Eye-Tracking Data in Visual Search Tasks: A Hallmark of Cognitive Function

Vicente Pallarés*, Mar Hernández†, and Laura Dempere-Marco‡

Abstract

Vision is an active process in that detailed representations of our visual world are only built from actively scanning our eyes with a series of saccades and fixations. The process of actively scanning a visual scene while looking for something in a cluttered environment is known as visual search. The study of visual search processes not only offers a unique opportunity to gain fundamental insights into visual information processing in the human brain, but also opens new avenues to assess cognitive function. In this work, we show that it is possible to unveil the strategies pursued by subjects to solve visual tasks by investigating dynamical aspects inherent to eye-tracking data in a generalized N-dimensional feature domain

1 INTRODUCTION

The visual search paradigm has been key in attention research for investigating visual attention deployment, both overt and covert (e.g., [1]). Interestingly, visual search processes encompass different cognitive functions, which include, among others, working memory (responsible for the active maintenance of information during short periods of time), long-term memory (memory traces tend to facilitate viewing of similar scenes), and decision-making (involved in categorization tasks and eye-movement control). Thus, the visual search paradigm can be used to evaluate the cognitive state of a person, which is of outmost importance in neurorehabilitation as can be used to assess and monitor the evolution of patients.

With the advent of eye-tracking technologies, accurate and (importantly) non-intrusive measurements of eye movements have become available in the last few decades. The availability of eye

*V. P. is with the Department of Information and Communication Technologies of Pompeu Fabra University, Barcelona, Spain

†M. H is with the Polytechnic School of Pompeu Fabra University, Barcelona, Spain

‡L. D. is with the University of Vic-Central University of Catalonia, Vic, Spain (corresponding author) laura.dempere@uvic.cat

movement recordings has undoubtedly boosted the research in visual search by making it possible to have ground truth data to constrain and validate the visual search models while providing vast amounts of data, which in turn require from advanced techniques for their analysis [2].

In this work, we present a method for knowledge gathering from visual search. The proposed method is evaluated by means of a simple visual search task. By making use of a generalized feature domain, the dynamics of the eye movements patterns are analyzed in order to assess executive function. In particular, both the capability to adaptively create strategies and their dynamics are studied.

2 METHODS

2.1 Experimental set-up: double conjunction search

To assess the validity of the proposed method, we have considered a modified version of the double conjunction search task presented in Hu et al. [3]. The difference lies in the systematic manipulation of the image content with the following objectives: 1) to assess the emergence of optimal feature-based strategies during visual search, and 2) to assess the effect of target switching within a sequence. Overall, 30 healthy volunteers have participated in this study. Their eye movements have been recorded with an eye-tracker (Tobii X120, fs = 120 Hz) while performing the task. The subjects were seated 65 cm in front of a 17-inch monitor (screen resolution 1024 $\times$ 768 pixels) and a chin-rest was used to prevent head movements.

Figure 1 illustrates several test cards, together with one of the experimental protocols, used in this study. The goal of the task is to find the unique object (i.e. a square) that presents two specific features among a set of distractors (i.e. the rest of squares).

2.2 Hot-Spot Framework

In this work, we consider the Hot-Spot framework developed in Hu et al. [3] and the extension presented in Pallarés et al. [4]. The fixation events from the scan-path are mapped both onto the color feature domain and a spatial domain defined by saccade directions. Given a scan-path $\xi = \langle x_i, y_i, t_i \rangle$, where each fixation is centered at $\langle x_i, y_i \rangle$ and has an associated dwell time t_i, it is possible to extract the prevalence of a particular feature f_k as

$$T(f_k) = \sum_{\langle x_i, y_i \rangle \in \xi} \left(\sum_{\langle x_i, y_i \rangle \in \Omega(x_i, y_i)} t_i {P^i}_{f_k}(x, y) \right) \quad (1)$$

where $\Omega(x_i, y_i)$ is the foveal field of the fixation centered at $\langle x_i, y_i \rangle$ considering all the pixels within a 2° visual angle, and $P^i_{f_k}$ is a probability distribution function. In this framework, the relevance $\Gamma(f_k)$ of feature f_k is assessed by normalizing the total dwell time (T) by a factor that represents the absence of any predetermined strategy.

