A New Method for Categorizing Scanpaths from Eye Tracking Data

Michael J. Haass, Laura E. Matzen, Karin M. Butler, Mika Armenta*
Sandia National Laboratories

Abstract

From the seminal work of Yarbus [1967] on the relationship of eye movements to vision, scanpath analysis has been recognized as a window into the mind. Computationally, characterizing the scanpath, the sequential and spatial dependencies between eye positions, has been demanding. We sought a method that could extract scanpath trajectory information from raw eye movement data without assumptions defining fixations and regions of interest. We adapted a set of libraries that perform multidimensional clustering on geometric features derived from large volumes of spatiotemporal data to eye movement data in an approach we call GazeAppraise. To validate the capabilities of GazeAppraise for scanpath analysis, we collected eye tracking data from 41 participants while they completed four smooth pursuit tracking tasks. Unsupervised cluster analysis on the features revealed that 162 of 164 recorded scanpaths were categorized into one of four clusters and the remaining two scanpaths were not categorized (recall/sensitivity=98.8%). All of the categorized scanpaths were grouped only with other scanpaths elicited by the same task (precision=100%). GazeAppraise offers a unique approach to the categorization of scanpaths that may be particularly useful in dynamic environments and in visual search tasks requiring systematic search strategies.

Keywords: eye tracking, pattern analysis, scanpath, trajectory analysis method, GazeAppraise

Concepts: •**Applied computing** → **Psychology;** •**Theory of computation** → *Computational geometry;*

1 Introduction

Moment-to-moment changes in mind and brain processing are reflected in how a person moves their eyes through a scene. Most commonly, eye tracking data are partitioned into discrete observations of periods of eye stability (fixations) and eye movements (saccades). These parameters have proved useful for revealing mind processes, such as the operation of spatial attention [Butler and Zacks 2006], and for relating mind and brain [Henderson et al. 2015]. Analysis of the sequential dependencies between eye positions, i.e., scanpath analysis, has been more difficult though, in part, because of the computational complexity. Visual representations of fixations and saccades spatially mapped onto the visual stimulus suggest that capturing and characterizing the combination of spatial and temporal features may provide important insights into the mind. For example, Yarbus's [1967] images of the fixations and saccades associated with answering different questions about a painting (give the ages of the people, estimate the material circumstances of the family, etc.) suggested that the goals of the viewer could be discerned from the trajectory of eye movements.

Recently several groups have used various methods to analyze the combined spatial and temporal features of eye movement behavior (represented as a vector of features) in order to distinguish task performance [Borji and Itti 2014; Haji-Abolhassani and Clark 2014; Henderson et al. 2013], however Greene et al. [2012], using similar methods, were unable to distinguish between tasks. A limitation of these methods is that they do not capture the sequential dependencies between eye movements.

Several methods have been developed to quantify scanpath similarity, but these methods require preprocessing of the visual stimuli or the eye movement data. Many of them rely on specifying areas of interest within the visual stimulus [Cristino et al. 2010]. In recurrence quantification analysis (RQA), the scanpaths of individual viewers from individual stimuli are extracted by initially dividing the stimulus into an array of spatial locations and mapping the sequence of fixation positions onto the array [Anderson et al. 2013]. Dewhurst et al. [2012] presented a method for comparing scanpaths that capture the sequential dependencies of eye positions using geometric vectors with a method called MultiMatch. However, this approach requires that the eye movement samples be processed in several ways before comparisons can be made [Jarodzka et al. 2010].

We sought a method of extracting eye movement trajectory information that could be applied to minimally-processed eye movement data, and that could be applied without specifying areas of interest a priori. The research we present here represents a proof-of-concept that this new approach can be used with unprocessed eye tracking data. Tracktable [Rintoul et al. 2015] is a set of libraries (soon to be open source) that performs multidimensional clustering on geometric features derived from large volumes of spatiotemporal data. The Tracktable libraries were originally designed for application to geospatial trajectories and have been tested using air traffic data from the US Federal Aviation Administration Aircraft Situation Display to Industry (ASDI). Tracktable is able to rapidly identify flight trajectory patterns such as holding patterns, weather avoidance, and mapping activities where the aircraft raster-scans over a land area. Like air traffic data, eye tracking data are made up of time-ordered sequences of spatial position coordinates. Recognizing the need for similar pattern identification capabilities for both domains, we have investigated the application of the Tracktable methodology to smooth pursuit eye movement data. In this paper, we report GazeAppraise, our adaptation of Tracktable for application to eye tracking data. GazeAppraise calculates geometric features over temporal intervals at multiple scales for each scanpath in an input set of eye tracking data (for example from multiple subjects viewing multiple stimuli). GazeAppraise then performs clustering in feature space to categorize scanpaths by similarity. This approach is novel because it segments eye tracking data into temporal intervals that determine the boundaries for calculating the spatial features (as opposed to defining fixations and saccades). The sequential dependencies between the eye samples are reflected in the mapping of these features onto multidimensional space.

