Adding Neurally-inspired Mechanisms to the SceneWalk model improves Scan Path Predictions for Natural Images

Lisa Schwetlick (lisa.schwetlick@uni-potsdam.de)
Research Focus Cognitive Science, Experimental and Biological Psychology,
University of Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany

Lars O. M. Rothkegel (lars.rothkegel@uni-potsdam.de)
Research Focus Cognitive Science, Experimental and Biological Psychology,
University of Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany

Ralf Engbert (ralf.engbert@uni-potsdam.de)
Research Focus Cognitive Science, Experimental and Biological Psychology,
University of Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany

Abstract

The selection of fixation locations during natural scene viewing depends in large part on image-dependent and observer-dependent factors. However, eye movement data from different images, viewers, and experimental designs also consistently contain systematic tendencies such as pronounced saccade angle distributions, return saccade statistics, and dependencies of these measures on fixation duration. When modelling complete human scan paths during extended natural image viewing these systematic tendencies are critical. The SceneWalk model (Engbert et al., 2015) incorporates image-dependent information through saliency maps and uses attentional processing and inhibitory tagging mechanisms to dynamically generate scan paths. Currently, scan paths simulated with this approach only partially reproduce observed systematic tendencies. Here we propose adding several neurally-inspired mechanisms to the model to improve performance: pre-saccadic and post-saccadic attentional shifts as well as facilitation of return mechanisms. These mechanisms are well-established both in experiments and neurocognitive theories of vision. We find that this extension improves the model to generate scan paths which are in qualitative agreement with empirical data. As the model is firmly theory-based, all parameters are biologically interpretable and thus permit evaluations of theoretical predictions of behavior. We also discuss a fully Bayesian framework using adaptive Markov Chain Monte Carlo methods.

Keywords: eye movements; scan paths; dynamical modeling; parameter estimation; scene viewing;

Background

The human visual system depends crucially on the ability to move the eyes over a scene. As only a small central region of the visual field, the fovea, receives detailed high resolution input, humans scan their visual environment in a series of fast jumps, *saccades*, and periods of relative motionlessness, *fixations*. The sequence of fixation locations chosen as gaze postions is subject of extensive research, as it permits insight into the theories of visuomotor control and their impact on visual perception.

Eye movements are guided by a variety of different mechanisms. Firstly, the image itself contains regions which are inherently more informative than others. For example, objects attract fixations (Nuthmann & Henderson, 2010). This finding and other observations have inspired a class of models which use saliency maps to predict which regions in the image are particularly interesting and therefore likely to be fixated (e.g. Itti, Koch, & Niebur, 1998; Kümmerer, Wallis, & Bethge, 2016). The second category of mechanisms are observer-dependent top-down effects, stemming from task and motivation as well as individual differences (de Haas, Iakovidis, Schwarzkopf, & Gegenfurtner, 2019). Thirdly, there exists a category of mechanisms which is stable over both observers and images (Tatler, Vincent, et al., 2008). Examples of these are the central fixation bias, the distribution of inter-saccadic angles and dependencies of saccade length and fixation duration.

SceneWalk

The SceneWalk Model (Schütt et al., 2017; Engbert, Trukenbrod, Barthelme, & Wichmann, 2015) uses attentional mechanisms coupled with inhibitory tagging to dynamically generate scan paths. Both streams are motivated by well-documented findings from the field of visual perception research. Attention is guided by image-dependent information and the foveated nature of the input (Schütt et al., 2017; Engbert et al., 2015). Inihibitory tagging of previously fixated regions promotes image exploration (Mirpour, Bolandnazar, & Bisley, 2019).

In the model, the two streams exist as independent 2D activation maps which evolve over time and are later combined to form a target map from which fixations are selected probabilistically. We compute a Gaussian $G_{A/F}$ centered around the current fixation position for each stream ($A =$ attention map, $F =$ fixation map/inhibitory tagging). Both streams are implemented on an $L \times L$ lattice and evolve via coupled differential

equations, i.e.,

$$\frac{dA_{ij}(t)}{dt} = -\omega_A A_{ij}(t) + \omega_A \frac{S_{ij} \cdot G_A(x_i, y_j; x_f, y_f)}{\sum_{kl} S_{kl} \cdot G_A(x_k, y_l; x_f, y_f)} \quad (1)$$

$$\frac{dF_{ij}(t)}{dt} = -\omega_F F_{ij}(t) + \omega_F \frac{G_F(x_i, y_j; x_f, y_f)}{\sum_{kl} G_F(x_k, y_l; x_f, y_f)} . \quad (2)$$

S_{ij} is a saliency map of the image, for which we can use a map generated by another model or, in this case, the empirical fixation density on the image.

