A Novel Method for Disentangling Strategies from Visual Search

Vicente Pallarés[1], Lorena Rami[2] and Laura Dempere-Marco[1,3]
[1]*Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain*
[2]*Alzheimer's Disease and Other Cognitive Disorders Unit, IDIBAPS, Hospital Clinic, Barcelona, Spain*
[3]*Faculty of Sciences and Technology, Universitat de Vic-Universitat Central de Catalunya, Vic, Spain*

Keywords: Eye-Tracking, Feature Domain Analysis, Visual Search, Aging.

Abstract: The process of actively scanning a visual scene while looking for something in a cluttered environment is known as visual search. In this work, we show that it is possible to disentangle the strategies pursued by subjects to solve visual tasks by investigating dynamical aspects inherent to eye-tracking data. A novel method is proposed to characterize visual search strategies in a generalized N-dimensional feature domain, which allows us to investigate spatial-temporal aspects of the search as well as the subjects' reliance on visual cues. In order to validate the proposed method, we have developed an experimental paradigm based on a double conjunction search in which one the visual cues is systematically manipulated, which can induce feature-based strategies in the observers. On the basis of the preliminary evidence presented here, we argue that this characterization of visual search strategies opens new avenues to assess cognitive function and its relation to normal aging.

1 INTRODUCTION

It is well-established that we do not systematically scan the visual world with our eyes when looking for something in a cluttered environment. Instead, we allocate the gaze selectively so as to maximize the information that is relevant for solving the task at hand. Thus, much information can be retrieved from the way the brain processes visual information by studying the eye movements during the exploration of a particular scene. For this reason, visual exploration mechanisms have been investigated for many years, and from many different perspectives, e.g. psyphophysics (Treisman and Gelade, 1980; Wolfe, 1998), computational neuroscience (Deco and Rolls, 2004; Deco and Zihl, 2006), or statistical physics (Boccignone and Ferraro, 2004). One important technique that tries to exploit these mechanisms and has become more used during the last decades is eye-tracking. Several cognitive processes have been, and are being, studied by means of eye movements, like scene perception (Võ and Henderson, 2010), visual search (see (Yang et al., 2002; Dempere-Marco et al., 2011) for a review) or reading (Rayner, 2009).

Arguably, visual attention plays a critical role in strategy formation and maintenance. Interestingly, a close relation between eye movements and visual attention has been broadly accepted. In particular, two main mechanisms to reallocate attention during visual search have been identified. The first one, called overt attention, relies on shifting the gaze toward a new location, whereas the second one, known as covert attention, consists of paying attention to an area in the periphery of the foveal region but without redirecting the gaze. Evidence of the existence of both types of attention has been proven by means of psychophysics, electrophysiology and neuroimaging (Carrasco, 2011). The interaction between saccade programming and covert attention has also been confirmed (Deubel and Schneider, 1996; Hoffman and Subramaniam, 1995), even unveiling the existence of common neural mechanisms (Corbetta et al., 1998; Beauchamp et al., 2001; Grosbras et al., 2005). A similar link between fixational eye movements (especially microsaccades) and covert attention has been largely suggested (Engbert and Kliegl, 2003; Laubrock et al., 2010; Yuval-Greenberg et al., 2014). Of note, the visual search paradigm has been key in attention research for investigating visual attention deployment, both overt and covert (e.g. Deco and Rolls (2004)).

Hence, we also make use of the visual search paradigm to obtain insights into the observers' reasoning processes while solving a visual task. We

Pallarés, V., Rami, L. and Dempere-Marco, L.
A Novel Method for Disentangling Strategies from Visual Search.
DOI: 10.5220/0005821602710276
In *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2016) - Volume 4: BIOSIGNALS*, pages 271-276
ISBN: 978-989-758-170-0

argue that this approach opens up the possibility to study the eye movements as a hallmark of cognitive function. To this end, a new approach for knowledge gathering from visual search is presented and evaluated here, as well as applied to a simple visual search task. The novelty of the proposed work lies on the definition of a generalized feature domain in which the dynamics of the eye movements patterns are analyzed. This analysis includes both visual cues (features) of the stimulus and spatial characteristics of the patterns, which enrich the characterization of the eye movements providing a new generalized feature domain in addition to the more common event-related spatial domain (Holmqvist et al., 2011).

