Real Time Eye Movement Analysis Framework: Objective-Based Systematic Approach

Ziho Kang
School of Industrial and Systems Engineering
University of Oklahoma
Norman, OK, USA.
zihokang@ou.edu

Abstract **— As eye tracking is becoming available in wearable or portable devices, eye movement (EM) analysis will be used to provide real time alerts, feedback, and/or assistive information for diversified objectives when performing tasks. However, we lack the methodologies to effectively analyze real time EM attributes including eye fixation numbers, durations, and especially visual scanpath sequences (i.e. time-ordered eye fixation sequences). Therefore, a systematic approach to analyze real time EM data is introduced through proposing the concepts of hierarchical areas of interest (HAOIs) and objective-based EM attribute thresholds. The framework will facilitate the automatic analysis of real time EM data with minimal human intervention or analysis efforts.**

Keywords – eye movement, eye tracking, gaze, scanpath, scanpath sequences, wearable technology, real time analysis.

I. INTRODUCTION

Eye movement (EM) analysis, also known as eye tracking or gaze analysis, will soon become ubiquitous in our everyday lives as low cost wearable and portable eye tracking devices are being delivered into the market. Furthermore, with the development of prototype devices that can track eye movements within wearable virtual reality (VR) devices, we will be able to explore how humans interact with other virtual humans or objects within the VR environment. Real time EM analysis in both real and VR environments can be used as an alerting method when the operators are involved in high risk related tasks. In addition, timely feedback or assistive information can be provided to the trainees who are struggling to perform complex tasks at hand or understand complex procedures or instructions.

However, EM data analysis results can vary based on the EM research objectives when conducting tasks, and the results are affected by how we define the areas of interest (AOIs) on the display and how we configure the thresholds to extract the eye fixation numbers, durations, and visual scanpath sequences. We currently lack conceptual designs and methodologies to process and analyze real time EM data when considering those factors, therefore, a systematic and integrated framework to process and analyze real time EM data is required. Details are as follows.

II. OVERVIEW AND ISSUES

EM analysis has been applied to investigate human's cognitive processes or human performance in diverse areas such as air traffic management [1-2], consumer behavior [3], driving behavior [4], and training/education [5-6]. Many EM analysis methods are based on post-processing of EM data, meaning that the EM data is analyzed after a task has been completed. A thorough literature review of the post-processing analysis methods and algorithms provided in [2] highlights the various approaches (e.g. string edit algorithm, Markov chain, vector-based approach, MTAHC) that we can apply to analyze complex EM data, but these methods are not yet tailored for real time analysis. Although there have been recent advancements in real time EM analysis [7-9], the analysis methods have been either geared towards using EMs as a control method (e.g. eye writing) [7] or applying EM data using a single threshold such as 250ms to detect an eye fixation [8] or using twelve recent eye fixations for real time analysis [9]. Furthermore, the researches explained above use nonhierarchical AOIs that are mapped one to one with the areas or objects on the display, and we might need to tailor the AOIs based on whether we need conduct the EM analysis in an aggregate level or a more detailed level.

Therefore, when analyzing real time EM data, we need to (1) determine how to define the important areas of interest (AOIs) within the display or scenery based on the research objective(s), (2) utilize the EM data including multiple EM attributes, and (3) determine which recursive time frames or amounts of accumulated EM data should be applied for analysis. Some details to the challenging issues are as follows:

- EM data only return pixel based spatial coordinates of the time-ordered eye fixations, and the analyst needs to define (i.e. plot) the AOIs around a data block or an object shown on the computer or VR display. During post-processing of EM data, the analyst can freely modify the shapes, sizes, or spatial locations of the AOIs to better investigate the EM characteristics; however, in real time, we need to have a systematic approach to minimize the intervention of the analyst.
- Raw EM data are usually collected between 30hz (low-end devices) and 300hz (high-end devices), and the raw data have to be further processed through algorithms such as I-VT or I-DT [10-11] to produce what we call "eye fixations." However, there are no unified approaches to define the threshold when applying such algorithms, and thresholds such as 50ms [12], 100ms [13-14], or 250ms [8] have been

978-1-5386-0706-0/17/$31.00 ©2017 IEEE

used based on the analyst's discretion. Furthermore, EM research articles typically do not explain in detail why such threshold was applied.

- For effective real time or near real time analysis, we need to determine the amount of EM data to analyze at hand. In [9], a single threshold of using twelve most recent eye fixations was applied, but it is possible to apply other thresholds to analyze most recent eye fixation numbers, durations, and/or scanpath sequences.

