Focusing on the Essential: Considering Attention in Display Design

Patrick Baudisch

Doug DeCarlo

Andrew T. Duchowski

Wilson S. Geisler

Larger screens and higher resolution enhance the viewing experience by allowing for deeper immersion. Recent research shows that a wider field of view can lead to increased performance in productivity tasks [2]. Over the past years, industry has been addressing the resulting demand by offering displays of steadily increasing resolution, reaching resolutions of over 9 million pixels (IBM T220 display). Although high resolution is desirable for a variety of applications, it results in an ongoing challenge for creators of rendering hardware, as the large number of pixels makes these displays especially hungry for computational resources. Displaying computationally intense graphics, such as flight simulation or interactive scientific visualization requires considerable rendering effort. It is important to note that when computing power is insufficient to support the task of the user, any benefits of large screen technology to user productivity may be negated. In many cases, this issue can be addressed with parallel rendering hardware; a display system consisting of an array of projectors, for example, is often driven by an array of PCs or a PC "cluster". Parallel hardware, however, leads to substantially increased costs and space requirements. Furthermore, in the case of projector array based displays, the increased need for rendering hardware is accompanied by the cost of the projector technology. So when people make decisions about display sizes and resolutions, not only do cost and space requirements for rendering hardware and display play an important role: user productivity itself is at stake.

Several research projects have addressed the demand for rendering power and display resolution by taking the user's attentional focus into account. In this article, we will use the term attentive displays to refer to this class of techniques. Attentive displays address the demand for rendering power and display resolution, but their approach is different from the parallel display approach described above. Instead of requiring additional hardware, attentive displays make more out of available hardware by directing display and computation resources to where it counts most. As we'll see in this paper, a single user can only be attending to a relatively small part of a display at a time (while multiple users can be attending to more than one location, presentations to groups are more likely to motivate more substantial computational resources). So, instead of rendering information at the same detail everywhere, these displays track the user's attention and render information in full detail only at the user's current focus of attention, while reducing information in peripheral areas. By shifting computational power from peripheral regions to the

region located in the user's focus of attention, attentive display systems can provide faster response times with higher subjective display quality than systems distributing their resources equally across the screen.

In order to achieve the focal effect, attentive displays need an add-on—a device informing the display about the user's current focus of attention. While other related approaches use models of attention based on properties of the displayed scene [5], most of the approaches we survey in this article use an eye tracker for this purpose. While eye tracking has long involved complex technology, recent technological progress in this area, as well as comparably moderate accuracy requirements, allow attentive displays to use relatively simple and inexpensive trackers. For details about eye tracking, see Zhai's article in this section. A general survey of eye tracking techniques can also be found elsewhere [4].

Attentive display prototypes have been applied to a variety of visually demanding tasks, spanning a wide range from driving simulators to advertisements and art [4]. In this article, we review five different approaches to degrading the resolution of the screen in peripheral regions. The presented techniques cover a variety of different methods, encompassing customized display hardware and software and range from real-time animation to artistic applications. The first four of the presented displays aim to match the subjective quality of a non-degraded display. To prevent users from noticing the drop in peripheral resolution, i.e., in order to design an *imperceptibly degraded* display, the size of the foveal regions of these displays is designed to at least match the extent of the user's *perceptual span* (see box). In order to do so, these displays use a model of size and resolution of foveal and peripheral vision. The last of the five presented displays, however, does not try to obtain an imperceptibly degraded display. Instead, it quite noticeably removes image content to achieve a different effect; by presenting viewers with only the most important information, it aims at reducing the viewer's cognitive load.