*e-mail:mjhaass@sandia.gov
Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of the United States government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only. © 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ETRA 2016, March 14 - 17, 2016, Charleston, SC, USA
ISBN: 978-1-4503-4125-7/16/03
DOI: http://dx.doi.org/10.1145/2857491.2857503

2 Method

2.1 Participants

Forty-one employees (Males = 22, Females = 16, 3 participants did not self-identify; Age: M = 27.0, SD = 11.5, Range = 17 to 65 years) were recruited from Sandia National Laboratories via email messages distributed to members of the workforce. Participants were paid their typical wage for their time spent participating in this study.

2.2 Apparatus

Eye movements were tracked using Seeing Machines FOVIO running at 60 Hz. The FOVIO was interfaced with EyeWorks Record 3.12 software running on a DELL Precision T3600 and using the Windows 7 operating systems on an Intel Xeon CPU E5-1603 0 @ 2.80 GHz with 8 GB of RAM. Movie files of a moving dot were presented using a script created in EyeWorks Design 3.12. All stimuli were presented on a DELL 19" LCD monitor set at a resolution of 1280 $\times$ 1024.

2.3 Materials

The stimuli consisted of four movie files in .avi format in which a white dot (22 $\times$ 20 pixels) moved across a black background. The four stimuli were created such that the white dot entered each quadrant of the visual display and so that some of the stimuli had similar curving geometric forms. The shapes traced by the white dot included a star, an S, an O, and a swirl starting from the center of the screen spiraling outward. The video dimensions were 1024 $\times$ 640 and the movie files were 23, 18, 14, and 14 seconds in length, respectively. At an average viewing distance of 78 cm the dot moved at 6.0 degree of visual angle/second, a speed that would allow participants to use smooth pursuit eye movements to track the dot. During stimulus presentation each video was preceded by a white fixation cross of 87 $\times$ 93 pixels on a black background presented for 2 seconds.

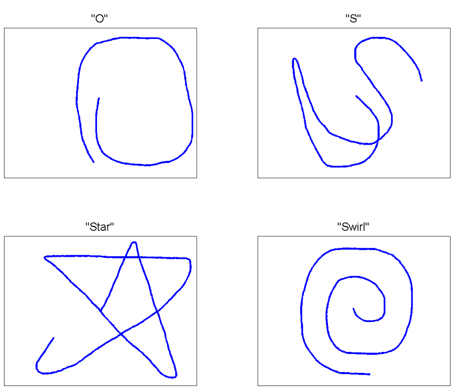

Figure 1: *Four shapes traced by white dot in the smooth pursuit task.*

2.4 Procedure

This study was approved by the Sandia National Laboratories Human Subject Review board. Informed consent was obtained from all participants. Participants were seated in a quiet and darkened room at a distance of 54 to 92 cm from the monitor. Before beginning the eye tracking tasks, the FOVIO was calibrated using a five-point calibration screen. Stimulus presentation was self-paced. Participants were instructed to look at the fixation cross when it appeared and then to follow the white dot as it moved across the screen. The 41 participants generated 164 scanpaths.

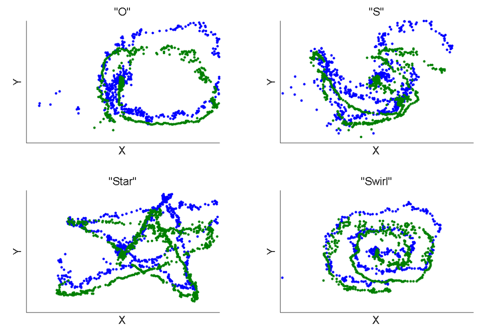

Figure 2: *Sample scanpaths from two randomly chosen subjects for each of four shapes used in smooth pursuit task.*

