SANDIA REPORT
SAND201X-XXXX
Unlimited Release
Printed Month and Year

Eye Tracking for Dynamic, User-Driven Workflows

Laura A. McNamara, Max Chen, Kristin Divis, J. Daniel Morrow
Sandia National Laboratories
Albuquerque, NM 87185

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology and Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

SAND201X-XXXX
Printed November 2018
Unlimited Release

SAND Report Title

Laura A McNamara, *PI*
Kristin Divis, and J. Daniel Morrow, *Technical Team Members*
ORG-05346, ISR Advanced Exploitation and Human-Systems Integration

Maximilian Chen, *Technical Team Member*
ORG 058xx
Sandia National Laboratories
P. O. Box 5800
Albuquerque, New Mexico 87185-MS-0519

Abstract

This three-year Laboratory Directed Research and Development (LDRD) project aimed at developing a developed prototype data collection system and analysis techniques to enable the measurement and analysis of user-driven dynamic workflows. Over 3 years, our team developed software, algorithms, and analysis technique to explore the feasibility of capturing and automatically associating eye tracking data with geospatial content, in a user-directed, dynamic visual search task. Although this was a small LDRD, we demonstrated the feasibility of automatically capturing, associating, and expressing gaze events in terms of geospatial image coordinates, even as the human "analyst" is given complete freedom to manipulate the stimulus image during a visual search task. This report describes the problem under examination, our approach, the techniques and software we developed, key achievements, ideas that did not work as we had hoped, and unsolved problems we hope to tackle in future projects.

TABLE OF CONTENTS

[On odd-numbered page]

FIGURES

TABLES

EXECUTIVE SUMMARY

Human-information interaction researchers rely on user interaction logs to study cue-driven behaviors associated with information foraging workflows in electronic environments – for example, when you are using your desktop computer to search for references in a cloud-hosted library. Mouse events (mouse movement, mouse pauses, mouse clicks) are used as proxy indicators of user attention and decision-making during the navigation of an interface, database portal, or web site. Eye tracking systems that generate gaze metrics, such as fixations, dwells, and scan paths, could provide valuable complementary evidence for understanding the perceptual and cognitive processes of information foraging in online environments. However, eye tracking technologies evolved in a research paradigm that assumes highly constrained, laboratory-type tasks that use static stimuli to test carefully structured hypotheses about perception and attention. Having evolved in a highly deductive scientific paradigm, eye trackers are less suited for inductive data collection and analysis. Eye tracking study design, data collection and analysis software tools are simply not suited for studying the complicated, unpredictable, user-driven workflows that characterize information foraging "in the wild."

This three-year Laboratory Directed Research and Development (LDRD) project aimed at developing a developed prototype data collection system and analysis techniques to enable the measurement and analysis of user-driven dynamic workflows. The project was inspired by our experience on the PANTHER Grand Challenge (FY2013-FY2015), which enabled us to capture, measure, and analyze streams of human-software-imagery interactions as trained Synthetic Aperture Radar (SAR) analysts searched SAR images for signatures associated with improvised explosive device (IED) emplacements. Missing from the otherwise detailed workflow analysis we developed in PANTHER were any measures derived from our eye tracker. This is because eye trackers are simply not designed for easy association of data samples gaze events (e.g., fixations) with dynamically-changing content, such as a stack of SAR images, in a desktop system.

Over three years, our team received $600K ($150K in FY16/FY17; $300K in FY18) to assess the feasibility of capturing and automatically associating eye tracking data with geospatial content, in a user-directed, dynamic visual search task. In FY16, we developed stimulus display and eye tracking data collection software that would allow us to capture gaze position, while simultaneously tracking changes in the relative position of a SAR image being moved (panned) in a display space. We then integrated our 60Hz Fovio eyetracker (distributed by Eye Tracking, Inc.) and recruited human participants to help us generate a controlled but feature-rich sandbox dataset for exploratory analysis and algorithm development. We spent FY17 and FY18 developing "bottom up" or inductive analysis methods to describe, analyze, and compare search styles and strategies among our group of image "analysts."