2 METHODS

2.1 Experimental Set-up

To assess the validity of the proposed method, we have considered a double conjunction search task akin to that presented by Hu et al. (2003). We have, however, modified such task in this study by systematically manipulating the image content in order to facilitate the emergence of optimal feature-based strategies during visual search. Figure 1 illustrates several test cards, together with the experimental protocol, used in this study. The goal of the task is to find the unique object (i.e. a square) that presents two specific features (i.e. the colors blue and yellow) among a set of distractors (i.e. the rest of squares).

Two groups of subjects have been considered in this study: 7 young participants (2 male/ 5 female, ages [22-38]), and 7 elderly subjects (2 male/ 5 female, ages [58-68]). Their eye movements have been monitored with an eye-tracker (Tobii X120, $f_s = 120$ Hz) while performing the task. The subjects were seated 65 cm in front of a 17-inch monitor (screen resolution 1024×768 pixels) and a chin-rest was used to prevent head movements.

2.2 Extending the Hot Spot Framework to a Generalized Feature Domain

A conceptual framework was presented in (Hu et al., 2003) based on the hypothesis that there is a direct relation between visual attention and oculo-motor action. In this approach, which received the name of hot-spot framework, the fixation events from the scan-path (originally defined in the spatial domain) are projected onto a new feature domain (in this particular case, the color domain). The prevalence of such features in the visual scan-path —defined as a sequence of fixations and saccades— are determined, which can unveil the existence of the underlying strategies pursued by the observer.

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 with a dwell time t_i, it is possible to extract the prevalence of a particular feature f_k as

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

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.

Those features that are relevant, not only from a bottom-up perspective but also in conjunction with top-down aspects, can be extracted by defining, in the feature domain F, a density function Γ. To do this, the prevalence of a feature point f_k must be normalized by a factor that represents the absence of any predetermined strategy. This normalization factor is computed by considering the exploration of the whole card with no over/under scanning of any part, and it is calculated as

$$T_0(f_k) = \sum_{\langle x_i,y_i\rangle\in Image}\left(\sum_{\langle x_i,y_i\rangle\in\Omega(x_i,y_i)} P^i{}_{f_k}(x,y)\right) \qquad (2)$$

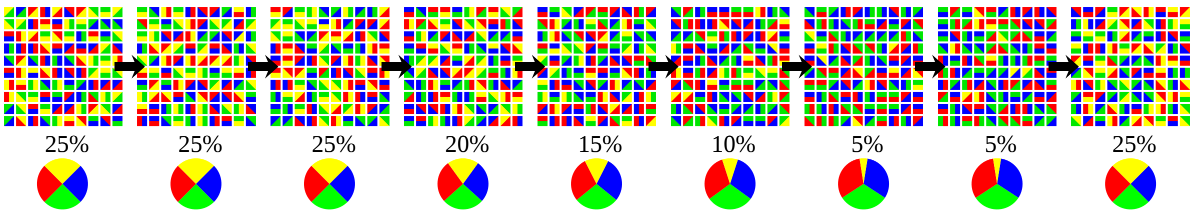

Figure 1: The test cards are composed of squares distributed in a 10×10 grid with a gap between adjacent squares. Each of the squares in the stimulus card is defined by two out of four possible colors: red, green, blue and yellow. A sequence of nine different cards is presented to the participants following the above protocol, and each card is displayed until the target is found. As illustrated in the example, four out of the nine 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.

T_0 describes, then, the prevalence of a feature in a scan-path which shows a systematic exploration of the test card with no preference for any features of the stimulus. The density function is finally defined as

$$\Gamma(f_k) = \frac{\dfrac{T(f_k)}{\sum_k T(f_k)}}{\dfrac{T_0(f_k)}{\sum_k T_0(f_k)}} = \frac{T'(f_k)}{T'_0(f_k)} \quad (3)$$

$T'(f_k)$ characterizes the prevalence of the feature f_k in the observers' scan-path, as Equation (1), but now considering the specific properties of the stimulus.