3 GazeAppraise for Scanpath Analysis

The 164 scanpaths consisting of the x and y position of each sample recorded at 60 Hz were processed using GazeAppraise. In our analysis, we chose four temporal scales, resulting in 10 temporal intervals: (1) the entire scanpath, (2 - 3) the first and second halves of the scanpath, (4 - 6) thirds of the scanpath, and (7 - 10) quarters of the scanpath. Note that the total number of temporal intervals, T, for the number of temporal scales, n, follows the triangle number series,

$$T_n = \frac{n(n+1)}{2}.$$

We began with four temporal scales that had been shown in previous work to minimize computational complexity while providing sufficient resolution to differentiate aircraft trajectories. We found this choice of temporal scales to also be effective in this application to eye movement patterns. Following the Tracktable method of Rintoul et al. [2015] let $SP(t)(t \in [0\ 1])$ represent the entire scanpath, then the set of scanpath temporal intervals is:

(1) $SP(t)(t \in [0\ 1])$

(2 - 3) $SP(t)(t \in [0\ \frac{1}{2}])$ and $SP(t)(t \in [\frac{1}{2}\ 1])$

(4 - 6) $SP(t)(t \in [0\ \frac{1}{3}])$, $SP(t)(t \in [\frac{1}{3}\ \frac{2}{3}])$ and $SP(t)(t \in [\frac{2}{3}\ 1])$

(7 - 10) $SP(t)(t \in [0\ \frac{1}{4}])$, $SP(t)(t \in [\frac{1}{4}\ \frac{1}{2}])$, $SP(t)(t \in [\frac{1}{2}\ \frac{3}{4}])$ and $SP(t)(t \in [\frac{3}{4}\ 1])$.

One or more features can be calculated over each temporal interval. For the smooth pursuit task, we calculated a two dimensional feature at each temporal interval: the median x and y position of the gaze. This metric was chosen because it is a robust statistic; it is less sensitive to noise in the eye tracking samples introduced by the specific eye tracking system or study environment conditions (such as subjects free to move in the eye tracker's head box volume). Let $md(SP[t0\ t1])$ be the median x and y location of the scanpath samples contained in the temporal interval $[t0\ t1]$, then the set of 10, two-dimensional features describing the scanpath is:

(1) $md(SP)[0\ 1]$

(2 - 3) $md(SP)[0\ \frac{1}{2}]$ and $md(SP)[\frac{1}{2}\ 1]$

(4 - 6) $md(SP)[0\ \frac{1}{3}], md(SP)[\frac{1}{3}\ \frac{2}{3}]$ and $md(SP)[\frac{2}{3}\ 1]$

(7 - 10) $md(SP)[0\ \frac{1}{4}], md(SP)[\frac{1}{4}\ \frac{1}{2}], md(SP)[\frac{1}{2}\ \frac{3}{4}]$ and $md(SP)[\frac{3}{4}\ 1]$.

The median calculation is implemented using the BOOST C++ library (www.boost.org) using a P2 quantile estimation algorithm.

To illustrate the feature calculation process, Figure 3 shows the vertical position (y axis) of the gaze of an ideal viewer versus elapsed time for the star smooth pursuit pattern. For this example, an ideal viewer would produce gaze coordinates that exactly match those of the stimulus dot as it moves over the screen. The horizontal lines at the top of the figure show three of the temporal scales used to calculate the median gaze y location at each of six temporal intervals. The triangles indicate the median gaze location feature value calculated at each of the corresponding temporal intervals. The 10, two-

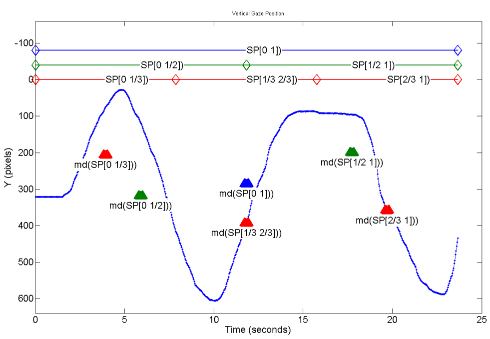

Figure 3: *Vertical position of the gaze of an ideal viewer versus elapsed time for the star stimulus. Three upper, horizontal lines show temporal intervals used to calculate features. Triangles show feature values at each temporal interval.*