Although this was a small LDRD, our team was able to demonstrate the feasibility of automatically capturing, associating, and expressing gaze events in terms of geospatial image coordinates. Our prototype data collection and analysis pipeline enables us to express gaze measures in terms of the content the user is engaging, even as the human "analyst" is given complete freedom to manipulate the stimulus image while searching for the targets. We can also integrate calculated gaze events (fixations, dwells, scanpaths) with image manipulation events

(panning direction and speed) into a series of “discrete” user-image interaction events, or tokens. Tokenizing these interaction events is an important step for integrated analysis

1. INTRODUCTION

Human-information interaction researchers rely on user interaction logs to study cue-driven behaviors associated with information foraging workflows in electronic environments – for example, when you are using your desktop computer to search for references in a cloud-hosted library. Mouse events (mouse movement, mouse pauses, mouse clicks) are used as proxy indicators of user attention and decision-making during the navigation of an interface, database portal, or web site. Eye tracking systems that generate gaze metrics, such as fixations, dwells, and scan paths, could provide valuable complementary evidence for understanding the perceptual and cognitive processes of information foraging in online environments. However, eye tracking technologies evolved in a research paradigm that assumes highly constrained, laboratory-type tasks that use static stimuli to test carefully structured hypotheses about perception and attention. Having evolved in a highly deductive scientific paradigm, eye trackers are less suited for inductive data collection and analysis. Eye tracking study design, data collection and analysis software tools are simply not suited for studying the complicated, unpredictable, user-driven workflows that characterize information foraging "in the wild."

This three-year Laboratory Directed Research and Development (LDRD) project aimed at developing a developed prototype data collection system and analysis techniques to enable the measurement and analysis of user-driven dynamic workflows. The project was inspired by our experience on the PANTHER Grand Challenge (FY2013-FY2015), which enabled us to capture, measure, and analyze streams of human-software-imagery interactions as trained Synthetic Aperture Radar (SAR) analysts searched SAR images for signatures associated with improvised explosive device (IED) emplacements. Missing from our otherwise detailed workflow analysis were any measures derived from our eye tracker, which was not capable of associating gaze events (e.g., fixations) with dynamically-changing content in the desktop displays we were using.

Over three years, our team received $600K ($150K in FY16/FY17; $300K in FY18) to assess the feasibility of capturing and automatically associating eye tracking data with geospatial content, in a user-directed, dynamic visual search task. This report explains why user-driven, dynamic workflows are so difficult to study; and outlines our team's approach, the solution we developed, analytic techniques, and lessons learned. Further details are available in the conference papers and presentations that were funded under this LDRD (listed in the appendix).

In the following pages, we describe the class of problems that interest us and why they are important. We then explain why eye tracking data collection systems and analysis techniques are poorly suited for the visual cognitive workflows that our team is examining. Over the past three years, we developed a prototype system for presenting geospatial imagery in a way that allowed us to track the elements of image content being gazed at, while also allowing the user to freely pan the image in unconstrained visual search task. Enabling the automatic association of gaze events with elements of content opened the door to "bottom-up" analysis of study participants' visual search strategies.

2. MOTIVATING PROBLEM

Like many LDRD projects, ours was borne out of frustration with existing technology. In this section, we explain how our work with imagery analysts in the Sandia Copperhead program led to a richer appreciation for the rapidly changing nature of imagery analysis in the United States Intelligence Community. As we discuss below, the advent of so-called *softcopy imagery analysis*, represented by the rapid adoption of digitized information systems and electronic displays in the intelligence community, has led to explosive growth in the diversity and quantity of image products that analysts have at their disposal. Coupled with complementary data sources – such as user log files, observational research, and focused experiments – eye tracking data could provide a powerful tool for studying how imagery analysts learn to allocate attentional resources in high-throughput workflows.

