Modeling the spread of object-based attention during free viewing

Nicolas Roth[1, 2], Olga Shurygina[1, 3], Flora M. Muscinelli[1], Klaus Obermayer[1, 2], and Martin Rolfs[1, 3]

[1]Cluster of Excellence Science of Intelligence, Technische Universität Berlin, Marchstraße 23, 10587 Berlin, Germany
[2]Institut für Softwaretechnik und Theoretische Informatik, Technische Universität Berlin, Marchstraße 23, 10587 Berlin, Germany
[3]Institut für Psychologie, Humboldt-Universität zu Berlin, Rudower Chaussee 18, 12489 Berlin, Germany

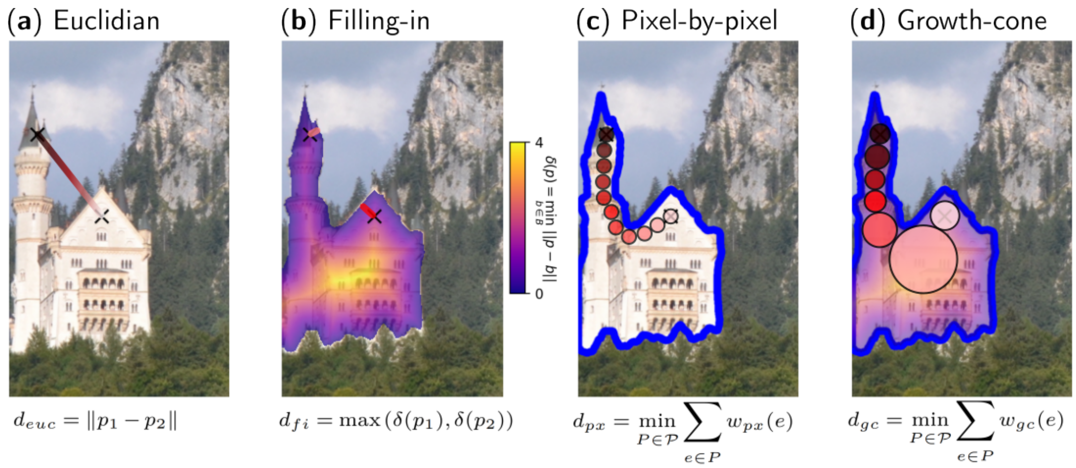

Fig. 1. Overview of the investigated models for how attention might spread when making a saccade from point p_1 to point p_2 (black crosses) within the same object. (a) The *Euclidean model* measures the distance d_{euc} between the points independent of the object. (b) In the *filling-in model*, attention spreads inwards from the object boundary B. The distance d_{fi} is then given by the larger distance from the boundary δ (plotted from blue to yellow for all points in the object) of the two points, $\delta(p_1)$ and $\delta(p_2)$. (c) The *pixel-by-pixel model* finds the shortest path P from the set of all possible paths $\mathcal{P}$ that connect the points p_1 and p_2, and lie within the interior of the object boundary B (shown in blue). Attention spreads with a fixed speed, with weights $w_{px}(e) = 1$ for horizontally and vertically aligned pixels, and $w_{px}(e) = \sqrt{2}$ for diagonally aligned pixels. The distance d_{px} is computed as the combined weight of all edges $e \in P$ and converted into degrees visual angle (dva). (d) The *growth-cone model* also finds the shortest path P within the object boundary but assumes that attention spreads faster in homogeneous image regions due to perceptual grouping occuring on multiple spatial scales. The grouping speed at an edge e between two neighboring pixels n_1 and n_2 is then determined by $\delta(n_1)$ and $\delta(n_2)$, which correspond to the radii of the largest circles that can be inscribed in the object boundary around these pixels, resulting in weights of $w_{gc}(e) = \frac{w_{px}(e)}{\delta(n_1)+\delta(n_2)}$. The distance d_{gc} is measured in number of growth-cones.

The time course of how object-based attention spreads provides important insights into the mechanisms of perceptual organization. Neurophysiological experiments of curve tracing have demonstrated that object-based attention spreads gradually over object representations and slows down in the presence of distractor objects [1]. These results were explained by a growth-cone model, which has since also described the serial grouping across simple 2D objects in behavioral studies: Jeurissen et al. [2] showed that during fixation, the perceptual grouping as reflected in the growth-cone model described reaction times to a cue presented at different eccentricities within the same object best, as long as the objects are not too complex.

Here, we investigated if fixation durations during free-viewing of real-world scenes also reflect a serial spread of attention that is fastest across large areas and is slowed down at locations that require detailed visual analysis. Based on the assumption that the spread of attention within an object determines the speed of visual processing, we reasoned that the fixation duration before a saccade should correlate with the time it takes for attention to spread from its start to endpoint. Although fixation durations vary considerably due to factors like scene content or saccade kinematics, the best model of how attention spreads under ecologically valid conditions should explain more variance in the fixation duration data than alternative models. We used free-viewing eye-tracking data on real-world images from the Potsdam Scene-Viewing Corpus [3] as well as our own recordings, from which we extracted the saccades in which the start and endpoints are located within the boundaries of the same objects. We compared four different models for how attention might spread across objects, each measuring the distance, over which the saccade had to be prepared, in different ways: (a) Euclidean distance (irrespective of object boundaries), (b) Filling-in distance (inwards from object's boundaries), (c) Pixel-by-pixel distance (the shortest path within the object region), and (d) Growth-cone distance (the shortest path within the object, but depending on the distance from object's boundaries). Using generalized linear mixed models, we assessed to what extent the distance of a saccade, as measured by the four models of spread, can account best for the variance in fixation duration. Among these models, the number of growth-cones between saccade start and endpoints best explained fixation duration within objects. Our results suggest that fixation duration in free viewing is affected by a serial spread of object-based attention, and that the growth-cone model can provide a better understanding of this attention spread and perceptual grouping, also in ecologically valid settings.

REFERENCES

1. Arezoo Pooresmaeili and Pieter R Roelfsema. A growth-cone model for the spread of object-based attention during contour grouping. *Current Biology*, 24(24):2869–2877, 2014.
2. Danique Jeurissen, Matthew W Self, and Pieter R Roelfsema. Serial grouping of 2d-image regions with object-based attention in humans. *Elife*, 5:e14320, 2016.
3. Heiko H Schütt, Lars OM Rothkegel, Hans A Trukenbrod, Ralf Engbert, and Felix A Wichmann. Disentangling bottom-up versus top-down and low-level versus high-level influences on eye movements over time. *Journal of vision*, 19(3):1–1, 2019.