Gaze-Contingent Displays

Our first example is a gaze-contingent display (GCD) [10]. GCDs degrade the resolution of peripheral image regions in order to reduce computational effort during image transmission, retrieval, or display. Fig. 1 shows an example of a movie scene rendered using a GCD. As the user focuses on the face of the shot's main character, all other display content is rendered at reduced resolution, substantially reducing

Figure 1.   Gaze-contingent display showing a scene from the movie The Gladiator. As the user focuses on the face of the shot's main character, all other display content is rendered at reduced resolution. This type of display can be used for gaze-contingent compression purposes or for the study of human visual perception—in this case the display is used to study glaucoma patients. (Original image © 2000 DreamWorks SKG and Universal Studios; gaze-contingent rendering and resolution map courtesy of Bill Geisler and Jeff Perry.)

the rendering effort for this frame. As the movie plays, the high-resolution region moves with the user's focus of attention, so that the spot at the user's focus of attention is always rendered in high-resolution. This effect is achieved by tracking the user's gaze with an eye tracker.

By compressing peripheral image information not resolvable by the user, GCDs help increase display speed. Applications include flight and driving simulators, virtual reality, infrared and indirect vision, remote piloting, robotics and automation, teleoperation, and telemedicine; image transmission and retrieval, and video teleconferencing [10]. In addition to these applications, gaze-contingent displays have been invaluable for the purpose of studying perception, e.g., in order to obtain measurements of the human perceptual span like those presented in the box.

Designing an imperceptibly degraded GCD, i.e., one indistinguishable from a full-resolution display, is desirable but difficult [10]. However, for certain tasks, such as visual search, the reduction in resolution may not necessarily interfere with user performance, even if the peripheral degradation of a GCD is quite noticeable. Given sufficient accuracy in tracking the user's eye movements and a sufficiently fast moving foveal region, a GCD with two resolution regions can moderately degrade the peripheral region, while still producing search performance comparable to a full-resolution display—a foveal region of a 5° viewing angle was found to be sufficient for this purpose [10].

Prior research of gaze-contingent displays has mainly addressed lossy resolution compression of peripheral image re-

gions. New research extends gaze-contingent displays to support arbitrary resolution maps, which allow exploring two new aspects of GCDs [8]. First, this enhancement allows creating foveal regions of arbitrary shape and size with peripheral regions degraded by arbitrary means, e.g., color or contrast reduction, not only resolution. The decoupling of resolution degradation from rendering allows the generation of high quality images with minimal artifacts at real-time display frame rates [8]. Second, this GCD system allows the display of multiple foveal regions at the same time. Multiple foveal regions provide a suitable display strategy for future systems capable of predicting the user's next point of focus.

Gaze-contingent displays have been successfully deployed to save rendering effort [4]. However, while peripheral content is rendered in low resolution, the display hardware on which it is displayed is still the same resolution as any other part of the screen surface. On large screens, where the greater part of the screen surface maps to the user's peripheral vision, this seems especially wasteful of the hardware. This aspect is addressed by focus plus context screens—an attempt to make better use of display hardware.

Focus Plus Context Screens

Focus plus context screens achieve a high-detail/low-detail effect by combining a wall-sized low-resolution display with an embedded high-resolution screen [1]. The installation shown in Fig. 2 uses an LCD inset combined with projection for generating the low-resolution context. The shown version uses a fixed-position high-resolution focus

The Perceptual Span: A Rule of Thumb

High resolution, color vision relies on cone photoreceptors located primarily in the fovea (the retina's small central region). Human ocular physiology places a limit on the range of the perceptual span: fine-grained visual acuity is limited by the dimension of the fovea to roughly 2° visual angle, or about 1-3% of the field of view. A good "rule of thumb" to remember is that the area of high resolution projected onto the fovea is about as large as one's thumbnail at arm's length. At a typical reading distance (30 cm), the foveal region is only about 2 cm in diameter—about as large as the red dot underlying this box. However, although the quality of the information extracted by the visual system declines with eccentricity, the "useful" field of view may extend considerably further, depending on the task. For tasks that require detailed discrimination, e.g., when reading a newspaper, the useful field of view may cover only a fraction of the fovea. For readers of alphabetical orthographies, such as English or French, for example, the span extends from 3-4 letters to the left of fixation to about 14-15 letter spaces to the right of fixation [9].

