P.: Visual search: a retrospective. J. Vis. **11**, 1–36 (2011)
2. Matzen, L.E., Haass, M.J., McNamara, L.A., Stevens-Adams, S.M., McMichael, S.N.: Effects of professional visual search experience on domain-general and domain-specific visual cognition. In: 17th International Conference on Human-Computer Interaction. Springer (2015)
3. Wolfe, J.M., Horowitz, T.S., Van Wert, M.J., Kenner, N.M.: Low target prevalence is a stubborn source of errors in visual search tasks. J. Exp. Psychol. Gen. **136**, 623–638 (2007)
4. Wolfe, J.M., Van Wert, M.J.: Varying target prevalence reveals two dissociable decision criteria in visual search. Curr. Biol. **20**, 121–124 (2010)
5. Fiedler, K.: Beware of samples! a cognitive-ecological sampling approach to judgment biases. Psychol. Rev. **107**, 659–676 (2000)
6. Fiedler, K., Brinkmann, B., Tilmann, B., Wild, B.: A sampling approach to biases in conditional probability judgments: beyond base rate neglect and statistical format. J. Exp. Psychol. Gen. **129**, 399–418 (2000)

7. Drury, C.G., Saran, M., Schultz, J.: Effect of fatigue/vigilance/environment on inspectors performing floursecent penetrant and/or magnetic particle inspection: Year 1 interim report (2004)
8. Drury, C.G., Watson, J.: Good practices in visual inspection
9. Gallwey, T.J., Drury, C.G.: Task complexity in visual inspection. Hum. Factors **28**, 595–606 (1986)
10. Lansdale, M., Underwood, G., Davies, C.: Something overlooked? How experts in change detection use visual saliency. Appl. Cogn. Psychol. **24**, 213–225 (2010)
11. Panjwani, G., Drury, C.G.: Effective interventions in rare event inspection. In: Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting (2003)
12. See, J.E.: Visual inspection: A review of the literature. Sandia Technical Report, SAND2012-8590 (2012)
13. Spencer, F., Schurman, D.: Reliabiity assessment at airline inspection facilities, vol. III. Results of an eddy current inspection reliability experiment. Technical report for the Federal Aviation Administration, DOT/FAA/CT-92-12, III (1995)
14. Spitz, G., Drury, C.G.: Inspection of sheet materials: test of model predictions. Hum. Factors J. Hum. Factors Ergon. Soc. **20**, 521–528 (1978)
15. Trumbo, M.C., Matzen, L.E., Silva, A., Haass, M.J., Divis, K., Speed, A.: Through a scanner quickly: elicitation of event-related potentials in transportation security officers following rapid image presentation and categorization. In: 17th International Conference on Human-Computer Interaction. Springer (2015)
16. McNamara, L.A., Cole, K., Haass, M.J., Matzen, L.E., Morrow, J.D., Stevens-Adams, S.M., McMichael, S.: Ethnographic method for experimental design: case studies in visual search. In: 17th International Conference on Human-Computer Interaction. Springer (2015)
17. Haass, M.J., Matzen, L.E., Stevens-Adams, S.M., Roach, A.R.: Methodology for knowledge elicitation in visual abductive reasoning tasks. In: 17th International Conference on Human-Computer Interaction. Springer (2015)
18. Speed, A., Silva, A., Trumbo, D., Stracuzzi, D.J., Warrender, C., Trumbo, M.C., Divis, K. Determining the optimal time on X-ray analysis for transportation security officers. In: 17th International Conference on Human-Computer Interaction. Springer (2015)
19. Biggs, A.T., Cain, M.S., Clark, K., Darling, E.F., Mitroff, S.F.: Assessing visual search performance differences between transportation security administration officers and non-professional visual searchers. Vis. Cogn. **21**, 330–352 (2013)
20. Silva, A.R., Emmanuel, G., McClain, J.T., Forsythe, J.C., Matzen, L.E.: Measuring expert and novice performance within computer security incident response teams. In: 17th International Conference on Human-Computer Interaction. Springer (2015)
21. Stracuzzi, D.J., Speed, A., Silva, A., Haass, M.J., Trumbo, D.: Exploratory analysis of visual search data. In: 17th International Conference on Human-Computer Interaction. Springer (2015)