All in all, by turning $T(f_k)$ and $T_0(f_k)$ into probability density functions, Γ can be explained in terms of signal detection theory (Hu et al., 2003). In particular, $T'(f_k)$ can be interpreted as the detected signal, which also contains a noise signal $T'_0(f_k)$. The likelihood function in our case corresponds to the ratio between $T'(f_k)$ and $T'_0(f_k)$. Thus, $\Gamma > 1$ would be indicative of certain feature preference in that the likelihood of paying attention to a particular feature would be higher than that expected by chance. In this work, we extend the original formulation in order to capture strategies developed on any feature domain, which also includes spatial patterns present in the visual scan-paths. Moreover, our approach can be applied to any eye movement event, not only fixations.

For this particular task, two specific feature domains have been analyzed. On one hand, the prevalence of color-based patterns is assessed in the color domain, as in (Hu et al., 2003). This is particularly relevant in our study given the decrease in the amount of one of the two target colors, which entails the emergence of an optimal strategy based this feature. On the other hand, and in contrast to the original work, spatial-temporal patterns have also been extracted from the scan-path and are projected onto a spatial feature domain. This has allowed us to reveal and characterize the emergence of certain stereotypical behaviors such as reading patterns or column-wise exploration.

2.2.1 Color Domain

The emergence of strategies in the feature domain is assessed from Equation (1) by considering color as the visual cue which defines such domain. In this study, the probability function $P^i_{f_k}(x,y)$ is defined such that for each fixation $\langle x_i, y_i \rangle$, $P^i_{f_k}(x,y) = 1$ $\forall (x,y) \in \Omega(x_i, y_i)$, and else $P^i_{f_k}(x,y) = 0$. As previously stated, this implies that the relevance of each color is proportional to the attention received (see Equation (1)), while it also depends on the characteristics of the scene (see Equation (2)). Therefore, the density function of each color is defined as in Equation (3). For these particular stimuli $1 \leq k \leq 4$, representing the four colors (i.e. red, green, blue and yellow).

2.2.2 Spatial Domain

Similarly, the emergence of spatial-temporal patterns is explored in this study. It is worth noting that saccade direction conveys information about the geometrical structure of the scan-paths, which in turn, may reflect the existence of underlying systematic behaviors. Moreover, the analysis based on the consideration of spatial positions (i.e. the $\langle x_i, y_i \rangle$ fixation coordinates) is not traslation invariant. By using saccade direction instead, the outcome is less sensitive to translation and thereby less dependent on the local characteristics of the stimulus. Thus, in order to explore this scenario, the angles of the saccades present in the scan-path are analyzed, which give rise to a newly defined feature domain of saccade directions.

To project the saccade events of the scan-path onto this feature domain, Equation (1) and Equation (2) are recalled. The saccade direction is calculated with respect to the horizontal. The angular range of 360° is discretized into 12 intervals spanning 30° and centered at 0°, 30°, 60°, 90°, ..., 330°. Each pair of opposed angles is considered as the same spatial direction, and defines one out of 6 possible directions of the eye movements. Since the saccade time length is substantially short and remains largely constant, t_i is taken as a constant. The prevalence of each saccade direction in the scan-path is then described by

$$T(f_k) = \sum_{\langle sacc_i \rangle \in \xi} P^i_{f_k} \quad (4)$$

where $sacc_i$ are the saccades extracted from the scan-path ξ, and $P^i_{f_k}$ is a probability distribution function, which is defined such that $P^i_{f_k} = \delta(d - f^i_k)$ (i.e. $P^i_{f_k} = 1$ if the direction of saccade i is k, and else $P^i_{f_k} = 0$). This prevalence must be normalized again by a factor that takes into account all the possible directions for the saccades in the scan-path. This operation will weight the saccade direction f_k according to its dominance when considering a strategy-free scan-path. Then, Equation (2) can be rewritten following an analogous reasoning to that described in previous sections as

$$T_0(f_k) = \sum_{\langle sacc_i \rangle \in Image} P^i_{f_k} \quad (5)$$

In this occasion, the lack of strategy is modeled as a homogenous probability of saccadic programming in

