Spatiotemporal windows for fixation detection

Tyler Thrash and Iva Barisic

Chair of Cognitive Science, ETH Zurich
{tyler.thrash,iva.barisic}@gess.ethz.ch

Abstract. Eye fixations are periods of relative stability derived from continuous eye position (or eye movement) data. In order to define eye fixations, researchers often assume that the eye(s) will not move beyond a particular spatiotemporal window (i.e., a spatial area towards which the eye is directed within a particular period of time). However, exact specifications of this window vary from field to field and even from one experiment to another. Efforts to standardize these specifications have assumed (either implicitly or explicitly) that there is one appropriate window size for describing eye behavior. The present paper explores an alternative approach. Specifically, we provide a method for determining the most appropriate spatiotemporal window that can vary from participant to participant (or task to task). This approach may also be extended to provide a metric for detection algorithm comparison.

Keywords: eye tracking • fixation detection • scene perception

1 Introduction

In order to be meaningful, eye tracking data needs to be classified into periods of movement (e.g., saccades) and periods of stability (e.g., fixations). During periods of movement, visual stimuli are usually considered inaccessible to the human observer. This phenomenon is called *saccadic suppression* [1]. Most of visual perception is based on information that is accessible during periods of stability [2]. Fixation detection algorithms attempt to determine what information is perceptually available by inferring which eye tracking data points represent periods of stability [3].

All of these algorithms essentially rely on the definition of what we call a “spatiotemporal window” (i.e., a spatial area towards which the eye is directed within a particular period of time). Some detection algorithms (e.g., dispersion-based algorithms; [4]) emphasize the two spatial dimensions of this window by evaluating possible fixations in terms of the dispersion of data points around possible foci. However, these algorithms also typically incorporate lower and upper bounds for the “reasonable” duration of a fixation. Other detection algorithms (e.g., velocity-based algorithms; [4]) emphasize the temporal dimension of this window by classifying eye tracking data in terms of velocity and/or acceleration. These algorithms also typically include lower and upper bounds for the size of a fixation along spatial dimensions. Thus, the three-dimensional spatiotemporal window is a critical consideration for the implementation of both dispersion-based and velocity-based algorithms.

ET4S 2014, September 23, 2014, Vienna, Austria

One assumption underlying most efforts to standardize specifications of the spatiotemporal window is that one set of parameters can be used to describe the eye behavior of all healthy adults (e.g., [5]), even though there is a good deal of variability in this behavior both within an individual and across individuals [6] [7] [8]. The variability not described by this set of parameters is typically considered “noise” (e.g., as resulting from the imprecision of the eye tracking equipment). Even algorithms that can be adapted to different noise profiles (e.g., [4]) assume the same spatiotemporal window for defining fixations. In contrast, the current approach allows for variability in the size of the spatiotemporal window across individuals and tasks.

The specification of spatiotemporal windows is especially critical when it is difficult to define the direction of a stimulus from the observer objectively (i.e., without relying on designations by other observers). This scenario is common for investigations of naturalistic scene perception and navigation because of the lack of clear boundaries between objects and/or the dynamic nature of the stimuli [9]. Except for sophisticated computational vision algorithms, there are no established methods for determining the objective “truth” to which a set of detected fixations (e.g., resulting from different detection algorithms) can be compared in these scenarios. The current approach extends a common technique for comparing mathematical models without needing to presuppose any particular objective truth.

2 Current approach

There are two primary applications of the current approach: the specification of the spatiotemporal window for different observers/tasks and the comparison of different detection algorithms.

2.1 Specification of the spatiotemporal window

Our general approach for specifying the most appropriate spatiotemporal window is to calculate error in the data points relative to the nearest detected fixation. Error, in this case, represents variability in the gaze data that is within the defined spatiotemporal window but cannot be explained by the set of fixations detected by a particular algorithm.

At most, six parameters are needed to describe spatiotemporal windows that reflect plausible (and interpretable) fixations. Researchers should start by defining the sizes of spatial and temporal intervals. The spatial and temporal interval parameters determine which data points are used for calculating the error term of each detected fixation. Data points are only included in the following calculations if they fall within both spatial and temporal intervals for any detected fixation. The distance function is calculated using the following equation:

$$d(p_1, p_2) = [w_1(x_1 - x_2)^m + w_2(y_1 - y_2)^m + (1 - w_1 - w_2)(t_1 - t_2)^m]^{\frac{1}{m}} \quad \text{(1a)}$$

Here, x_1 and x_2 represent the locations of two points along the horizontal axis, y_1 and y_2 represent the locations of two points along the vertical axis, t_1 and t_2 represent the locations of two points along the temporal dimension, the two *w*'s represent the relative weighing of the two spatial dimensions with respect to the temporal dimension, *m* represents the type of Minkowski distance metric, and $d(p_1,p_2)$ represents the distance between two points. For most applications, *m* should be constrained to be either *2* (resulting in a Euclidean distance metric) or *1* (resulting in a city-block distance metric). A city-block distance metric may be appropriate if researchers consider errors along *x* and *y* dimensions as independent of each other. Other values for *m* are possible but difficult to interpret. The parameters w_1 and w_2 also need to be constrained so that each weight is greater than *0* and that their sum is less than *1*. Larger values for the *w*'s indicate larger relative contributions for deviations along the corresponding spatial dimensions to the fit of the resulting model. Note that this distance function may need to accommodate differences in visual angle if, for example, two participants are fixed at different distances from the stimulus.