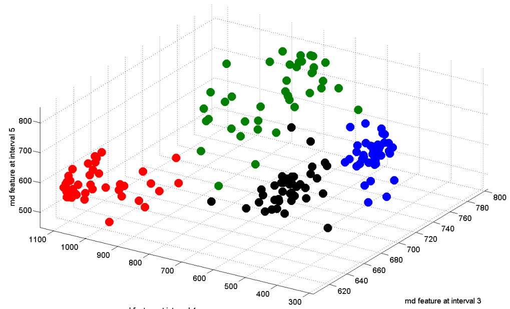

Figure 4: *Three-dimensional view of feature data used in unsupervised clustering. Color indicates cluster membership identified by GazeAppraise.*

dimensional features calculated for the smooth pursuit tasks were represented in 20-dimensional space. Unsupervised cluster analysis was performed using a scale-insensitive approach based on the well-known density based spatial clustering algorithm, DBSCAN [Ester et al. 1996]. For density based clustering, the total number of clusters does not need to be specified a priori. Instead, two intuitive parameters, the minimum number of members required to form a cluster (minPts) and the neighborhood radius (Eps), influence cluster identification. We set minPts equal to 10 (˜1/4th of the total number of subjects) and Eps equal to 200 pixels (˜1/6th of the full horizontal screen width). To illustrate the multidimensionality of these metrics, Figure 4 displays a subset of data. It depicts the x dimension data from three features calculated for the four different stimuli and each participant. Color indicates cluster membership identified by GazeAppraise. This figure also illustrates the need for a density based clustering algorithm rather than other approaches such as k-nearest-neighbors. Clusters are clearly present in the feature set, but the complexity of the cluster boundaries increases with increasing dimensionality of the feature space.

4 Results

Table 1 presents the results of applying the GazeAppraise algorithm to the 164 scanpaths. Unsupervised cluster analysis revealed that 162 of the scanpaths were categorized into one of four clusters and the remaining two scanpaths were considered outliers and not categorized, resulting in a recall/sensitivity score of 98.8%. All of the categorized scanpaths were grouped only with other scanpaths elicited by the same task for a precision = 100%.

Table 1: *Number of scanpaths assigned to each cluster in unsupervised clustering of 164 scanpaths.*

	Stimuli			
Cluster	O	S	Star	Swirl
1	40			
2	0	41		
3	0	0	40	
4	0	0	0	41
Outlier	1	0	1	0

5 Discussion

When GazeAppraise was applied to unprocessed eye tracking data to extract spatiotemporal features, the resulting multidimensional data were clustered into categories that reflected the differences between the original stimuli. This study represents a proof-of-concept: GazeAppraise successfully categorized raw eye tracking samples into distinct scanpaths that reflected the stimulus constraints, but in the absence of stimulus information to constrain the categorization.

One advantage of GazeAppraise is that, unlike previous scanpath analysis techniques (e.g., Multimatch; Dewhurst et al., [2012]), GazeAppraise does not require preprocessing of the eye movement data into fixations and saccades. Calculating these parameters from eye movement data requires assumptions that define which samples are part of fixations and which samples are from saccades. Parsing the eye movement record into discrete units (fixations and saccades) becomes more complex in dynamic environments where fixating a visual stimulus may require smooth pursuit eye movements, or when saccades may not be required to "fixate" a new object because the visual scene has changed.

Another advantage of GazeAppraise is that the approach does not require defining areas of interest or arrays of spatial locations a priori. Rather it classifies similar scanpath shapes together in the absence of stimulus information or knowledge. This ability is important because as visual stimuli become more cluttered and dynamic, the requirement of characterizing the spatial location of important information, or a relevant spatial array, becomes more onerous. Indeed, GazeAppraise can categorize the spatial dependencies between eye movement samples in the absence of a visual stimulus, thus providing a means of characterizing eye movements that are related to visual imagery and mindwandering.

In their guide to eyetracking, Holmqvist and colleagues identified several scanpath comparisons that could be useful [Holmqvist et al. 2011]. Of the seven listed, GazeAppraise has the potential to address four of them: (1) overall shape comparison, (2) similar shape that differs in scale, (3) similarity in position but reversal of order, and (4) differences in the speed of execution of a scanpath.

In this paper, we demonstrated that GazeAppraise can categorize scanpaths from raw eye tracking data, even when those data include samples collected with variations in calibration precision, tracking consistency, and viewer performance. Future work will need to explore how much and in what ways shapes can differ yet still be categorized together. Similarly, scaling algorithms applied to the calculation of x and y features could allow similarly-shaped scan paths that differ in scale to be clustered together, while representation of the reversed order of a set of positions at different temporal scales would also be relatively straightforward to implement.

Although not tested here, it can be mathematically shown that GazeAppraise will cluster together similar scanpaths that vary in temporal duration when the differences in time are distributed evenly across the eye movement samples relative to the duration of the scanpath. It remains to be demonstrated how robust GazeAppraise is to uneven distribution of these temporal differences across a viewing event. For example, it is expected that there would be more temporal variation across individuals in eye movement samples collected during cognitively guided viewing than during saliency guided viewing.