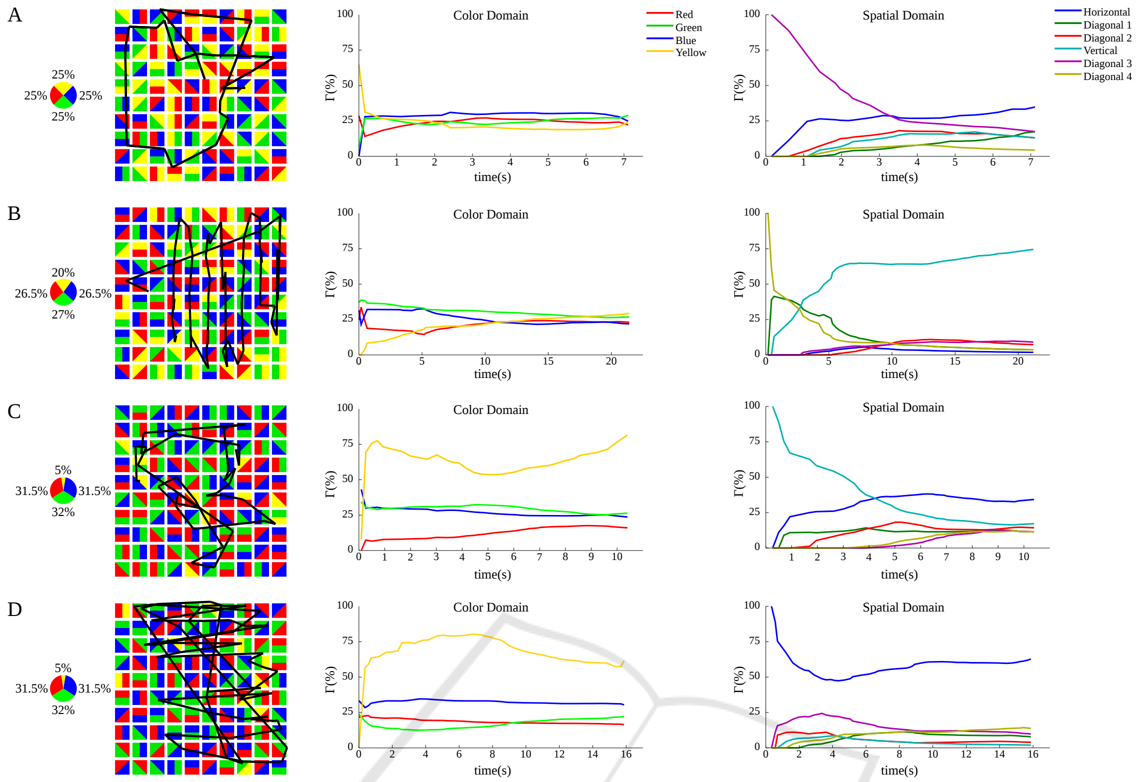

Figure 2: Illustration of four scan-paths that have been analyzed by projecting the information from fixations and saccades in the color and spatial domains, respectively. [Left] stimulus cards with the saccades trajectories superimposed; [center] the density function corresponding to each color in the feature domain; [right] the density function corresponding to each saccade direction for the 6 possible directions. A density value over 50% is accepted as an indicator of pursuing a strategy.

any of the 6 possible directions. By again turning, $T(f_k)$ and $T_0(f_k)$ into probability density functions a probability density function Γ akin to that in Equation (3) has been defined.

3 RESULTS

The analysis of the eye-tracking data in both domains (i.e. color and spatial feature domains) has allowed us to identify a number of stereotypical strategies, which are shown in Figure 2. Interestingly, the proposed method permits us to investigate both dynamical aspects of the visual search strategies as well as the aggregate behavior which accounts for the complete scan-path. In the first case, the density function $\Gamma(t)$ is evaluated by considering all of the fixations from the scan-path occurred up to time t. Thus, it is the dynamical evolution of the function $\Gamma(t)$ what is depicted in Figure 2. In order to normalize the curves into a common scale, the density functions have been divided by the number of features n, which conform the feature domain at hand, i.e. $n = 4$ for color and $n = 6$ for saccade directions.