III. PROPOSED FRAMEWORK

In this work, a framework is introduced that can analyze the real-time EM data through multiple paths. The multiple paths are created through (1) creating hierarchical areas of interest (HAOIs) based on the task objectives and (2) controlling the thresholds of eye fixation numbers and durations over short time periods to sort out relatively more important EM data. The overall simplified process is provided in Fig. 1 followed by some detailed descriptions and examples in the following sub-sections. In Fig. 1, the minimum threshold is set based on the 30hz device, and Tc, Nc, and Sc refers to the duration times of each eye fixation, eye fixation numbers, and scanpath sequence length up to the current point in time, respectively. A scanpath sequence is ~~be~~ composed of at least two eye fixations. An example sequence can be "ABCDE" in which each letter indicates an eye fixation on an AOI (i.e. "A" through "E"). In some cases, if we do not consider consecutive eye fixations that occurred within the same AOI, we can use a "collapsed" scanpath sequence (e.g. "AABBBCCCCDEE" is transformed to "ABCDE").

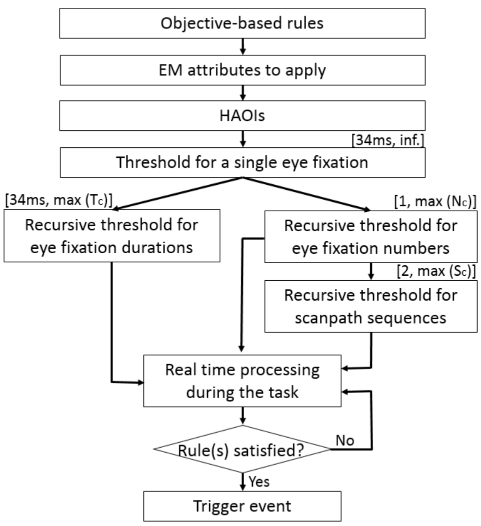


Fig. 1. Simplified process of real time EM analysis.

A. *Designing hierarchical areas of interests (HAOIs)*

HAOIs refer to AOIs that have levels of meanings, and lower level HAOIs can only belong to single higher level HAOI (Fig. 2). The concept of HAOI is developed based on the objective(s), and can enable faster real-time EM analysis since only some necessary HAOIs can be used for a certain objective. For example, consider the overall display as the car driver's front window view, the higher level HAOI as the pedestrian crossing the road, and the lower level HAOIs as the pedestrian's head, torso, arms, and legs. We would be interested in whether the driver is paying attention to the pedestrian as a whole (i.e. higher level HAOI) rather than which lower level HAOIs were observed, therefore, real time analysis would be performed only using the higher level HOAIs (e.g. pedestrians).

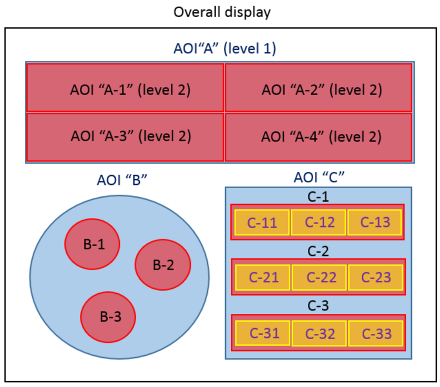


Fig. 2. HAOIs: Three important AOIs within the overall display is defined as "A," "B," and "C" that are higher level (or level 1) HAOIs. Within each higher level HAOI, lower level (or level 2) HAOIs are defined if more specific EM analysis is required. More levels can be created if needed.

In some cases, we might need to analyze different HAOI levels based on multiple objectives within a single task. For example, when performing a reading task, we might be interested in which paragraph (level 1 HAOI), sentence (level 2 HAOI), or word (level 3 HAOI) has been and/or is being read. As another example, for an air traffic control (ATC) task, we might be interested in (1) which areas within the radar screen (e.g. terminal, southwest approach, or north enroute areas) the specialist has been observed or interrogated, and (2) which aircraft pairs within a specific area has been interrogated. The framework will enable faster processing of the multiple objectives by only utilizing certain HAOI levels instead of utilizing all HAOI levels.

B. *Controlling thresholds to determine an eye fixation.*

Eye fixations based on small thresholds (50ms or 100ms) can show seemingly stochastic EMs due to involuntary jerking and/or adjusting movements that can occur when locating and/or reading a data block. By systematically increasing the threshold that defines an eye fixation, we can identify more important eye fixations and their durations, and furthermore obtain crisp scanpath sequences (Fig. 3).

