The statistical challenge of scan-path analysis

Moreno I. Coco
School of Informatics
University of Edinburgh
M.I.Coco@sms.ed.ac.uk

Abstract

By knowing what people look at while performing tasks, we would be able to better design devices that facilitate users' interaction. Eye tracking is a useful technique to study visual attention but the data are often spatially and temporally complex to analyze. To this aim more sophisticated analytical tools are needed to give meaning to data, especially when both the spatial and temporal components are to be taken in synchrony. In this paper we review statistical methods for spatial and temporal analysis of eye-movement data pinning down advantages and disadvantages while proposing new directions of research.

1. Introduction

Understanding where people focus their attention while performing tasks has became particularly crucial for the field of human-system interaction. By knowing how attentional resources are deployed it is possible to design efficient systems that minimize the effort of interaction. Web navigation is the domain in which the necessity to minimize the users' effort is greatest. The massive amount of similar informational resources available on internet means websites compete for users' attention. Therefore, the principal goal of web-design is to provide the information needed while minimizing attentional efforts. An optimal interaction is achieved when the user is able to visualize and grasp the information of the message communicated at first glance.

2. Eye-tracking research in HSI

A particularly powerful experimental technique to observe the shift of visual attention during the performance of tasks is eye-tracking. Eye-tracking is a monitoring technique used to sample spatial (coordinates) and temporal (time course in milliseconds) data of oculomotor movements (fixations and saccades).

Eye-movements are an explicit measure of visual attention that can be used as an implicit indicator of cognitive processing. Different types of eye-tracking studies have been conducted on the usability of the web, spanning different topics, from user related differences like gender, social status and instruction [8], [1], to structural organization of information [4], [6], [10]. The results have been obtained using a range of diverse statistical analysis techniques to provide explanation of specific problems. Due to the complexity of eye-movement data, the analysis often collapses the dynamism of temporal (e.g. gaze duration) and spatial (e.g. **R**egion **O**f **I**nterest) measures within a static frame (e.g mean of gaze duration over different spatial regions).

Therefore, despite the relevance of the results obtained, a common problem of these and many other eye-tracking studies has been the lack of a shared statistical methodology that is standardly used to visualize and inferentially analyze the visual behavior.

An ideal analysis of eye-movement data should treat synchronously spatio-temporal information, generalizing the visual responses of different viewers on the same task.

This ideal measure is known in visual research as *scan-path*. The statistical analysis of scan-paths is a core problem of visual research that is still unsolved.

The aim of this study is to briefly review some of the methods applied within the human-system interaction literature to tackle the problem of scan-paths presenting advantages and limits, while advancing ideas for future research.

3. Scan-paths

A scan-path represents the exact spatial sequence of eye-movements performed by a participant during a task. The scan-path reflects overtly the unfolding of visual attention over time, indicating precisely which contents of the visual context are attended. A scan path for a single participant is very easy to derive and most of the commercial software dealing with eye-movements (e.g. Data Viewer) have already

978-1-4244-3960-7/09/ $25.00 2009 IEEE

implemented a function that shows the scan path produced by a participant for a given single trial. However, the complexity of the problem grows exponentially as soon as we have to generalize the scan path of different subjects on the same trial. Each subject may perform a completely different scan-path and there are no clear methods to average them. The difficulty of averaging is due to both spatial and temporal causes. Spatially, we can have regions of interests that share the same semantic characteristics (e.g. drop down menu of air lines) but that are located at different coordinates (e.g. two different air line internet sites). Temporally, the participants can perform the same task (e.g clicks on the menu) but at different points in time. Independent solutions to spatial and temporal problems have been advanced but none of the proposals has managed to analyze jointly both types of information.

In the next section I review spatial and temporal solutions to the scan-path analysis, proposing methods that try to account for both.

4. Spatial data: heatmap

An interesting solution proposed to visualize an average spatial distribution of fixations is the statistical tool of 'heatmap' [7], [11].

The heatmap is a function to represent in a two-dimnesional colored map the values of a given dependent variable (fixation frequencies in our example). Gradations of colors (e.g from blue to red) are used to represent the underlying distribution of the values of variables (from low to high values).

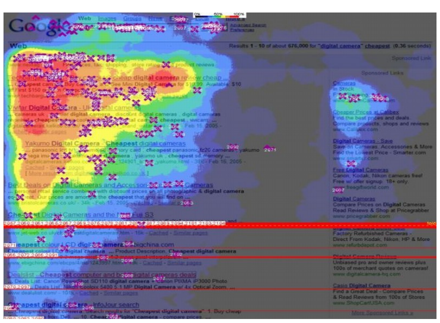

Figure 1. Example of a heatmap applied to a Google page (extract from www.fruition.net).