2.3 Feature transients and steady states

Eye movements dynamics can be investigated from the perspective that dynamical systems offer, when the task at hand involves close scrutiny of both target and distractors. Such an approach was already pursued in [5] by means of the so-called Transient Fixation Moments. In this work, we present an alternative method. In particular, when the system settles in a particular state in which certain properties remain unchanged in time, a so-called *steady state* is reached. In many systems, it is common that the steady state is only achieved after some time. During this period, while the system explores different configurations, it is said to be in a *transient state*. In our case, reaching the steady state would mean that the subject has developed a particular strategy and remains in it until the target is found. As previously discussed, the emergence of a feature-based strategy is characterized through the function $\Gamma(f_k)$. Thus, it is possible to assess the time needed

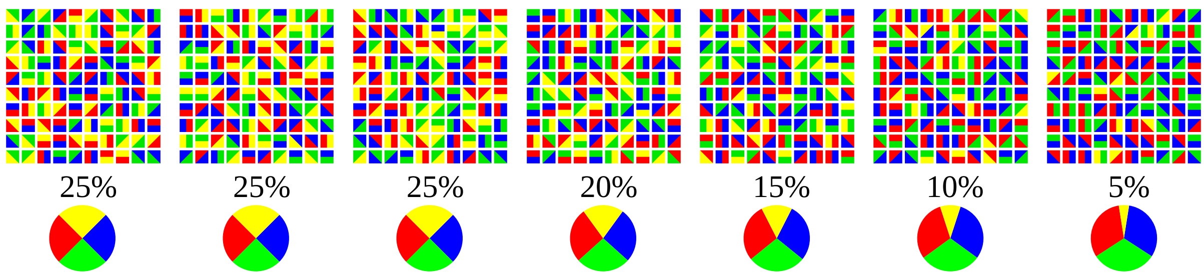

Figure 1: The test cards are composed of squares distributed in a 10×10 grid with a gap between adjacent squares. Each of the squares in the stimulus card is defined by two out of four possible colors. A sequence of different cards is presented to the participants following a particular protocol. Overall, 3 different protocols have been considered: 1) the proportion of one of the colors dimishes gradually while preserving the target object, 2) different color prevalences appear in a random order while preserving the target object, 3) switch of the target object in one of the cards. In this example, three out of the seven cards have an equal global amount of color each (i.e. 25% prevalence) whereas in the other five cards, the amount of yellow diminishes gradually, from 25% to 5% prevalence in 5% steps. Such decrease intends to elicit the emergence of color-based search strategies as a result of its increasing visual saliency.

to develop a given strategy by considering the time needed to reach the steady-state:

$$\frac{\partial \Gamma(f_k)}{\partial t} = 0 \tag{2}$$

3 RESULTS

The analysis of the eye-tracking data in the two domains (i.e. color and spatial feature domains) has allowed us to identify the emergence of several stereotypical visual search strategies. On one hand, we have found spatial-based strategies in which the subjects follow specific spatio-temporal patterns of saccades. Figure 2(A) illustrates a trial in which the subject mainly relies on horizontal movements to systematically scan the image until the target is found. On the other hand, we have also found feature-based strategies in which the subjects specifically relies on a particular feature to find the target. An example of such kind of strategies is shown in Figure 2(B). Note that, in this example, the relevance of the yellow color is prominent and remains largely constant until the end of the search. As in the feature-based strategy, the relevance of the horizontal saccades is substantially higher than that of the remaining spatial-temporal features.

Both spatial-temporal and feature-based patterns are studied by considering a common methodological framework. The method has been evaluated on 30 subjects and has allowed us to: 1) characterize the strategies that are employed, and 2) study the emergence, temporal deployment and dominance of such strategies.