The two pathways are each shaped by an exponent λ or γ, respectively. Then we subtract the weighted (C_F) inhibition path from the attention path.

$$u_{ij}(t) = \frac{(A_{ij}(t))^{\lambda}}{\sum_{kl} (A_{kl}(t))^{\lambda}} - C_F \frac{(F_{ij}(t))^{\gamma}}{\sum_{kl} (F_{kl}(t))^{\gamma}} \quad (3)$$

As this operation can cause negative activation, in the next step we take only the positive component of the map,

$$u^*(u) = \begin{cases} u, & \text{if } u > 0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

and finally add noise (ζ).

$$\pi(i,j) = (1-\zeta)\frac{u^*_{ij}}{\sum_{kl} u^*_{kl}} + (\zeta)\frac{1}{\sum_{kl} 1} \quad (5)$$

As the SceneWalk model implements concrete theory-based mechanisms, the parameters of the model have clear biological interpretations.

The Challenge: Systematic Tendencies

The performance of a scan path model can be quantified by the likelihood of empirical data given the model. In our model, the target map $\pi(i,j)$ for fixation selection can be used to directly read out the fixation likelihood for an upcoming experimentally-observed fixation (Schütt et al., 2017). In addition to likelihood-based inference, however, it is important to evaluate how well the model-generated data compare to the experimental data with respect to the empirically observed effects.

Fixation behavior produced by the SceneWalk model already resembles empirical scan paths on several important metrics such as the saccade amplitude distributions (see Fig. 1, top) or more complicated statistics like the pair correlation function of fixation locations (Engbert et al., 2015). Other systematic tendencies, however, are currently not reproduced by the model.

An example of an important statistic not reproduced by the SceneWalk model is the angle distribution of subsequent saccades (see Fig 1, middle). The characteristic "W"-shape of the empirical data shows that saccades are more likely to either continue in the same direction as the previous saccade or return in the direction of origin than to continue in any other direction. Not surprisingly, neither the SceneWalk model nor density sampling or homogeneous point processes capture this dynamic of saccades. The same is true of the relationship between fixation duration and change in saccade direction (Fig. 1, bottom).

Thus, these tendencies are caused by mechanisms present in the visual system, but not implemented in previous versions of our model. Adding new mechanisms to the model can significantly improve the agreement between simulated and experimental data, as shown by the successful addition of a Central Fixation Bias mechanism to the model (Rothkegel, Trukenbrod, Schütt, Wichmann, & Engbert, 2017). The following section will expand on how we used previously proposed features of visuomotor control and attention to motivate the update of the SceneWalk model.

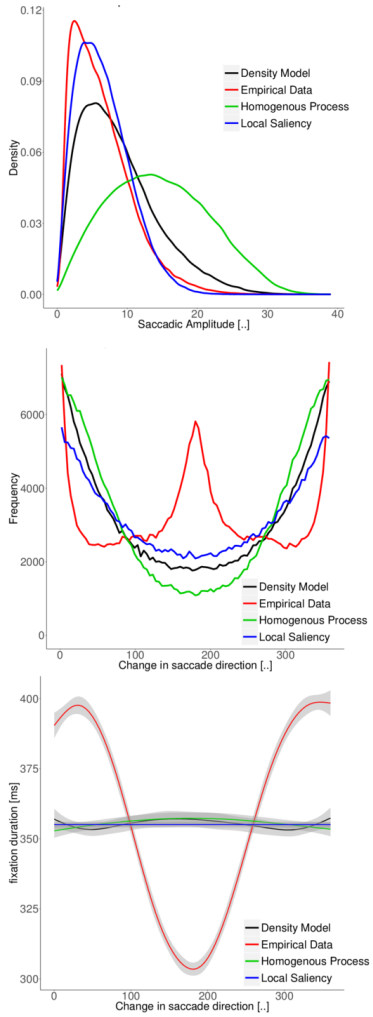

Figure 1: The figure outlines three systematic effects found in eye movement: saccade amplitude distribution, saccade angle distribution and the relationship of fixation durations and saccade angles. We compare empirical scan paths to simple sampling from a density, a homogeneous process and the local saliency as implemented in the SceneWalk model.