Figure 2 shows several paradigmatic examples which correspond to individual visual scan-paths over four different test cards. The first aspect that must be noted is the existence of a transient period, which is usually followed by a stationary state. It is during this stationary state that the strategy becomes stable and can be readily identified. Out of the four cases illustrated, Figure 2.A (a card presenting a 25% of each color), does not reveal the presence of any strategy, i.e. neither color-based nor any spatial-temporal pattern. In fact, after a short transient period, the density function for all four colors converges to a 25%, thus revealing the absence of any feature preference. Similarly, in the spatial domain no feature preference has been identified, which is in agreement with the visual scan-path (see left panel) in which no clear strategy can be identified.

In contrast, Figure 2.B shows another trial (20% of yellow) in which a spatial-temporal pattern emerges from the visual scan-path. In this case, a column-wise reading pattern can be identified (see left panel). The proposed method clearly reveals such a strategy (note the overall dominance of the density function corresponding to the vertical saccade direction). Notably,

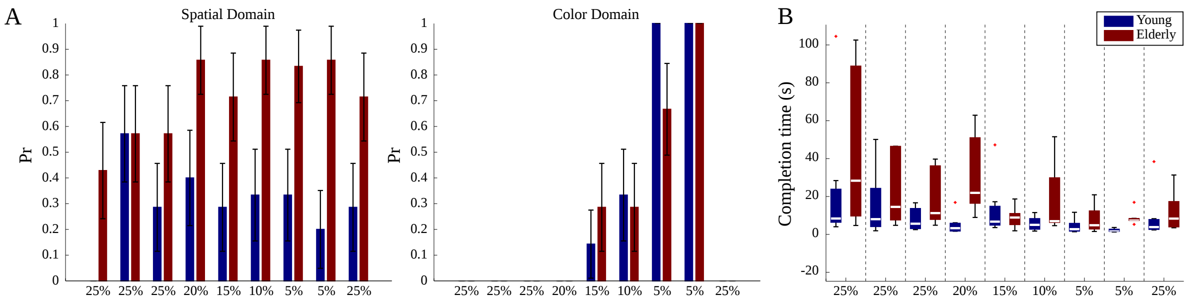

Figure 3: (A) Probability estimates of developing spatial-based and color-based strategies; the emergence of a strategy is considered whenever a 50% density value is exceeded for young (blue) and elderly (red) populations. (B) Task completion times for young (blue) and elderly (red) subjects.

this method does not only allow to identify whether a particular strategy occurs but also *when* it emerges and *how* (or whether) it is related to other strategies. Similarly, the trial in Figure 2.C, which presents a different proportion of colors, shows a color-based strategy. In this case, the density function corresponding to yellow grows rapidly revealing a preference for this color whereas no spatial-temporal pattern can be identified. Finally, Figure 2.D shows another trial (also 5% of yellow), in which a combination of color-based and spatial-based strategies emerges. The analysis in the color domain reveals a clear prevalence of yellow in the visual scan-path while, at the same time, a row-wise scanning pattern (i.e. dominance of horizontal direction saccades) becomes apparent in the spatial domain. The observer covered a large area of the card pursuing this spatial-temporal pattern, but with a tendency to focus on the most relevant feature given the task at hand. This is an example of an optimal strategy in that the most informative areas are fixated upon while following a systematic strategy which allows the subject to easily keep track of the items which have already been visited.

Finally, the eye-tracking data from both populations is considered to assess possible differences because of aging. For each trial, two coefficients have been computed, one describing spatial-based strategies (i.e. horizontal or vertical saccade direction dominance) and the other color-based strategies (i.e. yellow dominance). To this end, the stationary state and its accompanying strategy in each of these domains is characterized by the averaged density function value during the last 500 ms of the trial. In the case of the spatial domain, the maximum between the horizontal and vertical saccade density functions is considered. Whenever $\Gamma > 50\%$, we consider that a strategy has been actively pursued. Each trial of the experiment has been considered as a trial of a Bernoulli process interpreted as a success whenever $\Gamma > 50\%$. This has allowed us to evaluate for a particular strategy the probability of being pursued by a population.