Equation 1a also assumes that the distribution of data points that represent each fixation is uniform rather than Gaussian (see, e.g., [10]). The utility of the uniformly distributed distance function can be compared empirically to the utility of the following normally distributed (and Euclidean) distance function:

$$d(p_1, p_2) = \sqrt{\left(\frac{1}{s\sqrt{2\pi}}\right)\left\{\begin{matrix} w_1\left[1 - e^{-\frac{(x_1-x_2)^2}{2s^2}}\right] \\ +w_2\left[1 - e^{-\frac{(y_1-y_2)^2}{2s^2}}\right] \\ +(1-w_1-w_2)\left[1 - e^{-\frac{(t_1-t_2)^2}{2s^2}}\right] \end{matrix}\right\}} \tag{1b}$$

Here, the only additional parameter is *s*, which represents the "steepness" of the normally distributed distance function. Note that *s* does not necessarily correspond to the standard deviation of the distribution of resulting distances. The *w*'s should be constrained in the same manner as for the uniformly distributed distance function.

In order to determine which of several possible specifications is most appropriate for a particular detection algorithm, we then need to calculate the error term for each fixation:

$$e(fixation) = \frac{\sum d(p_i, \bar{p})}{n_p} \tag{2}$$

Here, *p* represents a data point with index i, $\overline{p}$ represents the centroid for all of the data points within the spatiotemporal window, *d* represents the distance metric from Equation 1a or 1b, n_p represents the number of data points within the spatiotemporal window for a detected fixation, and *e(fixation)* represents the error term for the detected fixation (i.e., the mean of the distances from the centroid to each data point within the spatiotemporal window).

If researchers are comparing sets of detected fixations with spatiotemporal windows of the same size and shape, then sums of *e(fixation)* across sets of detected fixations are sufficient for comparing different detection algorithms. Across any range of spatial and temporal intervals, the smallest sum of *e(fixation)* will reveal the most appropriate spatiotemporal window for any given detection algorithm.

However, in order to compare spatiotemporal windows with different shapes or sizes, the error term needs to be converted into a measure that accounts for the number of free parameters or the number of detected fixations, respectively. Towards this end, the summed and squared error terms for all of the detected fixations of a given spatiotemporal window can be converted to Bayes' information criterion *(BIC*):

$$BIC = \left\{ n_f \times \ln \left[\frac{\sum e(fixation)^2}{n_f - 1} \right] \right\} + \left[k \times \ln(n_f) \right] \qquad (3)$$

Here, n_f represents the number of detected fixations, *k* represents the number of free parameters, *ln* represents the natural logarithm function, and *e(fixation)* represents the error term from Equation 3. We consider each interval as only one parameter because the location of the fixation along a particular dimension and both boundaries of each interval are completely constrained by the determination of the size of the interval and the data.

2.2 Detection algorithm comparison

The *BIC* can also be used in order to compare different fixation detection algorithms using Equations 1-3. The primary challenge for comparing different detection algorithms thus becomes determining which parameters are free to vary (see [11]). The *BIC* should be used to penalize the fit of any parameter that could have changed in order to improve the fit of the model to the data. Notably, this method does not require any assumptions regarding the "true" foci in the stimulus.

3 Future validation studies

Future investigations can attempt to validate or invalidate our approach in at least two ways. First, following [5], researchers can direct participants to focus on individual stimuli at known coordinates. This procedure is often used by eye tracking software for calibrating eye movement data before an experiment [12]. For validation purposes, fixations may be considered the periods of time during which a participant was asked to focus on a particular stimulus. The veracity with which the *BIC* metric determines the most appropriate spatiotemporal window (or best-performing detection algorithm) should then be reflected by similar patterns in other metrics (e.g., number of detected fixations; [5]).

Second, the mean spatiotemporal window specified across individual participants may approximately correspond to established recommendations already in the literature (e.g., [5] [13]). This may occur if the primary advantage of the current approach is to account for additional variability, but this procedure could also be misleading if the current approach actually produces more accurate fixation detection than previous approaches.

4 Conclusions

The present paper provided a novel approach to the specification of spatiotemporal windows for fixation detection algorithms. This approach may also be applied to the comparison of different detection algorithms. Two future studies for potentially falsifying this approach are also briefly described.

5 References

1. Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin, 81,* 899-917.
2. Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7(11),* 498-504.
3. Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Eye Tracking Research and Applications Symposium,* 71-78.
4. Nyström, M. & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods, 42,* 188-204.
5. Komogortsev, O. V., Gobert, D. V., Jayarathna, S., Koh, D. H., & Gowda, S. M. (2010). Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on Biomedical Engineering, 57*, 2635-2645.
6. Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372-422.
7. Hyönä, J., Lorch Jr, R. F., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology, 94(1)*, 44.
8. Rayner, K., & Raney, G. E. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review, 3(2)*, 245-248.
9. Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. *Eye guidance in reading and scene perception, 11*, 269-293.
10. Santella, A., & DeCarlo, D. (2004). Robust clustering of eye movement recordings for quantification of visual interest. *Proceedings of the Eye Tracking Research and Applications Symposium,* 27-34.
11. Lewandowsky, S. & Farrell, S. (2010). *Computational Modeling in Cognition: Principles and Practice*. Thousand Oaks, CA, USA: Sage Publications.
12. Hornof, A. J., & Halverson, T. (2002). Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behavior Research Methods, Instruments, & Computers, 34(4)*, 592-604.
13. Salthouse, T. A., & Ellis, C. L. (1980). Determinants of eye-fixation duration. *The American Journal of Psychology, 93,* 207-234.