Toward A Pervasive Gaze-Contingent Assistance System

Attention and Context-Awareness in Augmented Reality

Kenan Bektaş
University of St. Gallen
Institute of Computer Science
kenan.bektas@unisg.ch

Figure 1: A gaze-enabled system may assess user's attention and provide contextual help when the user appears to need it. The real-world view is augmented with contextual overlays: the remaining time and distance to the next tram, and an inset map. For a less cluttered view the contrast of the inset (©OpenStreetMap contributors) is reduced in a gaze-contingent manner.

ABSTRACT

Mobile devices with high-speed connectivity provide us with access to gigabytes of high resolution images, videos, and graphics. For instance, a head-worn display can be used to augment the real view with digitized visual information (Figure 1). Eye tracking helps us to understand how we process visual information and it allows us to develop gaze-enabled interactive systems. For instance, foveated gaze-contingent displays (GCDs) dynamically adjust the level of detail according to the user's point-of-interest. We propose that GCDs should take users' attention and cognitive load into account, augment their vision with contextual information and provide personalized assistance in solving visual tasks. Grounded on existing literature, we identified several research questions that need to be discussed before developing such displays.

CCS CONCEPTS

• Human-centered computing → Ubiquitous and mobile computing systems and tools; **• Computing methodologies → Mixed / augmented reality**; **Perception**.

ETRA 2020, June 02, 2020, Stuttgart, Germany

ACM ISBN 978-1-4503-7135-3/20/06…$15.00
https://doi.org/10.1145/3379157.3391657

KEYWORDS

Gaze-Contingent Displays, Foveation, Pervasive Eye Tracking, Augmented Reality, Visual Complexity, Attention, Context-Awareness.

ACM Reference Format:
Kenan Bektaş. 2020. Toward A Pervasive Gaze-Contingent Assistance System: Attention and Context-Awareness in Augmented Reality. In *Symposium on Eye Tracking Research and Applications (ETRA '20 Adjunct), June 2–5, 2020, Stuttgart, Germany*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3379157.3391657

1 INTRODUCTION AND BACKGROUND

There is a growing interest in augmented reality (AR) applications that can overlay various forms of multimedia (e.g., image, video, graphics) onto a user's physical environment [Grubert et al. 2016; Orlosky et al. 2017]. However, finding relevant information in digital displays can become a problem due to information load and visual complexity. Various Level-of-Detail (LOD) management solutions exist to address such computational and perceptual problems. For instance, Focus+Context (F+C) displays [Baudisch et al. 2003] filter relevant information based on the manual interaction. Alternatively, depending on the user's eye movements, foveated GCDs retain high-resolution information within the foveal visual field and discard details toward out-of-focus regions [Bektaş et al. 2015; Duchowski and Çöltekin 2007]. Eye tracking becomes a default feature in many head-mounted displays, because it enables foveated rendering, depth-of-field simulation, and studying the user's viewing behavior [Stengel and Magnor 2016]. By mimicking human vision, GCDs can substantially improve rendering of virtual scenes, and they are relevant to many virtual reality (VR) and AR applications in the future [Guenter et al. 2012; Kim et al. 2019].

We consider developing a novel GCD that, while assessing user's attention and cognitive load, provides them a perceptually optimized display and an effective assistance when solving visual tasks in a specific context (e.g., navigation, driving, piloting, medical care). But how can we create visual tools that provide assistance while keeping the user's attention and cognitive load in check?

1.1 Pervasive Eye Tracking and AR

In real and simulated settings, eye tracking is considered to be useful in two ways. First, it can be used to assess user's attention, situation awareness, cognitive load, and performance in real time or within a specific time frame [Borojeni et al. 2016; Duchowski et al. 2019, 2018; Huang et al. 2019; Kiefer et al. 2016; Lim et al. 2018; Lindlbauer et al. 2019; Toet 2006]. Second, user's eye movements can guide an instant adjustment of the information load on a display [Anthes 2019; Duchowski 2018]. Recently, head-mounted eye trackers with automated assessment capabilities allow for an unobtrusive interaction that is usable in many daily activities, i.e., *pervasive eye tracking* [Bulling and Gellersen 2010; Kassner et al. 2014]. Gaze-enabled systems become feasible in AR applications that can continuously adapt to the requirements and constraints of the user's context, i.e., *pervasive AR* [Grubert et al. 2016].