Although this application of GazeAppraise used the median x and y position as features, the metrics used for each feature are flexible. In fact, each feature can have a different units scale, i.e. one feature measured in degrees of visual angle, another measured in milliseconds and another measured in pixels. Thus, features can be any quantity calculable from the eye tracking samples in each temporal interval. Other features that may be useful for scanpath categorization include, but are not limited to, mean and variance of point-to-point distances, mean nearest neighbor distance (randomness of points), total length of scanpath, area and centroid of the convex hull encompassing scanpath points, etc. For example, metrics based on point-to-point distances would implicitly encode the proportion of fixation to saccade activity over the temporal interval. Total scanpath length could measure the amount of the visual display that was viewed which may be important for assessing systematic search processes like visual inspection. Convex hull metrics could measure the amount of the peripheral visual display that is viewed.

The application of GazeAppraise to eye movement analysis is nascent; the eye tracking samples were collected under highly constrained viewing conditions (smooth pursuit eye movements constrained by the stimulus characteristics) not typical of everyday eye movement patterns. It remains to be demonstrated that more typical eye movement trajectories with fixations and saccades, that are influenced to a greater extent by top-down processes, can be categorized. The contribution of this research is to demonstrate the application of a new set of spatiotemporal trajectory libraries to raw eye tracking data, an application we refer to as GazeAppraise. Categorization of eye tracking data collected while viewing four different, but constraining, stimuli was highly successful. Future work will validate the usefulness of this approach by applying the algorithm to eye tracking data from systematic search tasks.

Acknowledgements

We wish to thank Danny Rintoul and Andy Wilson of Sandia National Laboratories, Albuquerque, NM, USA, for their assistance adapting Tracktable to eye tracking data.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energys National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND_2016_0144_C.

References

ANDERSON, N. C., BISCHOF, W. F., LAIDLAW, K. E., RISKO, E. F., AND KINGSTONE, A. 2013. Recurrence quantification analysis of eye movements. *Behav Res Methods 45*, 3, 842–856.

BORJI, A., AND ITTI, L. 2014. Defending yarbus: eye movements reveal observers' task. *Journal of Vision 14*, 3.

BUTLER, K. M., AND ZACKS, R. T. 2006. Age deficits in the control of prepotent responses: evidence for an inhibitory decline. *Psychol Aging 21*, 3, 638–643.

CRISTINO, F., MATHOT, S., THEEUWES, J., AND GILCHRIST, I. D. 2010. Scanmatch: a novel method for comparing fixation se-quences. *Behav Res Methods 42*, 3, 692–700.

DEWHURST, R., NYSTROM, M., JARODZKA, H., FOULSHAM, T., JOHANSSON, R., AND HOLMQVIST, K. 2012. It depends on how you look at it: scanpath comparison in multiple dimensions with multimatch, a vector-based approach. *Behav Res Methods 44*, 4, 1079–1100.

ESTER, M., KRIEGEL, H., SANDER, J., AND XU, X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD-96 Proceedings. Second International Conference on Knowledge Discovery and Data Mining*, 226–231.

GREENE, M. R., LIU, T., AND WOLFE, J. M. 2012. Reconsidering yarbus: a failure to predict observers' task from eye movement patterns. *Vision Research 62*, 1–8.

HAJI-ABOLHASSANI, A., AND CLARK, J. J. 2014. An inverse yarbus process: predicting observers' task from eye movement patterns. *Vision Research 103*, 127–142.

HENDERSON, J. M., SHINKAREVA, S. V., WANG, J., LUKE, S. G., AND OLEJARCZYK, J. 2013. Predicting cognitive state from eye movements. *PLoS One 8*, 5.

HENDERSON, J. M., CHOI, W., LUKE, S. G., AND DESAI, R. H. 2015. Neural correlates of fixation duration in natural reading: Evidence from fixation-related fmri. *Neuroimage 119*, 390–397.

HOLMQVIST, K., NYSTRÖM, M., ANDERSSON, R., DEWHURST, R., JARODZKA, H., AND VAN DE WEIJER, J. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press.

JARODZKA, H., HOLMVQUIST, K., AND NYSTRÖM, M. 2010. A vector-based, multidimensional scanpath similarity measure. *Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications*, 211–218.

RINTOUL, M., WILSON, A., VALICKA, C., SHEAD, T., RODRIGUEZ CZUCHLEWSKI, K., KEGELMEYER, W., AND NEWTON, B., 2015. Panther: Trajectory analysis. Technical report.

YARBUS, A. L. 1967. Eye movements and vision. *Plenum Press*.