The threshold for a single eye fixation can be controlled based on the objective type and/or through a step-wise increase of the threshold. One of the many ways to control the threshold is to identify whether only the spatial location of the object was observed or whether specific information of the object was read. Within our visual sensory system, the spatial location of an object is governed by our dorsal stream (that process information of "where") which is processed faster, and the content of the object is governed by the ventral stream (that process the information of "what" such as forms or words) which takes more time to process [15]. However, one size does not fit all, meaning that even if we apply the above theory, each person should be applied with adjusted thresholds since the EM characteristics can be idiosyncratic among the individuals. Applying the stepwise threshold increase for each individual as part of a sensitivity analysis (as shown in Fig. 3) might address such an issue.

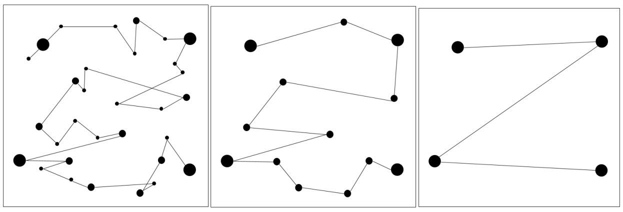

(a) (b) (c)

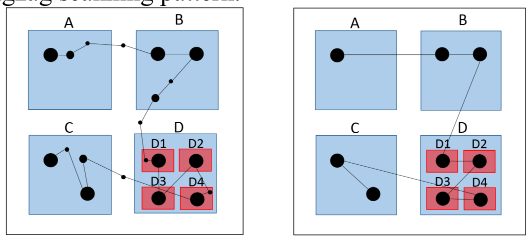
Fig. 3. EM data output based on increasing the eye fixation threshold: The dots indicate the eye fixations, the size of the dots is proportional to the eye fixation duration, and the lines indicate the saccades among the eye fixations. (a) shows the case when all eye fixations are considered based on a 50 ms threshold. (b) shows a more refined EM data ouput when threshold is increased (i.e. 100ms). (c) shows a clear "Z"-type zigzag pattern and the four important longer duration eye fixations as the threshold is further increased (i.e. 200ms).

For example, during driving, an objective is to only know the location of the pedestrians to avoid an accident, therefore, a shorter threshold should be applied. However, for a reading task, we might not be interested in the detailed reading pattern of each reader, but a simplified pattern of how certain groups of words or mathematical expressions that the reader has been and/or is gazing at for longer periods. For an ATC task, the specialists read detailed information on the radar display such as the altitude and/or speed of each aircraft while the spatial location of each aircraft is registered when reading such data, therefore a longer threshold will provide more pertinent EM data through filtering out short eye fixations.

C. *Controlling thresholds of EM attributes to determine the amount of accumulated data to analyze*

Which EM attribute to analyze and for how long is controlled based on the rules implemented according to the task objectives rather than using fixed values as in noted in section II. It is not necessary to use all the EM attributes, and only the necessary attributes can be used based on the objectives. Note that scanpath sequences can become complex even for a short time duration such as tens of seconds [2].

Fig. 4 shows an example of how we can control the amounts of EM data to analyze. The thresholds can be selected based on different rules. For example, if the objective is to identify the visual scanning pattern of the four larger regions within the display, the analysis starts once an eye fixation occurs at one of the regions until eye fixations occurred on all four regions and re-visited any other region. Note that this process is recursive. In the case shown in Fig. 4, the rule is applied using level 1 HAOIs, increased eye fixation threshold (200ms), and only the scanpath sequence. The collapsed scanpath sequence in Fig. 4 (b) transforms "ABBDDDDCC" to "ABDC," therefore, returns an output of a circular (i.e. counterclockwise) scanning pattern. Similarly, if the objective is to identify the visual scanning pattern within only the region "D," the analysis starts once the eye fixation occurs within "D" and ends when the eye fixation occurs on another level 1 HAOI (i.e. "A," "B," or "C"). The collapsed scanpath sequence "D1D2D4D4," therefore, returns an output of a "Z"-type zigzag scanning pattern.



(a) (b)

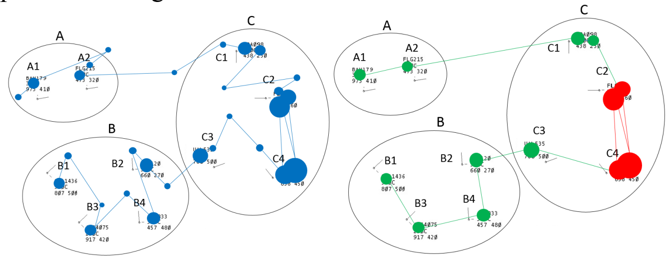
Fig. 4. Scanpath sequence analysis example: (a) shows the scanpath outputs using a 50ms eye fixation threshold. (b) shows the cleansed scanpath based on the increase threshold (i.e. 200ms).