2.1. SAR Imagery Analysis

Under the PANTHER Grand Challenge (2012-2015; Kristina Czuchlewski, PI), we had the opportunity to study how trained Synthetic Aperture Radar (SAR) imagery analysts in the Copperhead program interacted with stacks of SAR image products to detect, evaluate, and make decisions about potential threat signatures associated with improvised explosive devices (IEDs) [1]. As we discuss below, imagery analysis these days relies on specialized software called *electronic light tables* (ELT) that enable imagery analysts to retrieve, interact with, annotate, and produce intelligence with digital image products. The Copperhead program had developed a customized, program-specific ELT package. The fact that this ELT was developed locally enabled us to customize a version of the platform with a logging system that generated event messages whenever a user manipulated a SAR image in a workflow.

Coupled with interviews and observational research we conducted with the imagery analysts in the program, this custom Copperhead ELT enabled PANTHER researchers to do relatively naturalistic research on the analysis process. We used interviews and ethnographic observation to generate a basic descriptive model of the imagery analysis workflow, then developed a realistic task analysis protocol involving anomaly detection and differentiation challenges. As they performed the analysis task using our instrumented system, we captured analysts' panning and zooming behaviors; points when they transitioned to analysis of historical imagery to compare current scene features with previous features; and the keyboard-based "flickering" behaviors, which generated an illusion of movement that facilitates detection of scene differences. Analysis of the ELT log files revealed differences in analysts' search strategies: for example, we found that experienced analysts replied more heavily on a specific subset of image products than their novice counterparts, who looked at a wider range of image types [2].

To better understand differences between experienced and less-experienced Copperhead personnel, PANTHER researchers developed eye tracking experiments examining how domain experience influenced tacit search strategies and target detection performance. The eye tracking research revealed that experienced imagery analysts unconsciously directed their gaze to image areas where threat signatures were most likely to be visible – a good example of how experiential knowledge, in the form of "top-down attention," facilitates task efficiency in a demanding, high-throughput workflow. Not surprisingly, the less experienced analysts were both less efficient in their search strategies, spending more time examining less valuable regions of an image; *and* were less accurate than their experienced counterparts [2-4].

Given the resources that PANTHER and the Copperhead program afforded us, one might reasonably expect that we would have used eye trackers to capture imagery analysts gaze patterns, and to integrate the gaze datasets with the ELT log files. Capturing the details of imagery analysts' visual interactions with SAR images, in a relatively unstructured, naturalistic workflow, would have been a powerful complement to the richly detailed log files we were collecting. Unfortunately, today's eye trackers do not lend themselves to naturalistic research designs, for reasons we discuss below [5, 6]. Moreover, naturalistic research requires analysis methods that enable us to characterize and compare complex information foraging behaviors in high-dimensional image spaces. Doing so would enable us to richly characterize the complex visual cognitive workflows of imagery analysis in the era of electronic light tables and softcopy sensor products [7, 8]. We discuss each of these topics in turn below.

3. STATE OF THE (EYE TRACKING) ART

Researchers have been measuring eye movements to study visual behavior, perception and cognition for well over a century. Eye trackers developed in the late 19th/early 20th century were physically invasive, requiring research participants to wear contact lenses to which lightweight styluses were mounted; or with embedded metal coils that measured electromagnetic fluctuations associated with eye movements. Fortunately, eye tracking technology has become much less invasive: these days, eye tracking systems consist of image processing algorithms/software and a small infrared video camera/illumination source. The infrared source illuminates the eyeball, generating two signatures that can be used to calculate eye position: a small, sharp reflection from the surface of the cornea; and a larger, disk-shaped bright reflection from the pupil. Basic commercial systems sample these reflections at 60-120Hz (intervals between 16.67 and 8.83 milliseconds), while high-end laboratory systems provide sampling rates up to 2000Hz. Eye trackers may be "head mounted" glasses worn by research participants, or remote cameras mounted alongside the stimulus/display surface to illuminate a participant's face from a short distance (usually within a few feet) [9, 10].

Eye tracking data collection systems[*] have evolved tremendously over the past few decades, becoming smaller, lighter, and less expensive, while providing greater measurement accuracy and precision. As eye trackers have become less intimidating, they have migrated from the experimental confines of visual laboratories into a widening range of application domains, most notably in human-computer interaction research. Even as eye trackers have moved "into the wild," however, neither the systems nor their associated analysis methods have evolved much beyond the conceptual model of experimental psychology [10].