4 CONCLUSION

In this work, we provide evidence supporting the notion that it is possible to unveil, and mathematically characterize, the strategies pursued by subjects while solving complex visual tasks. We suggest that the use of eye-tracking technology may provide important insights into cognitive assessment and monitoring. With the availability of detailed spatio-temporal visual scan-patterns obtained from eye-tracking experiments, further novel and accurate tests that serve as behavioral surrogates of cognitive function can be devoloped. It is worth noting that traditional neuropsychological tests (e.g. Trail Making Test A and B) are routinely used in clinical practice to assess cognitive performance. However, in contrast to the type of measures that are commonly obtained in such tests (e.g., overall completion time of the task, regarded as reaction time) eye-tracking recordings offer unrivaled spatiotemporal resolved data, which makes it possible to untangle the

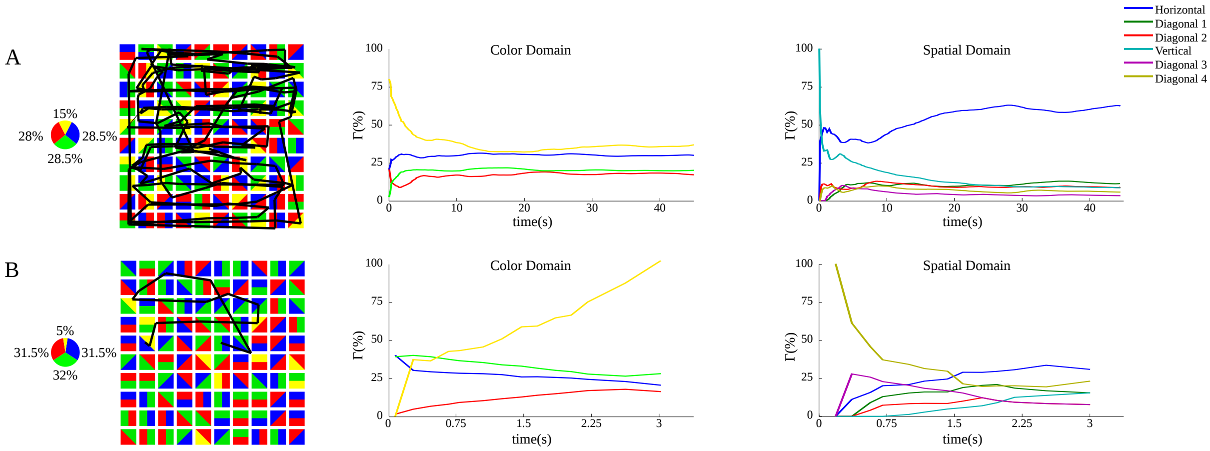

Figure 2: (A) Illustration of a scan-path in which a spatio-temporal based strategy emerges. As can be seen in the right panel, the relative attention received by the horizontal saccades is superior to that of the rest of the directions, despite the likelihood of making a saccade in any given direction is the same for all of the directions. (B) Illustration of a scan-path in which a color-based strategy emerges early during the trial. As can be seen in the middle panel, the relative attention received by the yellow color is superior to that of the rest of the colors, despite its low prevalence in the card. In contrast, no spatio-temporal pattern has been observed.

reasons why, for instance, such reaction times are obtained. The proposed methodology initiates a series of necessary steps towards establishing ”cognitive signatures”, and identifying very subtle changes in such signatures. This is particularly useful in the area of neurorehabilitation since it can be used as a diagnostic tool to assess the cognitive state of patients and to monitor the progression of the therapeutical interventions.

ACKNOWLEDGMENTS

The authors acknowledge funding from the Spanish Ministry of Economy and Competitiviness through the project TIN2013-40630-R.

References

[1] E. T. Rolls, and G. Deco, *Computational Neuroscience of Vision*. Oxford University Press: Oxford, 2002.

[2] L. Dempere-Marco, X.P. Hu, G.Z. Yang (2011), A novel framework for the analysis of eye movements during visual search for knowledge gathering. Cognitive Computation 3 (1), 206-222

[3] X.P. Hu, L. Dempere-Marco, and G.Z. Yang (2003). Hot spot detection based on feature space representation of visual search. IEEE Transactions on Medical Imaging, 22(9):1152-1162.

[4] V. Pallares, L. Rami, L. Dempere-Marco (2016), A Novel Method for Disentangling Strategies from Visual Search, BIOSIGNALS 2016: 9th International Joint Conference on Biomedical Engineering Systems and Technologies, Rome, Italy

[5] L. Dempere-Marco, X.P. Hu, G.Z. Yang ,Visual scan-path analysis with feature space transient fixation moments Proc. SPIE 5034, Medical Imaging 2003: Image Perception, Observer Performance, and Technology Assessment, p. 503-512