Extending SceneWalk

The existing literature includes evidence for attentional shifts directly preceding saccades (Deubel & Schneider, 1996) as well as attentional remapping immediately following saccades (Golomb, Chun, & Mazer, 2008). There are also indications that in addition to inhibition of return, in certain time frames there is also a facilitation of return (e.g., Smith & Henderson, 2009.

The proposed extensions of the SceneWalk model are based on splitting each fixation into three distinct phases of attention and saccade control:

- The **fixation phase** is implemented exactly as in the original SceneWalk model. At the end of this phase the upcoming fixation location is selected and at the beginning effects of the previous saccade are important.
- The **pre-saccadic phase** begins shortly before saccade onset. The attention Gaussian precedes the eye movement to the next fixation location while the inhibition Gaussian remains centered around the current position.
- The **post-saccadic phase** immediately follows the saccade. During this phase the attention Gaussian is shifted in the direction of the saccade, emulating a retinal remapping system.

To enable facilitation of return saccades in the model, we assume that there is a prolonged activation in the attention map at recently fixated locations. Mathematically, we implemented a location-dependent decay of the attention map, where a small window around the previous fixation location on the attention map decays slower (ω_{shift}) than on average for the map (ω_A).

In the next section, we report some qualitative analyses of the consequences of these model modification for scan path statistics.

Results

Using the extended SceneWalk model, we generate data and compare experimental scan paths from human participants with model-simulated data. As shown in Fig. 3, the model-simulated scan paths now qualitatively reproduce the shape of the saccade angle distribution. Furthermore, for the complex relationship between fixation durations and saccade angles we observe good qualitative agreement between experimental and simulated data (Fig. 4).

Our results lend support to the idea that pre- and post-saccadic attention shifts are responsible for some of the dynamics found in eye movement data. Thus, we find that neurally-inspired mechanisms are highly compatible with scan path generation when implemented within our dynamical framework of the SceneWalk model.

Outlook: Likelihood-based Parameter Inference

We set out to implement neurally-inspired visuomotor control principles to improve a model of scan path generation. Results reported here suggest that, with the modifications, the

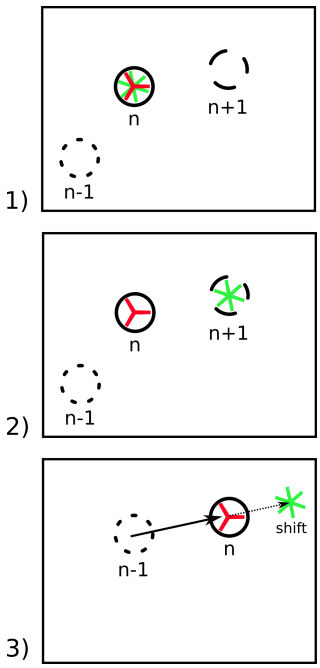

Figure 2: The three phases of the SceneWalk model. The circles indicate fixation positions, with n being the current. The green, 5-pointed star is the center of the attention Gaussian while the red, 3-pointed star is the center of the inhibitory Gaussian. Panel 1 shows the main fixation phase, where fixation position and the Gaussians align. In Panel 2 the center of the attention Gaussian moves to the upcoming fixation position, modelling a presaccadic attention shift. Panel 3 shows the post-saccadic attention shift where attention shifts in the direction of the saccade after fixation onset. After the post-saccadic phase, attention moves to the current fixation position, establishing the same situation as in Panel 1.

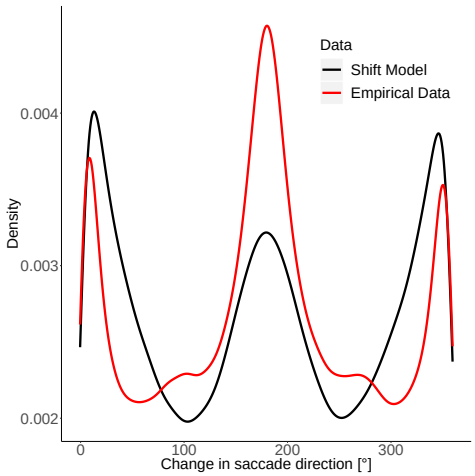

Figure 3: The extended SceneWalk model simulates data with a saccade angle distribution which qualitatively resembles the empirical distribution.