1.2 Models for Gaze-Contingent Viewing

A GCD would be indistinguishable from a conventional display, provided that the underlying models are implemented in a perceptually plausible way [Parkhurst and Niebur 2002; Reingold et al. 2003]. Individual visual perception models (VPMs) such as depth-of-field simulation and foveation have often been studied to address the computational and perceptual benefits of GCDs. For instance, foveation approximates the neuro-physiological limits for an optimal image coding, compression, rendering, and quality assessment [Floren and Bovik 2014]. Compared to such individual models, combined models can concurrently discard spatial and chromatic details from the out-of-focus parts of a display [Bektaş et al. 2019]. In particular cases, researchers found that GCDs can improve users' visual search performance [Bektaş et al. 2019; Murphy et al. 2009], and they can improve the reading capabilities of children with developmental dyslexia [Schneider et al. 2011]. Thus, the effect of *visual crowding* [Pelli et al. 2004] can be reduced with a GCD that masks peripheral distractors. In other cases, GCDs might attract, guide or enhance the capacity of users' attention [Loschky and McConkie 2002; Toet 2006]. Models of motion perception, foveal masking, transparency, stereoscopic vision; semantic or content-aware models; and F+C methods can be tested with GCDs [Bektaş and Çöltekin 2011; Bektaş 2018]. However, a systematic user evaluation of such models in GCDs seems to be missing from the literature.

1.3 Visual Complexity and Quality of GCDs

The more objects are present in our visual field, the more difficult it becomes to detect and identify relevant objects [Eckstein 2011; Wolfe et al. 2011], and such displays are often characterized as complex or cluttered [Rosenholtz et al. 2007]. There are objective and subjective ways of measuring the visual complexity on a display [Rosenholtz et al. 2007; Schnur et al. 2018]. Similarly, conventional image quality metrics [Wang and Bovik 2006] rely on the assumption that the image resolution is uniform [Wang and Bovik 2001], thus they are not directly applicable to foveated rendering [Wang et al. 2001]. Researchers proposed foveated image quality metrics, based on single [Floren and Bovik 2014] or multiple salient image features [Swafford et al. 2016]. These metrics appear to be applicable to virtual reality content. However, for VR and AR applications, we need foveated quality and clutter metrics that account for the temporal changes of user's point-of-interest (i.e., eye movements). Researchers studied how image complexity [Bonev et al. 2013] and cognitive load [Allsop et al. 2017] might affect eye movements on conventional displays. We need to understand to what extent foveated GCDs can reduce the visual complexity and the cognitive load while preserving the image quality.

2 RESEARCH QUESTIONS

Gaze-enabled systems allow for a better understanding of the visual perception and pave the way toward recreating "human visual search abilities in machines" [Eckstein 2011]. Motivated by the earlier research presented above, we want to address the following research questions:

- How can eye tracking help us to make an assessment of attention, cognitive load, situation awareness, and performance while the user performs a visual task (e.g., search, navigation or piloting)? Which metrics are suitable for a low-latency or even real-time assessment?
- Based on the requirements and constraints of the user, the environment and the system (i.e., context sources), which individual VPMs and their combinations can be enabled in a GCD for real-time LOD management?
- How can we estimate the visual complexity and image quality on a GCD (i.e., a fixation-based or frame by frame estimation)? How does a VPM (e.g., foveation) affect the saliency of virtual overlays and real objects?
- How does a GCD affect its user's situation awareness, cognitive load, attention or task performance compared to conventional displays?

To address these questions we want to develop a *pervasive GCD* that can serve both as an assessment and assistance tool in an AR setup. Concretely, the pervasive GCD should afford an *attention-aware* and *context-aware* assistance while reducing the visual complexity in a perceptually plausible way (Figure 1). Our envisioned pervasive GCD is able to:

- **Predict** users' cognitive load and attention based on existing eye tracking measurements such as pupil dilation and vergence.
- **Implement** individual VPMs and their combinations to discard perceptually redundant and contextually irrelevant details from the given display in real time.
- **Evaluate** the effect of the gaze-contingent viewing on users' situation awareness and task performance.

In this paper, we proposed preliminary research questions we aim to address in the future. We are looking forward to discuss them at the Workshop on Eye Tracking for Quality of Experience in Multimedia (ET-MM 2020).

REFERENCES

Jonathan Allsop, Rob Gray, Heinrich H Bülthoff, and Lewis Chuang. 2017. Eye movement planning on Single-Sensor-Single-Indicator displays is vulnerable to user anxiety and cognitive load. *Journal of Eye Movement Research* 10, 5 (2017), 8–1.