There are a few reasons for this. For one thing, the diffusion of eye trackers outside the psychology laboratory was largely driven by human-computer interaction researchers in the late 1990s and early 2000s. Gaze metrics developed by visual psychologists were easily translated to human-computer interaction research, with minimal interpretive tweaking. An excellent example of this is the widely cited (939 citations and counting!) Poole and Ball review of 2006. They provide a glossary of standard eye tracking terms and measures (e.g., fixations, saccades,

---

[*] We are not speaking here of eye tracking systems used as *input* devices (e.g., assistive technologies); only as data collection systems in basic and applied research.

regressions, scanpath, transitions, etc.) and provide guidance for applying those measures in software evaluation: for example, regressive saccades may indicate that users are having trouble interpreting an interface element [10].

There is nothing intrinsically wrong with repurposing tried-and-true analysis techniques for applied research in human-computer interaction. However, these measures emerged in the hypothetic-deductive world of the scientific laboratory, which presumes a controlled set of tasks, stimuli, and a structured experimental process to systematically develop and test hypotheses about visual perception and cognition. Controlled experimental studies are (by definition) not naturalistic studies, so it is hardly surprising that a technology that emerged from the former is not well suited to the latter.

Even as the technology has evolved into the mainstream, most eye trackers are fairly simple sensors that sample the location of individual's gaze in a 2-dimensional field at regular intervals. An eye tracking dataset is nothing more than a series of time-stamped spatial coordinates (x, y, t). If this sounds basic, that's because it is: eye trackers only log the *where* and *when* of an individual's gaze. Determining *what* an individual may be perceiving or attending to is a separate problem that involves the association of those (x, y, t) samples captured in the "display space" with the content that the display is actually rendering. Commercial eye tracking systems usually come with study design and stimulus presentation software that "tracks" both the gaze events and the content being rendered on the display. However, these systems typically presuppose content that is scripted by the experimenter, rather than being driven by the whims of the study participant or user.

A third challenge is the lack of support for integrating eye tracking systems with mainstream computer hardware and software. These days, just about anyone with a few thousand dollars and a desktop computer can purchase, install, and run an eye tracker. Most systems offer support for stimulus presentation, data collection, data management and analysis for standard eye tracking study designs – i.e., involving static stimuli and/or constrained tasks. Naturalistic research, however, is not well-supported. Most eye tracking packages do not interact with third-party applications (such as a browser), to automatically associate of an (x, y, t) gaze sample with the actual *content* rendered on the display. Instead, most eye tracking packages include study design and/or stimulus presentation unit. This application enables the researcher to use the eye tracking system to present a visual stimulus (an image, a film clip) to the study participant, so that the eye tracking software can keep track of what content is being rendered, when and where, in the display space, then associate the content with the captured gaze events. There are, of course, fully integrated software/hardware/sensor suites designed for specialized eye tracking data collection, but these are extremely expensive and are intended for basic research in visual perception and cognition – for example, collecting data to characterize how humans interact with letter shapes when reading text.

Interestingly, despite the fact that eye trackers are really not suited for more freeform, dynamic, realistic workflows of people interacting with digital information spaces, HCI researchers have not developed new methods for analyzing eye tracking data.

Inadequacies

Eye trackers generate an enormous amount of data, even at the low end of 60Hz. Automated analysis and processing is critical. Deductve experimental designs, hypothesis testing that replies on known targets and etc, are well supported by eye movement study designs and aanlysis techniques, softare. However, if you don't know where people are going to look, or how they're going to position the stimulus – if you put users in control -

4. IMAGERY ANALYSIS: FROM VISUAL INSPECTION TO VISUAL INFORMATION FORAGING

One of the most useful constructs in human factors and industrial engineering is the visual inspection model, which describes how people examine an

Over the past twenty years, the discipline of imagery analysis has experienced a quiet but rapid revolution in tools and techniques. Until fairly recently, imagery analysts in the United States Intelligence Community were trained in using physical films and illuminated surfaces, or *light tables*, to examine imagery for indicators of significant intelligence events. In the 2000s, however, the rapid expansion of affordable desktop computing, networking, and enterprise information management, collectively overthrew the hardcopy paradigm. By 2010, the light tables and film cabinets had been replaced by electronic displays (both LCD and CRT), digitized imagery databases, and electronic light tables, including commercial platforms such as SOCET GXP and RemoteView.

