Human Vision at a Glance

Ruth Rosenholtz[1]
Massachusetts Institute of Technology, Department of Brain & Cognitive Sciences, CSAIL, USA
http://persci.mit.edu/people/rosenholtz
rruth@mit.edu

Dian Yu
Massachusetts Institute of Technology, Department of Brain & Cognitive Sciences, CSAIL, USA
dyumit17@gmail.com

Abstract

Recent advances in human vision research have pointed toward a theory that unifies many aspects of vision relevant to information visualization. According to this theory, loss of information in peripheral vision determines performance on many visual tasks. This theory subsumes old concepts such as visual saliency, selective attention, and change blindness. It predicts the rich details we have access to at a glance. Furthermore, it provides insight into tasks not commonly studied in human vision, such as ability to comprehend connections in a network diagram, or to compare information in one part of a display with that in another.

2012 ACM Subject Classification Human-centered computing → Visualization design and evaluation methods; Human-centered computing → Visualization theory, concepts and paradigms

Keywords and phrases human vision, information visualization, attention, eye movements, peripheral vision, gist, ensemble perception, search, saliency

Digital Object Identifier 10.4230/LIPIcs.COSIT.2019.1

Category Invited Talk

Funding *Ruth Rosenholtz*: funded in part by NSF BCS-1826757
Dian Yu: funded in part by NSF/NIH/BMBF IIS-1607486 to R. Rosenholtz

1 Relevance of human vision to design

Designs, whether of information visualizations or of user interfaces, can be good or bad for both cognitive and perceptual reasons. At the bare minimum, designers strive to make relevant information easily perceptible. If a legend is not easily discriminable from the rest of the display, a user will not easily notice it. If icons with distinct meaning do not also have a sufficiently distinct appearance, a user will need to search for the icons of interest, an often slow process, and they will have trouble getting the "gist" of the display, e.g. the spatial distribution of the two types of icons. In addition, even if information is readily available to perception, a confusing design can make tasks cognitively difficult, for example when icons are clearly identifiable but have unclear meaning. Here we focus on some important perceptual issues:

1. How easy will it be for users to find information in a display?
2. Will users notice important and unexpected information, such as an alert?
3. In dynamic displays, will users notice and make sense of information that changes?
4. How difficult is it to connect related information in multi-view displays?
5. What information can the user get at a glance at a display, i.e. what "gist" can they easily extract?

[1] Corresponding author

14th International Conference on Spatial Information Theory (COSIT 2019).
Editors: Sabine Timpf, Christoph Schlieder, Markus Kattenbeck, Bernd Ludwig, and Kathleen Stewart;
Article No. 1; pp. 1:1–1:4
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

6. What distinguishes details that users can perceive at a glance from those they are more likely to miss?
7. Many visualizations like maps and graphs contain complex connections between nodes. Can the user easily make sense of these connections?

Classic vision science has studied these issues as essentially six separate problems: search, and saliency of "alerts" (items 1, 2); change or difference detection (3, 4); scene gist (5, 6); gist of an ensemble of items (5, 6); what details require "attention" to perceive veridically (3, 6); and visual cognition problems such as graph and maze connectedness (7). While vision science has elucidated a number of important phenomena, the results have often been unsatisfying in terms of translation to applications such as information visualization.

First, in the interest of pinpointing visual mechanisms, vision scientists have quite reasonably studied simple controlled displays. For instance, experiments have asked observers to locate a red item amongst a large number of homogeneous green items. Such experiments aimed to determine what feature differences (color, size, etc.) lead unusual items to "pop out" and "draw the observer's attention"; a question related to whether a user will notice an alert. These experiments have been interesting for probing vision, but such homogeneous displays have limited applicability to more complex real-world designs.

Second, much of the relevant vision science has largely remained at the level of descriptive enumeration of behavioral results, rather than transitioning to predictive models. For instance, suppose you have two alternative designs for a subway map, and want to know which makes it easier for observers to find a route between two stations. Vision science could provide some good rules of thumb such as "high contrast lines are better". But it cannot provide a direct answer to the original question: which design is better, and by how much? This reliance on descriptive rather than predictive models is particularly problematic since vision science has produced few behavioral results related to perception of complex applied displays.

Third, to the extent that vision science has unified these 7 key perceptual issues for designs, it has done so by relating them all to visual attention: many details are difficult to rapidly perceive because they require focused attention [18]; gist is easy because it does not require focused attention [17, 19]; change detection is difficult because it does require attention to the change [5, 9, 13]; tracing a path through a graph is difficult because attention needs time to spread along each route [6]. The problem with these theories, even if correct, is that attention is nearly impossible to measure using behavioral methods alone. Attempts to fix this problem by using eye movements as a proxy for attention have been modestly successful only in very limited situations [16].

New developments in vision science, over the last decade, have instead suggested that many perceptual issues of relevance to design largely depend upon a single factor: the information lost and maintained in peripheral vision (see [10] for a review). Human performance of visual tasks relies greatly on peripheral vision, the region outside the rod-free fovea, comprising more than 99% of the visual field. We mainly use our foveas for work that requires precision and ultra-fine details, such as reading a paragraph of text or making careful measurements. Peripheral vision, on the other hand, is critical for efficient processing of large portions of a display or scene, e.g. to get the gist of a scene, to notice an alert or find a target, or to compare two conditions in a plot or two views of the same data. In fact, one might often want to design for maximum information gathering with a minimal number of eye movements, making use of efficient peripheral processing to understand a display at a glance (or anyhow, a small number of glances).