To fulfill the objective of avoiding a pedestrian accident for a driving task, we would only need to analyze the eye fixations and/or durations when pedestrians are present within the projected location of the driver's car. For a reading task, we would only need to analyze a few eye fixation numbers and/or durations to provide immediate assistance or feedback when a single difficult word or phrase is read, or analyze the visual scanning patterns to find out which words or phrases were heavily interrogated when a difficult paragraph is read. For a complex ATC task, multiple objectives (such as whether all aircraft within a certain region(s) were observed within a short time frame, whether certain aircraft pairs or groups within a region were heavily interrogated, or whether different scanning patterns were used to avoid tunnel vision) that will require the application of different rules and EM attributes.

IV. CASE STUDY

A representative output is provided for the more complex task, the ATC task, to verify some possibilities of the proposed framework. A training environment was assumed, meaning that an instructor would receive warnings that would trigger in real time as a trainee is performing an aircraft conflict detection task (i.e. detecting aircraft pairs or groups that will have a collision or near-collision in the near future). Therefore, the HAOIs, characterized trainee's EMs, and the triggered warnings would only be shown/provided to the instructor.

Real time simulated data was used to mimic a trainee's EMs, and the simulated data was created from (1) analyzing the ATC specialists' EM data collected from previous experiments used for post-processing analysis and (2) their identified EM characteristics (i.e. fixations, durations, and scanpaths) [1, 16]. The simulated real time EM data were processed through the prototype analysis software developed in JAVA. The simulated data were created based on the same settings that were used to collect the specialists' data (i.e. 60hz sampling rate, 0.5^{o} degrees of visual angle accuracy, I-VT algorithm). Two HAOI levels were created. Specifically, three dynamic regions (level 1 HAOIs) on the radar display were created based on the proximity and the convergence of multiple aircraft, and ten aircraft and their associated data blocks were assigned as level 2 HAOIs. The level 2 HAOIs were designed based on the convex area approach introduced in [2]. The visualized output for a short period of 8 seconds is provided in Fig. 5.



(a) (b)

Fig. 5. Visualized EM data output: (a) shows the original scanpath with eye fixation threshold set at 50ms. (b) shows the scanpath when the eye fixation threshold was increased to 200ms (green) and 400ms (red).

Two pertinent objectives out of the many possible objectives explained above were analyzed. The first objective was: "If an eye fixation enters a certain dynamic region (i.e. "A," "B," or "C"), then identify whether the trainee has interrogated all aircraft within the region (i.e. whether at least one data of each aircraft such as altitude has been read) before moving on to another region." The rule was: "If the objective is not met, trigger a warning by indicating which aircraft was not interrogated." In this case, only the level 2 HAOIs linked to the currently observed single level 1 HAOI were used. A 200ms threshold was used to define an eye fixation based on the minimum threshold to read a data point such as 360 (or 36,000 ft.). Only the eye fixation numbers (i.e. at least one eye fixation) were evaluated. Analysis started when an eye fixation first occurred within a region (e.g. HAOI "A") and ended when it occurred in another the region (e.g. HAOI "B"). Warning was not triggered since all aircraft within the current region were interrogated before moving on to another region (Green and Red eye fixations in Fig. 5(b)).

The scenario was developed to create a certain conflict, and the trainee's task is to identify the conflict; therefor, the second objective was: "Identify whether the aircraft pairs that will have conflicts are heavily interrogated back and forth by the trainee." The rule was: "If not heavily interrogated (i.e. if there are no back and forth EMs), trigger an alert." In this case, only the conflicting pair within the level 2 HAOIs (i.e. C2 and C4) was used. The eye fixation threshold was set at 400ms since at least two or more data such as altitude and speed need to be read, and only the collapsed scanpath sequences were analyzed. Analysis started when one of the two HAOI is visited and end when an eye fixation occurs on a third HAOI. Warning was not triggered as the collapsed sequence was "C2C4C2C4" (Red eye fixations in Fig. 5(b)).

V. Conclusions and Future Work

The real-time EM analysis framework shows promise on timely application of appropriate EM attributes based on how we define the objectives rather than using all the EM attributes and fixed thresholds. Future work includes developing more detailed classification procedures for the objectives, HAOIs, and the EM attribute thresholds, and adapting post-analysis methods for real time analysis. The framework not only supports effective real time eye movement analysis, but can also serves as an integral part of the multimodal training methods.