Electronic light tables are software systems that digitize and render images on a desktop or other display panel. Throughout the intelligence workplace, ELTs have largely replaced the physical light tables and magnifying lenses that were once the workhorse artifacts of imagery analysts. The shift from hardcopy inspection to softcopy interaction has reshaped the imagery analysis workflow: where imagery analysts once examined individual or small sets of physical films, ELT software and electronic databases enable analysts to interact rapidly and fluidly with a greater range of image products. In addition, digitized geospatial data affords the application of new image processing and rendering algorithms to derive spatial indicators from images taken at different points in time ("change detection").

5. TOWARD GAZE INFORMED INFORMATION FORAGING MODELS

The problems our work sought to address were as follows:

Issue: pixel rendered imagery is high dimensional. Eye tracking data even at the relatively low rate of 60Hz is unwieldy and must be processed into useful measures of gaze behaviors, such as fixations and clusters of fixations (which we call “dwells”). Also movements between fixations, saccades.

Transforming a raw stream of eye tracking data into meaningful measures reduces the dimensionality of the dataset considerably, but the problem of associating calculated gaze events, such as a fixation, with image content, is challenging if you have not IF you can apply top-down filters, pre-identify objects that you expect or want people to be looking at , r you want to know if people are looking at them, that’s relatively straightforward – that’s the decudcitve approach, you can keyframe the element, to facilitate tracking its lcoatoin in the display space. However, doing so inductively, without being able to predict ahead of time what they’ll be looking at, is more difficult – have to back-segment the image into regions that are morphologically, semantically meaningful, not just a raster. You can reduce dimensionality with a raster, but it doesn’t necessarily map onto meaningful objects – in fact, we know it doesn’t map onto meanigfl objects in natural scenes.

Under PANTHER borrowed the superpixel algorithms, used to cluster gaze data into regions, found that fixations mapped more nicely onto objects in the image than when the image was rastered into regions (Haass, Matzen). but pixel space remains complex.
We don’t want tp be maping gaxe samples to individual pixels, then trying to figure out what object is being “looked” at.

Borrow form PANTHER – superpixel clustering algorithms, and megapixels (a superset of superpixels – use those to reduce the dimensionality of the data we’d be dealing with.

Three spaces:

Display space

The geospatial coordinates of an image

“application space, “ a subset of the display space that facilitates quick association of the display coordinates with the geospatial coordinates of the image.

Sandbox data creationAnalyse the task elements of the copperhead workflow – panning, studying targets, “flickering” to compare signatures between images.

Treat those as abstractions, then design tasks that represent those in a highly controlled form. Four tasks, five trials, all using the same SAR image, which has good properties for a sandbox dataset

We knew we wanted to use the superpixel algorithms to bound pixel areas in ways that approximate features in the iamge, so we needed an image with clearly definable, visible regions.

Roof image – distinct areas of bright returns, amenable to clustering and object bounding with superpixel algorithms. Enable association of fixations with objects at specific locations – the roof of a house at a specific location, marked by centroid of a superpixel.

Select image

Data curation, basic analysis, and quality metrics

CHALLENGE – dimensionality reduction – millions of pixels, thousands of gaze samples