In Figure 1, we show an example of a heatmap applied to a Google results page. The bright regions of the page are those mostly fixated by participants whereas cold regions are those fixated less. The crosses (X) represents the spatial coordinates of fixations. It appears that most of the fixations fell within the first three links. An obvious conclusion is that participants do not spend much time inspecting all the links on the page but trust directly the ranking provided by Google. However, even if the heatmap provides a powerful visualization of an averaged spatial 'scan-path', it entirely loses any information regarding with time. Moreover, by applying the heatmap indiscriminately to the whole page, we reduce the granularity of fixation analysis for specific image regions. For a deeper understanding of attentional resource deployment, we might want to have a more precise focus on specific parts of image. We propose a way to use heatmap that takes into account the idea of image regions while maintaining the temporal information.

We explain our method through eye-movement data gathered from a psycholinguistic visual world study [2]. In the study, the main interest was to see how certain visual and linguistic factors can influence patterns of eye-movements for different visual objects given syntactically ambiguous linguistic stimuli.

Figure 2 shows the results yielded applying the heatmap function on a set of visual ROI for two different experimental conditions.

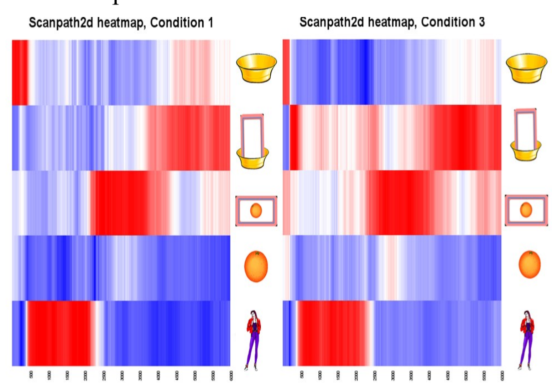

Figure 2. Heatmap of fixation frequencies on visual ROI over time. Comparing two different psycholinguistic experimental conditions

We build our heatmap using as variables: 1) the visual ROI and 2) the time course slices[1]; and as dependent measure the count of fixation frequencies. We sum, across subjects, the fixation frequencies for each time slice (10ms) on each target object per condition. Thus, we obtain for each 10ms bin, a global estimation of how much participants were looking at a certain object. In the heatmap, we display objects on the rows and the time course on the columns. The heatmap assigns for each time slice a color that is chosen by comparing the frequency value of that slice

1 The time requires segmentation in slices to form ROI. Fixations for each ROI are calculated counting the time slices.

to the global frequency distribution. If, for example, on object ``the orange' at time slice 310ms we have a frequency value of 10, and the highest frequency value of the distribution is 90, the heatmap will assign a blue color to it (when blue is used to map low values).

We clearly see that visual objects have different fixations frequencies at different time points. The difference is due to the synchroneous processing of the same visual (ORANGE) and linguistic ('*the orange*') referential entity (see [2] for more details).

Moreover, comparing the two conditions, we see different patterns of fixations to visual objects. Through this use of the heatmap we can better approximate to an ideal scan-path without being forced to collapse either spatial or temporal information. Moreover, we can pin down exactly the relative importance of different visual region of an image over time. In studies of web usability, this method can help to get a clear picture on how different parts of the information are accessed which is extremely useful to restructure the interaction of users toward the expected direction. However, a disadvantage of this method is the fragmentation of the whole image into several subregions that might discard important phenomena happening at different locations. Another a weak point of heatmaps is its failure to provide inferential analysis. We can visualize which object has been fixated and at which moment in time, but we can't have an exact estimation of how many subjects where synchronously fixating at the same object during the same time slice.

It may be possible to obtain a measure of synchronous fixation through cross-recurrence analysis. The idea is to use the cross recurrence measure resulted on each time slice as input to the heatmap.

4. Temporal alignment: Cross-recurrence analysis

The **C**ross-**R**ecurrence **A**nalysis (CRA) is an interesting technique used to compare different temporal sequences for the same type of data in order to discover intrinsic patterns and regularities. It has mostly been developed to study physical phenomena such as climatological or geological events [5] that show recursive patterns. The backbone of CRA is the recurrence analysis that seeks regular patterns by comparing a single time series at different temporal scales. The basic idea of RA is to generate from the same time series, different time series that are temporally shifted trying to find the alignment across series that minimizes the distance between points (for a more detailed explanation refer to [12]. CRA is an extended version of RA that simultaneously compares two different time series.

Recently CRA has found application also in psycholinguistic, syntactic similarities in dialogue [3] and eye-tracking research, gaze coordination of different viewers [9].

Our idea is to use CRA on eye-movement sequences from different participants to extract a global recurrence measure of fixations for each visual ROI.

The first attempt compares fixations from two different participants for the same trial.

In Figure 3 we see on the x and y axis the two different time series. The degree of convergence between the two time series is generated along the diagonal of the CRA matrix.