Acknowledgment

This research was funded by the FAA Center of Excellence: Project No. A17-0160 and A17-0162.

References

[1] Z. Kang and S.J. Landry, "An eye movement analysis algorithm for a multi-element target tracking task: Maximum transition-based agglomerative hierarchical clustering," *IEEE Trans. on Human-Machine Syst.*, vol. 45(1), pp. 13-24, 2015.

[2] Z. Kang, S. Mandal, J. Crutchfield, A. Millan, and S. McClung, "Design concepts and algorithms in eye tracking research for a multi-element objects tracking-and-controlling task to support human performance analysis," *Comput. Intell. and Neurosci.*, 1), vol. 1, pp. 1-18, 2016.

[3] R.V. Menon, V. Sigurdsson, N.M. Larsen, A. Fagerstrøm, and G.R. Foxall, "Consumer attention to price in social commerce: Eye tracking patterns in retail clothing," *J. of Business Res.*, vol. 69, no. 11, pp. 5008-5013, 2016.

[4] R. Zheng, K. Nakano, H. Ishiko, K. Hagita, M. Kihira, and T. Yokozeki, "Eye-Gaze Tracking Analysis of Driver Behavior While Interacting With Navigation Systems in an Urban Area." *IEEE Trans. on Human-Machine Syst.*, vol. 46, no. 4, pp. 546-556, 2016.

[5] M.L. Lai, M.J. Tsai, F.Y. Yang, C.Y. Hsu, T.C. Liu, S.W.Y. Lee, M.H. Lee, G.L. Chiou, J.C. Liang, and C.C. Tsai, "A review of using eye-tracking technology in exploring learning from 2000 to 2012," *Educ. Res. Review*, vol. 10, pp. 90-115, 2013.

[6] J.L. Rosch and J.J. Vogel-Walcutt, "A review of eye-tracking applications as tools for training," *Cognit., tech. & work*, vol. 15, no. 3, pp. 313-327, 2013.

[7] K.R. Lee, W.D. Chang, S. Kim, and C.H. Im, "Real-Time "Eye-Writing" Recognition Using Electrooculogram," *IEEE Trans. on Neural Syst. and Rehabilitation Eng.*, vol. 25, no. 1, pp. 40-51, 2017.

[8] H.J. Wee, S. W. Lye, and J.P. Pinheiro, "Real Time Eye Tracking Interface for Visual Monitoring of Radar Controllers," In *AIAA Modeling and Simul. Tech. Conf.*, Grapevine, TX, 2017, pp. 1317-1331.

[9] S.S. Parikh and H. Kalva, "Real Time Learning Level Assessment Using Eye Tracking." *Workshop on 2017 Comput. and Math. Models in Vision*, 2017. Retreieved May, 2017 from http://docs. lib.purdue. edu/modvis/2017/session04/.

[10] D.D. Salvucci and J.H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," In *Proc. of the 2000 Symp. on Eye tracking Res. & Appl.*, ACM, 2000, pp. 71-78.

[11] O.V. Komogortsev, D.V. Gobert, S. Jayarathna, D.H. Koh, and S.M. Gowda, "Standardization of automated analyses of oculomotor fixation and saccadic behaviors," *IEEE Trans. on Biomedical Eng.*, vol. 57, no. 11, pp. 2635-2645, 2010.

[12] F. Cristino, S. Mathôt, J. Theeuwes, and I.D. Gilchrist, "ScanMatch: A novel method for comparing fixation sequences," *Behav. Res. Methods*, vol. 42, no. 3, pp. 692-700, 2010.

[13] A. Çöltekin, S.I. Fabrikant, S.I., and M. Lacayo, "Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings," *Int. J. of Geograph. Info. Sci.*, vol. 24, no. 10, pp.1559-1575, 2010.

[14] J.H. Goldberg and X.P. Kotval, "Computer interface evaluation using eye movements: methods and constructs," *Int. J. of Indust. Ergo.*, vol. 24, no. 6, pp. 631-645, 1999.

[15] A.V. Barber, "Visual mechanisms and predictors of far field visual task performance," *Human Factors: The J. of the Human Factors and Ergonom. Soc.*, vol. 32, no. 2, pp. 217-233, 1990.

[16] S. McClung and Z. Kang, "Characterization of Visual Scanning Patterns in Air Traffic Control," *Comput. Intell. and Neurosci.*, vol. 1, pp. 19-35, 2016.