While reading tasks are comparably well understood, viewing tasks populate a much wider spectrum. Contrary to reading, there is, for example, no particular 'right way' to look at a picture. Context differences are generally at play and viewing behavior and eye movement patterns change as a function of the task. Viewing tasks range widely, from viewing art to performing visual search, such as in target detection, from driving to performing visual inspection or from looking at advertisements to viewing a user interface. For tasks that do not require detailed discrimination, such as visual navigation or detection of large moving objects, the useful field of view may extend to a full 180°. For attentive display design, it is therefore important to consider the dominant viewing task for which the display will be used.

screen; the iconic illustration at the bottom right shows where it is located. The callout shows the difference in resolutions between the focus and the context area. While the focus area offers enough resolution to allow users to see individual cars, the coarse pixels in the context area merely allow seeing larger objects, such as buildings.

In the example shown, the user is inspecting a specific neighborhood on a satellite image of San Francisco. If the user was using a regular-sized monitor showing the same level of detail as the shown setup, only the neighborhood of interest would be visible, without visual context. With residential areas looking very much alike, it would be hard for the user to tell where the shown portion of the satellite image is located within the city, potentially disorienting the viewer. Adding the low-resolution context screen space brings the Bay bridge and the piers into view, providing additional landmarks that simplify orientation. When the user moves the mouse, the entire display content pans, which al-

lows scrolling display content into the focus region in order to make it high resolution.

For tasks involving large maps or detailed chip designs, focus plus context screens were shown to allow users to work from 20 to 35% faster than when using displays with the same number of pixels, but in homogeneous resolution or with multiple views. For an interactive driving simulation, users' error rates were only a third of those in a competing multiple-view setup [1].

In applications that continuously draw the user's attention to the focus area, as is the case for example in the driving simulation used in the experiment, focus plus context screens with a fixed position focus succeed, because the display's focus and context regions cover the user's foveal and peripheral vision the same way a corresponding high-resolution screen does. This makes this type of focus plus context screen, which can be built from comparably inexpensive off-the-shelf components, a cost-effective alternative to complex multi-projector high-resolution screens. By slaving the focus display to the user's gaze, future versions may obtain high resolution wherever the user looks, thereby widening the applicability of focus plus context screens to applications where users continuously look around.
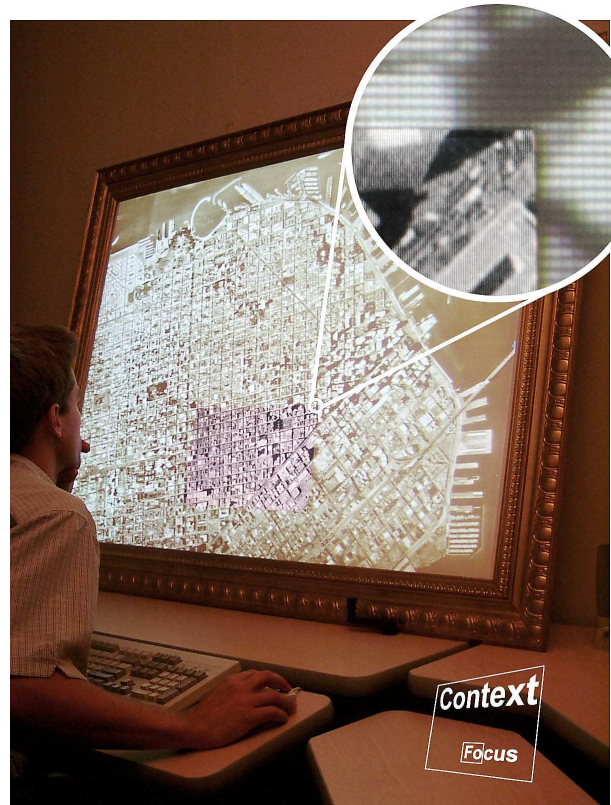