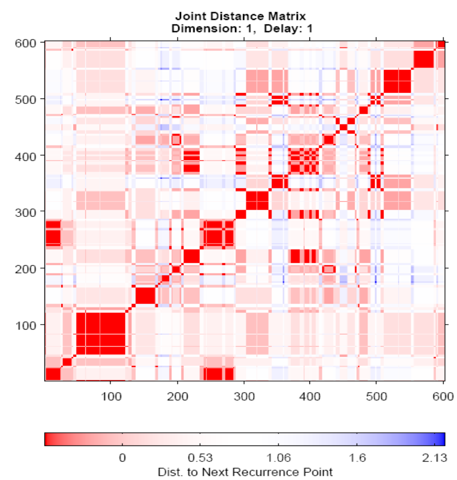

Figure 3. Joint recurrent plot of fixation frequencies of two different participants for the same trial.

In the plot, this measure is visualized through the size of squares. The bigger the square is, the more agreement there is between the two time series. We can see, for example, that during the first 1000ms fixations of the two participants were similarly aligned on the same visual objects. Besides visualizing, it is also possible to quantify the percentage of recurrence to obtain a deeper understanding of the agreement between participants [12].

However, this analysis applied to eye-tracking data presents several problems that seriously limit its applicability.

The first problem is the constraint on the number of time series that can be simultaneously compared. CRA can account for a maximum of two time series. Usually, eye-tracking experiments involve many participants, thus we need to compare multiple eye-tracking series at the same time. The second problem is the granularity of the measure of recurrence. In our case, we have calculated an overall measure of recurrence of the two eye-movement series by calculating the distance of fixations between different visual ROIs. Thus, we know when subjects have agreed by looking at the same object, but we don't know on which specific object they agreed.

The third problem is the number of different trials. We did our analysis looking only at a single trial, but in a standard eye-tracking experiment, there are usually more than a single trial.

At a first glance, it seems that CRA is difficult to apply to eye-tracking research but further research on methods of averaging across visual ROIs or trials are needed.

5. Conclusion

In the present study we have highlighted the need for sophisticated monitoring techniques to understand and improve human-system interaction. We have reviewed eye-movement literature on web-design, focusing on statistical methods of analysis. The concept of scan-path and its spatio-temporal intractability has been discussed, stressing the necessity of finding a statistical technique that can reliably preserve both spatial and temporal information. In this context we have reviewed the visualization method of heatmaps, presenting a way to subdivide the analysis into visual ROIs while preserving the time information. Finally, we have applied a statistical technique, CRA, to measure similarities between two different time-series highlighting advantages and disadvantages for the analysis of eye-movements data.

References

[1] Aula, A., Majaranta, P., & Raiha, K. Eye tracking reveals the personal styles for search result evaluation. *Proceedings INTERACT*, 2005.

[2] Coco, M. I. & Keller, F. Competition between visual and linguistic resources. *Presented as poster at AMLAP,14*, Cambridge, UK., 2008.

[3] Dale, R. & Spivey, M. Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and care givers in conversation. *Language Learning*, 56, 391–430, 2006.

[4] Habuchi, Y., Kitajima, M., & Takeuchi, H. Comparison of eye-movements in searching for easy-to-find and hard-to-find information in a hierarchically organized information structure. *Proceedings of the 2008 symposium on Eye tracking research and application*, 131–134, 2008.

[5] Marwan, N. & Kurths, J. Nonlinear analysis of bivariate data with cross recurrence plots. *Physics Letters A*, 302, 299–307, 2002.

[6] Privitera, C. & Stark, L.. Algorithms for defining visual region-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 970–982, 2000.

[7] Raiha, K., Aula, A., Paiva, M., Rantala, H., & Koivunen, K. Static visualization of temporal eye-tracking data. *Lecture Notes in Computer Science,* 3585/2005, 946–949, 2005.

[8] Rayner,K., Li,X., Williams,C. Cave, K. & Well. A. Eye movements during information processing tasks: Individual differences and cultural effects. *Vision research*, 47, 2714-2726, 2007.

[9] Richardson, D., Dale, R. & Kirkham, N.. The art of conversation is coordination: common ground and coupling of eye-movements during dialogue *Psychological science*, 18, 407-413, 2007.

[10] Santella,A. & DeCarlo, D. Robust clustering of eye movement recordings for quantification of visual interest. *Eye Tracking Research and application (ETRA) Symposium*, 2004.

[11] Spakov, O. & Mniotas, D. Visualization of eye gaze data using heat maps. *Electronics and electrical engineering*, 2, 55-58, 2007

[12] Zbilut,J. & Webber, C.. Recurrence quantification analysis. *In Wiley Encyclopedia of biomedical engineering*, 2006.