The results obtained (Figure 3.A) show some evidence of notable differences among the two populations. In particular, the probability of developing a spatial strategy is higher for the elderly population for almost any kind of card, while for young participants spatial-based strategies tend to decrease as color-based strategies emerge. In short, the elderly apparently undergo more systematic searches. Beyond visuo-motor aspects, and the large variability observed in the data, such systematic patterns tend to be accompanied by longer overall search times (as can be seen in Figure 3.B).

4 CONCLUSIONS

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. A novel experimental paradigm based on a double conjunction search, in which one of the visual cues is systematically manipulated, has been proposed. Both spatial-temporal and feature-based patterns are studied by considering a common methodological framework. The method has been evaluated on two groups of subjects, who differ in their age. The proposed method has allowed us to: 1) characterize the strategies that are employed, 2) study the emergence, temporal deployment and dominance of such strategies, and 3) assess differences between the two cohorts.

We suggest that the use of eye-tracking technology may provide important insights into aging research. There is evidence that a decline in various aspects of cognition accompanies the aging process. For instance, older adults show a loss of processing speed (Salthouse, 1996) and a decline in several aspects of executive function (working memory, task switching, inhibitory function), and reasoning, as reviewed in (Sperling et al., 2011). We, thus, hypothesize, and provide preliminary evidence, that the visual search paradigm can be used to probe such basic cognitive functions and their relation to normal aging and age-related neurodegenerative pathologies.

ACKNOWLEDGEMENTS

The authors acknowledge funding from the research project TIN2013-40630-R (Spanish Ministry of Economy and Competitiveness)

REFERENCES

Beauchamp, M., Petit, L., Ellmore, T., Ingeholm, J., and Haxby, J. (2001). A parametric fmri study of overt and covert shifts of visuospatial attention. *NeuroImage*, 14(2):310–321.

Boccignone, G. and Ferraro, M. (2004). Modelling gaze shift as a constrained random walk. *Physica A: Statistical Mechanics and its Applications*, 331(1-2):207–218.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51(13):1484–1525.

Corbetta, M., Akbudak, E., Conturo, T., Snyder, A., Ollinger, J., Drury, H., Linenweber, M., Petersen, S., Raichle, M., Van Essen, D., and Shulman, G. (1998). A common network of functional areas for attention and eye movements. *Neuron*, 21(4):761–773.

Deco, G. and Rolls, E. (2004). A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research*, 44(6):621–642.

Deco, G. and Zihl, J. (2006). The neurodynamics of visual search. *Visual Cognition*, 14(4-8):1006–1024.

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

Deubel, H. and Schneider, W. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, 36(12):1827–1837.

Engbert, R. and Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, 43(9):1035–1045.

Grosbras, M.-H., Laird, A., and Paus, T. (2005). Cortical regions involved in eye movements, shifts of attention, and gaze perception. *Human Brain Mapping*, 25(1):140–154.

Hoffman, J. and Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, 57(6):787–795.

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*. OUP Oxford.

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

Laubrock, J., Kliegl, R., Rolfs, M., and Engbert, R. (2010). When do microsaccades follow spatial attention? *Attention, Perception, and Psychophysics*, 72(3):683–694.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Q J Exp Psychol (Hove)*, 62:1457–506.

Salthouse, T. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103(3):403–428.

Sperling, R., Aisen, P., Beckett, L., Bennett, D., Craft, S., Fagan, A., Iwatsubo, T., Jack Jr., C., Kaye, J., Montine, T., Park, D., Reiman, E., Rowe, C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M., Thies, B., Morrison-Bogorad, M., Wagster, M., and Phelps, C. (2011). Toward defining the preclinical stages of alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's and Dementia*, 7(3):280–292.

Treisman, A. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136.

Võ, M.-H. and Henderson, J. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision*, 10(3):1–13.

Wolfe, J. (1998). Visual search. In Pashler, H., editor, *Attention*, Studies in cognition. University College London Press, London, UK.

Yang, G.-Z., Dempere-Marco, L., Hu, X.-P., and Rowe, A. (2002). Visual search: Psychophysical models and practical applications. *Image and Vision Computing*, 20(4):291–305.

Yuval-Greenberg, S., Merriam, E., and Heeger, D. (2014). Spontaneous microsaccades reflect shifts in covert attention. *Journal of Neuroscience*, 34(41):13693–13700.