Figure 2. Focus plus context screens complement a monitor-sized high-resolution area in the screen center with a large low-resolution space in the periphery.
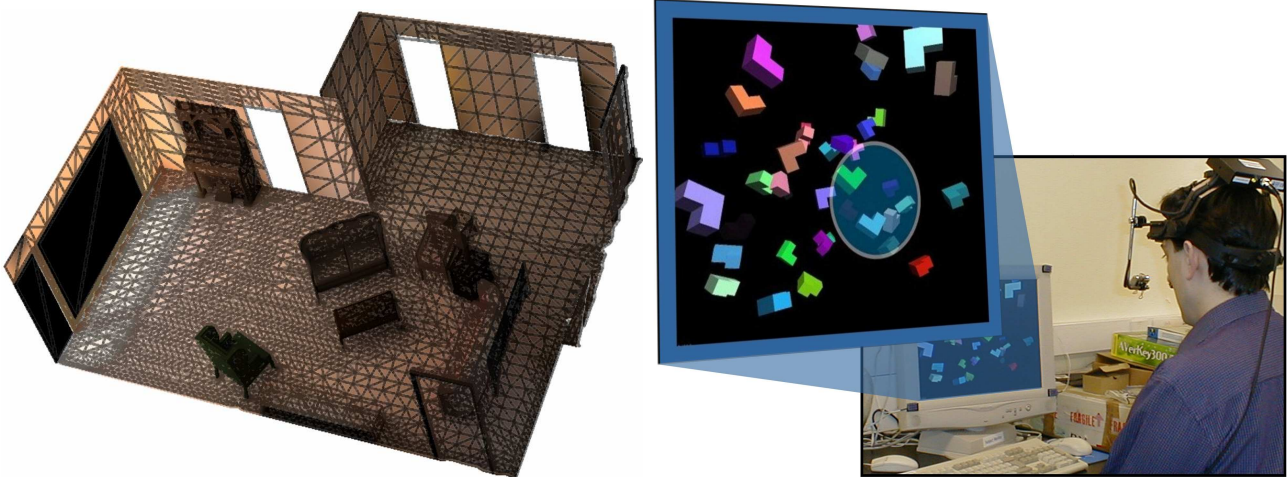
Figure 3. Gaze-contingent spatial and temporal LOD modeling. As the viewer focuses outside the room at the left of the rendering (image at left, courtesy of David Luebke), scene objects located at the right side of the room are rendered using a lower level of spatial detail, indicated by larger triangles (overlayed). Collisions between L-shaped objects (image at right, courtesy of Carol O'Sullivan and John Dingliana) are calculated at a higher level of temporal detail if located within the user's current focus of attention.

Real-Time 3D Graphics

Both GCDs and focus plus context screens degrade peripheral information by manipulating the image (i.e., pixel) properties of the display. In computer graphics research, rendering speed is a primary concern. Interactive applications, such as virtual reality, demand high frame rates in order to satisfy real-time interaction and display. For complex scenes consisting of a large number of polygons, such as virtual terrains containing significant topological detail, or when using computationally expensive rendering techniques such as ray tracing or radiosity, achieving an acceptable combination of surface detail and frame rate requires a substantial hardware effort. Researchers are therefore exploring attentive user interface techniques directing the bulk of system resources towards the scene components delivering the highest perceptual impact. One prominent example of an attentive 3D rendering engine varies the Level of Detail (LOD) at which an object is drawn based on the user's gaze [6]. This way, unattended scene objects are modeled with fewer polygons, even when they are not distant in the scene. Gaze-contingent LOD reduction is similar to the gaze-contingent display in that both techniques reduce the complexity of the displayed image, however, unlike GCDs, graphical methods do so at the object geometry level, rather than at the image level.

Gaze-contingent LOD reduction was found to lead to substantial performance improvements. In the example shown in Fig. 3 (left), a reduction of the number of triangles by 70% still leads to an imperceptibly degraded display [6].