Understanding of peripheral vision in fact appears to subsume understanding of many important visual phenomena, from saliency and search, through change detection and gist perception. Some items are salient because they can easily be seen in the periphery, due to distinct visual features or a lack of clutter [1, 8, 12, 14, 20]. The gist, or information readily available at a glance, derives from the information that survives peripheral vision. The details a user might miss are those less readily available in the periphery, sometimes leading to a failure of the user to detect changes to a dynamic display [15]. Peripheral vision can also inform us about the usability of network graphs and maps – topics of relevance to a number of applications, but little studied in vision science.

Human vision depends upon peripheral vision in spite of the fact that peripheral vision loses a great deal of information. The big loss of information has to do with phenomena known as "crowding": the degradation in visual performance in the presence of clutter. These losses are complex, stimulus-specific, and it is hard to get good intuitions about what information is lost and what is preserved.

Our lab has spent the last decade studying the effect of peripheral vision on many visual tasks. We have developed and extensively tested the state-of-the-art model of peripheral vision [2–4, 7, 11, 12, 20]. We call this the Texture Tiling Model because peripheral vision appears to compress and summarize large regions of the visual world by treating them as texture. This model makes predictions about what information is preserved and lost in peripheral vision, and does so in an intuitive way; it outputs visualizations of the information available to peripheral vision. To date, we have demonstrated that this model does well at predicting performance on over 70 visual tasks.

From the point of view of understanding one's users, it is good news that peripheral vision has proven a more powerful explanation than attention. Whereas attention is difficult to measure it is relatively easy to know where a user is pointing their eyes and therefore what portion of the visual field lands in the periphery. On the other hand, the importance of peripheral vision means that we must interpret eye-tracking results with care; users will often be processing visual input across large portions of the visual field, not just where they point their eyes.

To get intuitions about what tasks users can successfully complete with peripheral vision, one can use one's own visual system. Rather than the old advice to "squint" at a design, point your eyes at locations the user is likely to fixate (for example, on a button they must click), and introspect on what information is clear or confusing in the periphery. Alternatively, one can inspect the visualizations output by our Texture Tiling Model – with no need to fixate – to get intuitions about what information the model predicts is and is not preserved in peripheral vision. One need not even entirely trust the model, as again one can verify any predictions by using one's own visual system to introspect on the information available.

References

1. J. J. Andriessen and H. Bouma. Eccentric vision: Adverse interactions between line segments. *Vision Research*, 16(1):71–78, 1976.
2. B. J. Balas, L. Nakano, and R. Rosenholtz. A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, 9(12):13, 2009.
3. H. Chang and R. Rosenholtz. Search performance is better predicted by tileability than by the presence of a unique basic feature. *Journal of Vision*, 16(10):13, 2016.
4. K. A. Ehinger and R. Rosenholtz. A general account of peripheral encoding also predicts scene perception performance. *Journal of Vision*, 16(2):13, 2016.
5. A. Hollingworth and J. M. Henderson. Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception & Performance*, 28(1):113–136, 2002.
6. R. Houtkamp, H. Spekreijse, and P.R. Roelfsema. A gradual spread of attention. *Perception & Psychophysics*, 65(7):1136–1144, 2003. `doi:10.3758/BF03194840`.

7 S. Keshvari and R. Rosenholtz. Pooling of continuous feature provides a unifying account of crowding. *Journal of Vision*, 16(3):39, 2016.

8 F. L. Kooi, A. Toet, S. P. Tripathy, and D. M. Levi. The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision*, 8(2):255–279, 1994.

9 R. A. Rensink, J. K. O'Regan, and J. J. Clark. To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8:368–373, 1997.

10 R. Rosenholtz. Capabilities and limitations of peripheral vision. *Annual Review of Vision Science*, 2(1):437–457, 2016. doi:10.1146/annurev-vision-082114-035733.

11 R. Rosenholtz, J. Huang, and K. A. Ehinger. Rethinking the role of top-down attention in vision: effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology*, 3:13, 2012. doi:10.3389/fpsyg.2012.00013.

12 R. Rosenholtz, J. Huang, A. Raj, B. Balas, and L. Ilie. A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, 12(14):1–17, 2012. doi:10.1167/12.4.14.

13 B. J. Scholl. Attenuated change blindness for exogenously attended items in a flicker paradigm. *Visual Cognition*, 7(1-3):377–396, 2000. doi:10.1080/135062800394856.

14 M. Scolari, A. Kohnen, B. Barton, and E. Awh. Spatial attention, preview, and popout: Which factors influence critical spacing in crowded displays? *Journal of Vision*, 7(2):7, 2007.

15 M. E. Smith, L. Sharan, E. Park, L. C. Loschky, and R. Rosenholtz. Difficulty detecting changes in complex scenes depends in part upon the strengths and limitations of peripheral vision. *Journal of Vision*, under revision.

16 B. W. Tatler, M. M. Hayhoe, M. F. Land, and D. H. Ballard. Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, 11(5):5, 2011. doi:10.1167/11.5.5.

17 A. Treisman. How the deployment of attention determines what we see. *Visual Cognition*, 14:411–443, 2006.

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

19 J. M. Wolfe, M. L.-H. Vo, K. K. Evans, and M. R. Greene. Visual search in scenes involves selective and non-selective pathways. *Trends in Cognitive Sciences*, 15(2):77–84, 2011.

20 X. Zhang, J. Huang, S. Yigit-Elliot, and R. Rosenholtz. Cube search, revisited. *Journal of Vision*, 15(3):9, 2015.