Gary Anthes. 2019. Augmented reality gets real. *Commun. ACM* 62, 9 (2019), 16–18. https://doi.org/10.1145/3344293

Patrick Baudisch, Doug DeCarlo, Andrew T. Duchowski, and Wilson S. Geisler. 2003. Focusing on the essential: considering attention in display design. *Commun. ACM* 46, 3 (mar 2003), 60. https://doi.org/10.1145/636772.636799

Kenan Bektaş and Arzu Çöltekin. 2011. An Approach to Modeling Spatial Perception for Geovisualization. *Procedia - Social and Behavioral Sciences* 21 (2011), 53–62. https://doi.org/10.1016/j.sbspro.2011.07.027

Kenan Bektaş, Arzu Çöltekin, Jens Krüger, and Andrew T Duchowski. 2015. A Testbed Combining Visual Perception Models for Geographic Gaze Contingent Displays. In *Eurographics Conference on Visualization (EuroVis) – Short Papers*, E. Bertini, J. Kennedy, and E. Puppo (Eds.). 67–71. https://doi.org/10.2312/eurovisshort.20151127

Kenan Bektaş. 2018. *Gaze-Contingent Geovisualization for Level of Detail Management* . Ph.D. Dissertation. University of Zurich.

Kenan Bektaş, Arzu Çöltekin, Jens Krüger, Andrew T Duchowski, and Sara Irina Fabrikant. 2019. GeoGCD: Improved Visual Search via Gaze-Contingent Display. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (Denver, Colorado) *(ETRA '19)*. Association for Computing Machinery, New York, NY, USA, Article 84, 10 pages. https://doi.org/10.1145/3317959.3321488

Boyan Bonev, Lewis Chuang, and Francisco Escolano. 2013. How do image complexity, task demands and looking biases influence human gaze behavior? *Pattern Recognition Letters* 34, 7 (2013), 723–730. https://doi.org/10.1016/j.patrec.2012.05.007

Shadan Sadeghian Borojeni, Lewis Chuang, Wilko Heuten, and Susanne Boll. 2016. Assisting drivers with ambient take-over requests in highly automated driving. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 237–244. https://doi.org/10.1145/3003715.3005409

Andreas Bulling and Hans Gellersen. 2010. Toward mobile eye-based human-computer interaction. *IEEE Pervasive Computing* 9, 4 (2010), 8–12. https://doi.org/10.1109/MPRV.2010.86

Andrew Duchowski, Krzysztof Krejtz, Justyna Zurawska, and Donald House. 2019. Using microsaccades to estimate task difficulty during visual search of layered surfaces. *IEEE transactions on visualization and computer graphics* (2019). https://doi.org/10.1109/TVCG.2019.2901881

Andrew T Duchowski. 2018. Gaze-based interaction: A 30 year retrospective. *Computers & Graphics* 73 (2018), 59–69. https://doi.org/10.1016/j.cag.2018.04.002

Andrew T. Duchowski and Arzu Çöltekin. 2007. Foveated gaze-contingent displays for peripheral LOD management, 3D visualization, and stereo imaging. *ACM Transactions on Multimedia Computing, Communications, and Applications* 3, 4 (2007), 1–18. https://doi.org/10.1145/1314303.1314309

Andrew T Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, Martin Raubal, and Ioannis Giannopoulos. 2018. The Index of Pupillary Activity: Measuring Cognitive Load vis-à-vis Task Difficulty with Pupil Oscillation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 282. https://doi.org/10.1145/3173574.3173856

Miguel P Eckstein. 2011. Visual search: a retrospective. *Journal of vision* 11, 5 (jan 2011). https://doi.org/10.1167/11.5.14

Andrew Floren and Alan C. Bovik. 2014. Foveated Image and Video Processing and Search. In *Academic Press Library in Signal Processing: Volume 4 — Image, Video Processing and Analysis, Hardware, Audio, Acoustic and Speech Processing*. Chapter 14, 349–401. https://doi.org/10.1016/B978-0-12-396501-1.00014-5

Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. 2016. Towards pervasive augmented reality: Context-awareness in augmented reality. *IEEE transactions on visualization and computer graphics* 23, 6 (2016), 1706–1724. https://doi.org/10.1109/TVCG.2016.2543720

Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. 2012. Foveated 3D graphics. *ACM Transactions on Graphics* 31, 6 (nov 2012), 1. https://doi.org/10.1145/2366145.2366183

Michael Xuelin Huang, Jiajia Li, Grace Ngai, Hong Va Leong, and Andreas Bulling. 2019. Moment-to-Moment Detection of Internal Thought during Video Viewing from Eye Vergence Behavior. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. Association for Computing Machinery, New York, NY, USA, 2254–2262. https://doi.org/10.1145/3343031.3350573

Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication*. 1151–1160. https://doi.org/10.1145/2638728.2641695

Peter Kiefer, Ioannis Giannopoulos, Andrew Duchowski, and Martin Raubal. 2016. Measuring Cognitive Load for Map Tasks Through Pupil Diameter. *Geographic Information Science: 9th International Conference, GIScience 2016, Montreal, QC, Canada, September 27-30, 2016, Proceedings* (2016), 323–337. https://doi.org/10.1007/978-3-319-45738-3_21