Gaze-contingent modeling has also been applied to real-time temporal resolution degradation [7]. The degradable collision handling mechanism shown in Fig. 3 (right) evaluates object collisions inside the user's focus of attention with greater precision than collisions occurring in the user's periphery. The circle in the callout indicates the field of 4°

visual angle inside which collisions are processed at greater precision. Saving processing time for collisions outside this area allows spending extra processing time on collisions in the user's focus of attention, which results in an overall improvement in the perception of the simulation.

Figure 4. Aggregated fixations from 131 subjects viewing Paolo Veronese's Christ Addressing a Kneeling Woman. Subjects' gaze is drawn to the two main figures. (Original image © National Gallery, London, annotations © IBS, University of Derby, UK, courtesy of David Wooding.)

Easily Perceived Displays

While the approaches described above *follow* the user's attention, attentive displays have also been used to *direct* the viewer's attention, e.g., in the context of art. Artists have long been able to draw viewers' attention to specific artwork

regions. Consider the painting shown in Fig. 4. By controlling luminance, color contrasts, and depth cues, the painter is guiding the viewer's gaze towards the depictions of Christ and the kneeling woman. The artist's success is evidenced in a recent large-scale eye tracking study [11], which shows how only the main two figures in the image were fixated with the remainder of the image left largely unnoticed (see inset). Neuroscientists like Zeki [12] claim this lightens the viewer's perceptual burden, and enables them to look deeper into a piece of art, as the artist has left the viewer with simpler visual inferences to make.

Work in the field of non-photorealistic rendering [3] uses similar techniques to guide the viewer's attention and to allow computer generation of aesthetically pleasing images. Fig. 5 shows an example. This system employs a perceptual model that works from gaze recordings from a single user (see inset) to decide which parts of a photograph should be removed, as eye movement patterns are good indicators for what is important to the viewer [9]. Instead of blurring away detail where the user didn't look, the result is stylized using smooth black lines and colored regions. This produces a rendering that guides the viewers' attention to what the original user found important. This way, the incorporation of one viewer's gaze guides the attention of future viewers.
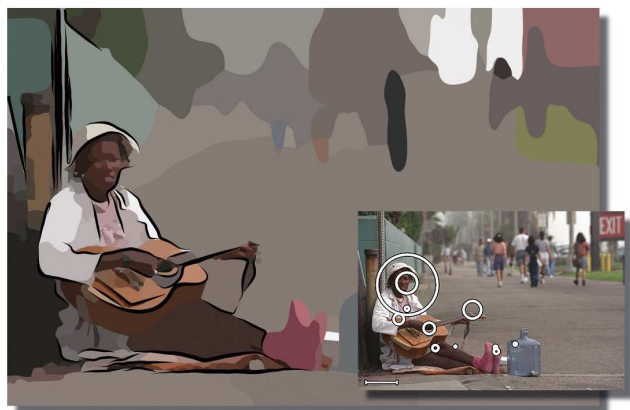

Figure 5. This gaze-based drawing was generated by transforming a photograph based on a user's fixations (inset). (From [3] © 2002 ACM, Inc., courtesy of Doug DeCarlo. Original photo courtesy http://philip.greenspun.com.)

Conclusion

In this article, we presented five examples of attentive displays. All five techniques have in common that they attempt to match the characteristics of computer displays to the characteristics of human vision, namely its distinction between foveal vision and peripheral vision. They all attempt to make better use of limited rendering resources by tailoring display content to the affordances of human vision. The presented techniques differ, however, in which resource they try to preserve and in their adopted strategies for achieving this goal. Each of the discussed techniques falls onto a different point in the spectrum of attentive displays. Gaze-contingent displays, as well as the two presented 3D approaches improve display frame rates and responsiveness given certain rendering hardware; focus plus context screens achieve better immersion and visual context with given display hardware; non-photorealistic rendering, finally, saves maybe the scarcest resource of all—the user's attention.