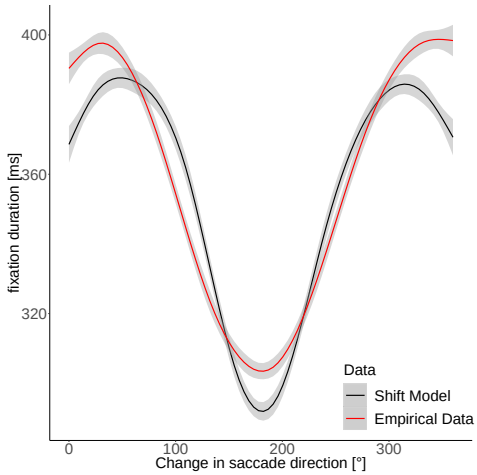

Figure 4: The relationship between saccade angle and fixation duration in the empirical data has a distinctive shape. The extended SceneWalk model produces a comparable relationship between the two measures.

SceneWalk model can be improved to include important systematic tendencies of eye-movement behavior. Ongoing work focuses on likelihood-based parameter inference for the extended version of the SceneWalk model.

The SceneWalk model generates a continuous-time evolution of a target map for upcoming fixations. For likelihood-based parameter inference, this target map provides an efficient tool to compute the likelihood for experimentally-observed fixation sequences. Thus, the likelihood can be computed numerically without approximation. Such models with a computable likelihood function are characterized by two considerable advantages. First, it is straightforward to estimate model parameters by maximizing the likelihood of the model given some empirical data. Moreover, since the model is implemented efficiently, the likelihood opens the door to a fully Bayesian framework (Schütt et al., 2017). Using a Differential Evolution Adaptive Metropolis Algorithm (Laloy & Vrugt, 2012) we obtained pilot results recently for the improved version of the model. Secondly, models with a likelihood function are more easily compared to competing framework, as comparisons does not have to rely on ad-hoc performance metrics that are, in most cases, motivated by experimental research but lack statistical rigor.

Finally, estimated parameters can then be fed back into the model to simulate data on the level of individual observers. The fit between simulated and experimental data will shed light on the dynamical system that produces fixation behavior, including interindividual differences in fixation behavior (de Haas et al., 2019) and the underlying visuomotor mechanisms.

Acknowledgments

This work is part (Project B05) of Collaborative Research Center 1294 *Data Assimilation* at the University of Potsdam, funded by Deutsche Forschungsgemeinschaft.

References

de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S., & Gegenfurtner, K. R. (2019). Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National Academy of Sciences*, 201820553. doi: 10.1073/pnas.1820553116

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837. doi: 10.1016/0042-6989(95)00294-4

Engbert, R., Trukenbrod, H. A., Barthelme, S., & Wichmann, F. A. (2015). Spatial statistics and attentional dynamics in scene viewing. *Journal of Vision*, *15*(1), 14. doi: 10.1167/15.1.14

Golomb, J. D., Chun, M. M., & Mazer, J. A. (2008). The native coordinate system of spatial attention is retinotopic. *Journal of Neuroscience*, *28*(42), 10654–10662. doi: 10.1523/jneurosci.2525-08.2008

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. doi: 10.1109/34.730558

Kümmerer, M., Wallis, T. S. A., & Bethge, M. (2016). Deepgaze ii: Reading fixations from deep features trained on object recognition. *ArXiv*.

Laloy, E., & Vrugt, J. A. (2012). High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing. *Water Resources Research*, *48*(1). doi: 10.1029/2011wr010608

Mirpour, K., Bolandnazar, Z., & Bisley, J. W. (2019). Neurons in FEF keep track of items that have been previously fixated in free viewing visual search. *The Journal of Neuroscience*, *39*(11), 2114–2124. doi: 10.1523/jneurosci.1767-18.2018

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*(8), 20. doi: 10.1167/10.8.20

Rothkegel, L. O. M., Trukenbrod, H. A., Schütt, H. H., Wichmann, F. A., & Engbert, R. (2017). Temporal evolution of the central fixation bias in scene viewing. *Journal of Vision*, *17*(13), 3. doi: 10.1167/17.13.3

Schütt, H. H., Rothkegel, L. O. M., Trukenbrod, H. A., Reich, S., Wichmann, F. A., & Engbert, R. (2017). Likelihood-based parameter estimation and comparison of dynamical cognitive models. *Psychological Review*, *124*(4), 505–524. doi: 10.1037/rev0000068

Smith, T. J., & Henderson, J. M. (2009). Facilitation of return during scene viewing. *Visual Cognition*, *17*(6-7), 1083–1108. doi: 10.1080/13506280802678557

Tatler, B. W., Vincent, B. T., et al. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*(2), 1–18. doi: 10.16910/jemr.2.2.5