Jonghyun Kim, Youngmo Jeong, Michael Stengel, Kaan Akşit, Rachel Albert, Ben Boudaoud, Trey Greer, Joohwan Kim, Ward Lopes, Zander Majercik, et al. 2019. Foveated AR: dynamically-foveated augmented reality display. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–15. https://doi.org/10.1145/3306346.3322987

Yixiang Lim, Subramanian Ramasamy, Alessandro Gardi, Trevor Kistan, and Roberto Sabatini. 2018. Cognitive human-machine interfaces and interactions for unmanned aircraft. *Journal of Intelligent & Robotic Systems* 91, 3-4 (2018), 755–774. https://doi.org/10.1007/s10846-017-0648-9

David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. 2019. Context-Aware Online Adaptation of Mixed Reality Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 147–160. https://doi.org/10.1145/3332165.3347945

Lester C. Loschky and George W. McConkie. 2002. Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display. *Journal of Experimental Psychology: Applied* 8, 2 (2002), 99–117. https://doi.org/10.1037/1076-898X.8.2.99

Hunter a. Murphy, Andrew T. Duchowski, and Richard a. Tyrrell. 2009. Hybrid image/model-based gaze-contingent rendering. *ACM Transactions on Applied Perception* 5, 4 (jan 2009), 1–21. https://doi.org/10.1145/1462048.1462053

Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. 2017. Virtual and augmented reality on the 5G highway. *Journal of Information Processing* 25 (2017), 133–141. https://doi.org/10.2197/ipsjjip.25.133

Derrick J Parkhurst and Ernst Niebur. 2002. Variable-resolution displays: a theoretical, practical, and behavioral evaluation. *Human factors* 44, 4 (2002), 611–29. https://doi.org/10.1518/0018720024497015

Denis Pelli, Melanie Palomares, and Najib J Majaj. 2004. Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of vision* 4, 12 (2004), 12. https://doi.org/10.1167/4.12.12

Eyal M Reingold, Lester C Loschky, George W McConkie, and David M Stampe. 2003. Gaze-Contingent Multiresolutional Displays: An Integrative Review. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 45, 2 (2003), 307–328. https://doi.org/10.1518/hfes.45.2.307.27235

Ruth Rosenholtz, Yuanzhen Li, and Lisa Nakano. 2007. Measuring visual clutter. *Journal of Vision* 7, 2 (aug 2007), 17. https://doi.org/10.1167/7.2.17

Nicolas Schneider, Michael Dorr, Laura Pomarjanschi, and Erhardt Barth. 2011. A gaze-contingent reading tutor program for children with developmental dyslexia. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization - APGV '11*. ACM Press, New York, New York, USA, 117. https://doi.org/10.1145/2077451.2077480

Susan Schnur, Kenan Bektaş, and Arzu Çöltekin. 2018. Measured and perceived visual complexity: A comparative study among three online map providers. *Cartography and Geographic Information Science* 45, 3 (2018), 238–254. https://doi.org/10.1080/15230406.2017.1323676

Michael Stengel and Marcus Magnor. 2016. Gaze-Contingent Computational Displays: Boosting perceptual fidelity. *IEEE Signal Processing Magazine* 33, 5 (2016), 139–148. https://doi.org/10.1109/MSP.2016.2580913

Nicholas T. Swafford, José A. Iglesias-Guitian, Charalampos Koniaris, Bochang Moon, Darren Cosker, and Kenny Mitchell. 2016. User, Metric, and Computational Evaluation of Foveated Rendering Methods. In *Proceedings of the ACM Symposium on Applied Perception* (Anaheim, CA) *(SAP '16)*. ACM, New York, NY, 7–14. https://doi.org/10.1145/2931002.2931011

Alexander Toet. 2006. Gaze directed displays as an enabling technology for attention aware systems. *Computers in Human Behavior* 22, 4 (2006), 615–647. https://doi.org/10.1016/j.chb.2005.12.010

Zhou Wang and Alan Bovik. 2001. Embedded foveation image coding. *IEEE transactions on image processing* 10, 10 (jan 2001), 1397–1410. https://doi.org/10.1109/83.951527

Zhou Wang, Alan Bovik, Ligang Lu, and Jack L Kouloheris. 2001. Foveated wavelet image quality index. In *Applications of Digital Image Processing XXIV*, Vol. 4472. International Society for Optics and Photonics, 42–52. https://doi.org/10.1117/12.449797

Zhou Wang and Alan C. Bovik. 2006. Modern Image Quality Assessment. *Synthesis Lectures on Image, Video, and Multimedia Processing* 2, 1 (jan 2006), 1–156. https://doi.org/10.2200/S00010ED1V01Y200508IVM003

Jeremy M Wolfe, George A Alvarez, Ruth Rosenholtz, Yoana I Kuzmova, and Ashley M Sherman. 2011. Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics* 73, 6 (2011), 1650. https://doi.org/10.3758/s13414-011-0153-3