As rendering and display hardware continue to increase in power and decrease in cost, users will continue to see improved rendering quality on their computer screens. Desktop PCs will be able to display real-time graphics at a quality corresponding to that of today's movies—graphics that today require hours of off-line rendering. As this happens, theater quality graphics will have advanced another step ahead, inching closer to the as yet distant goal of *photorealism*. But in the future, we will also be seeing more kinds of displays, in more places, and for more applications. The most effective use of these displays will seamlessly integrate the requirements of the task and the needs of the user. Despite rapid technological progress, however, users will be limited by their current hardware configuration, no matter what state of advancement it is in. There will always be a desire to stay one step ahead of the current state-of-the-art. The techniques described in this article will offer one possibility of doing so. These considerations suggest that attentive displays will be an enduring figure in the design of interactive computer systems.

References

[1] BAUDISCH, P., GOOD, N., BELLOTTI, V., AND SCHRAEDLEY, P. Keeping Things in Context: A Comparative Evaluation of Focus Plus Context Screens, Overviews, and Zooming. In *Proceedings of CHI '02* (Minneapolis, MN, 2002), ACM, pp. 259–266.

[2] CZERWINSKI, M., TAN, D. S., AND ROBERTSON, G. G. Women Take a Wider View. In *Proceedings of CHI '02* (Minneapolis, MN, 2002), ACM, pp. 195–202.

[3] DECARLO, D., AND SANTELLA, A. Stylization and Abstraction of Photographs. *Proceedings of ACM SIGGRAPH 2002 ACM Transaction on Graphics 21*, 3 (2002), 769–776.

[4] DUCHOWSKI, A. T. *Eye Tracking Methodology: Theory & Practice*. Springer-Verlag, Inc., London, UK, 2003.

[5] HORVITZ, E., AND LENGYEL, J. Perception, Attention, and Resources: A Decision-Theoretic Approach to Graphics Rendering. In *Proceedings of UAI '97* (San Francisco, CA, 1997), Morgan Kaufmann Publishers, pp. 238–249.

[6] LUEBKE, D., REDDY, M., COHEN, J., VARSHNEY, A., WATSON, B., AND HUEBNER, R. *Level of Detail for 3D Graphics*. Morgan-Kaufmann Publishers, San Francisco, CA, 2002.

[7] O'SULLIVAN, C., DINGLIANA, J., AND HOWLETT, S. Gaze-Contingent Algorithms for Interactive Graphics. In *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movement Research*, J. Hyöna, R. Radach, and H. Deubel, Eds. Elsevier Science, Oxford, England, 2002.

[8] PERRY, J. S., AND GEISLER, W. S. Gaze-Contingent Real-Time Simulation of Arbitrary Visual Fields. In *Human Vision and Electronic Imaging* (San Jose, CA, 2002), SPIE.

[9] RAYNER, K. Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin 124*, 3 (1998), 372–422.

[10] REINGOLD, E. M., LOSCHKY, L. C., MCCONKIE, G. W., AND STAMPE, D. M. Gaze-Contingent Multi-Resolutional Displays: An Integrative Review. *Human Factors* (2002). In Press.

[11] WOODING, D. Fixation Maps: Quantifying Eye-Movement Traces. In *Proceedings of ETRA '02* (New Orleans, LA, 2002), ACM.

[12] ZEKI, S. *Inner Vision: An Exploration of Art and the Brain*. Oxford University Press, 1999.

Patrick Baudisch is a research scientist at the Large Display User Experience Group at Microsoft Research in Redmond, WA.
Doug DeCarlo is an assistant professor of Computer Science with a joint appointment in Cognitive Science at Rutgers University, New Brunswick, NJ.
Andrew Duchowski is an associate professor of Computer Science at Clemson University, Clemson, SC.
Bill Geisler is a professor of Psychology at the University of Texas at Austin, Austin, TX.