Address this by grouping gaze samples into fixations

Associating fixations with superpixel centroids

6. CONCLUSION

As our remote sensing systems become increasingly sophisticated, and as we develop algorithms that enable the extraction of more and more geospatial information from remotely sensed datasets, the diversity and quantity of information products available to intelligence analysts will only grow. It is difficult to see the intelligence community making effective use of these burgeoning resources without automation; but what automated support for geospatial workflows entails is not immediately obvious. Understanding how imagery analysts use computational systems to interact with geospatial information products can help us identify what processes and functions can be automated, so that people can do meaningful work: pattern recognition, contextual interpretation, inference, nuanced communication. What constitutes “good” automation, however, depends very much on the context of work, the tasks and goals. Even imagery analysts working in the same organization may approach seemingly similar inspection tasks very differently, depending on the mission and context for which the imagery is being collected. Environmental monitoring, for example, may require an analyst to look for subtle changes in ground elevation using imagery collected over hundreds of square miles of terrain, on a monthly basis. In contrast, an analyst looking for evidence of illicit human or drug trafficking along a contested border in the very same region might search for activity signatures generated on a much shorter timescale, perhaps over a few tens of square miles. are data-rich and can be processed into a portfolio of image products that highlight and/or minimize different types of scene features, which are variably useful depending on the analyst’s goals.

[Blank page following section.]

REFERENCES [REFERENCE HEAD]

1. Author, *Title*, SANDXX-XXXX, Sandia National Laboratories, Albuquerque, NM, March 2017. [Reference]

[Blank page following section.]

APPENDIX X: TITLE [APPENDIX HEAD]

Appendix text here.

[Blank page following section.]

DISTRIBUTION

[Can go on an even or an odd page]

[List external recipient names and full mailing addresses]

4 Lawrence Livermore National Laboratory
Attn: N. Dunipace (1)
P.O. Box 808, MS L-795
Livermore, CA 94551-0808

[List in order of lower to higher Mail Stop numbers.]

1	MSXXXX	Name of Person	Org. Number
1	MSXXXX	Name of Person	Org. Number

[The housekeeping entries are required for all SAND reports.]

1	MS0899	Technical Library	9536 (electronic copy)

For LDRD reports, add:

1	MS0359	D. Chavez, LDRD Office	1911

For CRADA reports add:

1	MS0115	OFA/NFE Agreements	10012

For Patent Caution reports, add:

1	MS0161	Legal Technology Transfer Center	11500

Sandia National Laboratories

1 Clark, H.: ‘IED detector developed by Sandia Labs being transferred to Army
’, in Editor (Ed.)^(Eds.): ‘Book IED detector developed by Sandia Labs being transferred to Army
’ (Sandia National Laboratories, 2014, edn.), pp.
2 McNamara, L.A., Cole, K., Haass, M.J., Matzen, L.E., Morrow, J.D., Stevens-Adams, S.M., and McMichael, S.: ‘Ethnographic Methods for Experimental Design: Case Studies in Visual Search’, Proceedings of the International Conference on Augmented Cognition, 2015, pp. 492-503
3 Matzen, L.E., Hasss, M., Tran, J., and McNamara, L.: ‘Using Eye Tracking Metrics and Visual Saliency Maps to Assess Image Utility’. Proc. Human Vision and Electronic Imaging, San Francisco, CA2016 pp. Pages
4 Matzen, L.E., Haass, M.J., McNamara, L.A., Stevens-Adams, S.M., and McMichael, S.N.: ‘Effects of professional visual search experience on domain-general and domain-specific visual cognition’, in Editor (Ed.)^(Eds.): ‘Book Effects of professional visual search experience on domain-general and domain-specific visual cognition’ (Springer, 2015, edn.), pp. 481-491
5 Hansen, D.W., and Pece, A.E.: ‘Eye tracking in the wild’, Computer Vision and Image Understanding, 2005, 98, (1), pp. 155-181
6 Lappi, O.: ‘Eye Tracking in the Wild: The Good, the Bad and the Ugly’, Journal of Eye Movement Research, 2015, 8, (5), pp. 1-21
7 Pirolli, P.: ‘Information Foraging ’ (Oxford University Press, 2007. 2007)
8 Pirolli, P., and Card, S.: ‘Information Foraging’, Psychological Review, 1999, 106, (4), pp. 643-675
9 Duchowski, A.T.: ‘Eye Tracking Methodology: Theory and Practice’ (Springer-Verlag, 2007. 2007)
10 Poole, A., and Ball, L.: ‘Eye Tracking in HCI and Usability Research: Current Status and Future Prospects’: ‘Encyclopedia of Human-Computer Interaction ’ (IGI Global), pp. 211-219