Eye-Based Interaction in Graphical Systems: Theory & Practice

Andrew Duchowski
Clemson University

Part III
Potential Gaze-Contingent Applications

G Introduction

A wide variety of eye tracking applications exist, each variant increasingly relying on advanced graphical techniques. Eye tracking applications can broadly be described by two categories, termed here as *diagnostic* or *interactive*. In its diagnostic role, the eye tracker provides objective and quantitative evidence of the user's visual and (overt) attentional processes. As an interface modality, the eye tracker serves as a powerful input device which can be utilized by a host of visually-mediated applications.

In general, in their diagnostic capacity, eye movements are simply recorded to ascertain the user's attentional patterns over a given stimulus. Diagnostic applications are distinguished by the unobtrusive use of the eye tracking device. In some cases, it may even be desirable to disguise the eye tracker so that potential subjects are not aware of its presence. Furthermore, the stimulus being displayed may not need to change or react to the viewer's gaze. In this scenario, the eye tracker is simply used to record eye movements for post-trial, off-line assessment of the viewer's gaze during the experiment. In this way, eye movement data may be used to objectively corroborate the viewer's point of regard, or overt locus of attention. For example, studies which test the appearance of some aspect of a display, say the location of an advertisement banner on a web page, may be bolstered by objective evidence of the user's gaze falling (or missing) the banner under consideration. Typical statistical measurements may include the number of fixations counted over the banner during a 5-minute "web browsing" session. Recently, eye tracking has experienced a resurgence in interest in the area of usability testing. Diagnostic eye tracking techniques are applicable (but not restricted) to the fields of Psychology (and Psychophysics), Marketing/Advertising, and Ergonomics and Human Factors.

Equipped with an eye tracker as an input device, an interactive system is expected to respond to, or interact with, the user. Interactive applications are therefore usually expected to respond to the user's gaze in some *gaze-contingent* manner. The archetypical interactive eye tracking application is one where the user's gaze is used as a pointing device. This type of ocular interaction can be considered but one of a set of multimodal input strategies from the system's point of view [Hut93, Nie93, Sch93, Sch93b]. An example of a system relying solely on gaze as input has been shown to be an important communication tool for quadriplegics, where the eyes are used for positioning a cursor over an oversized, projected keyboard. Using gaze to aid communication has also been explored in multi-party computer supported collaborative work systems [Ver99]. Here, gaze was used as a color-coded pointer indicating participants' interest in locations on a shared document. Besides being used as a pointing device, knowledge of the user's gaze may be utilized to alter the display for speed-up purposes, as may be required in the rendering of complex virtual environments [MBD96]. Interactive eye tracking techniques are applicable (but not restricted) to the fields of Human-Computer Interaction, Digital Displays, and Computer Graphics.

H Psychology, Psychophysics, and Neuroscience

A wide assortment of eye tracking applications can be found in Psychology, Psychophysics, and Neuroscience ranging from basic research in vision science to the investigation of visual exploration in aesthetics (e.g., perception of art).

Basic vision science research explores the phenomenon of vision itself. A host of useful factual information has been derived through psychophysical testing (e.g., spatial acuity, contrast sensitivity function, etc.). These types of studies often rely on the display of basic stimuli, e.g., sine wave gratings, but studies are also being conducted on the visual perception of natural scenery [HNK96].

Another application of eye tracking in basic perceptual research involves studying eye movements during reading [Ray92]. Various issues are explored, ranging from the determination of the attentional window (e.g., a few words), patterns of eye movement (left-to-right for English readers), and perception of font types.

Neuroscientific investigation of eye movements includes the use of eye trackers during FMRI recordings and during performance of natural tasks [GT97, BHP95]. These types of studies investigate the cognitive loads during various visual tasks. Eye movement data may be compared with brain imaging data (e.g., FMRI) to identify active functional cortical regions.

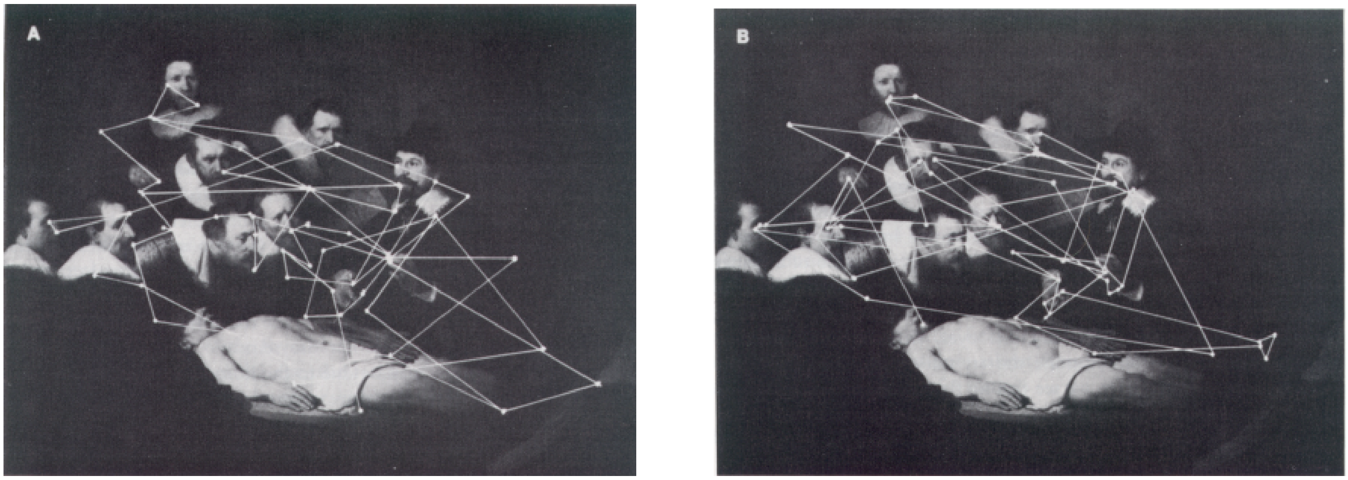
Eye tracking has also been employed in the fascinating study of perception of art [Sol99]. As observed by Yarbus, a viewer's intent influences eye movements and fixations over a scene. Eye movements over art have further refined the scope of these studies, examining differences in how trained viewers search for meaning and aesthetic qualities in fine art pieces. An example of such studies is shown in Figure 48, where small differences in scanpaths can be seen between groups of subjects viewing the artwork for its semantic meaning or for its aesthetic appeal. Remarkably, however, both sets of scanpaths are very similar in terms of fixated image features.

There is a good deal of potential for collaboration between computer scientists, eye tracking researchers, and scientists investigating visual perception. Inevitably, the study of visual perception will require increasingly richer sets of visual stimuli. Providing the capability of recording eye movements over this imagery will undoubtedly extend our knowledge of vision and visual perception.

I Ergonomics and Human Factors

Participants at a recent SIGCHI conference expressed significant interest in employing eye trackers in usability studies [KEJ99]. Indeed, it is believed that eye movements can significantly enhance the observation of users' strategies while using computer interfaces [GK99]. Among various experiments, eye movements have been used to evaluate the grouping of tool icons [GK99], compare gaze-based and mouse interaction techniques [SB00], evaluate the organization of click-down menus [BADM99], and test the effectiveness of electronic map designs for pilots' visual flight rules (VFR) navigation [OHR99].

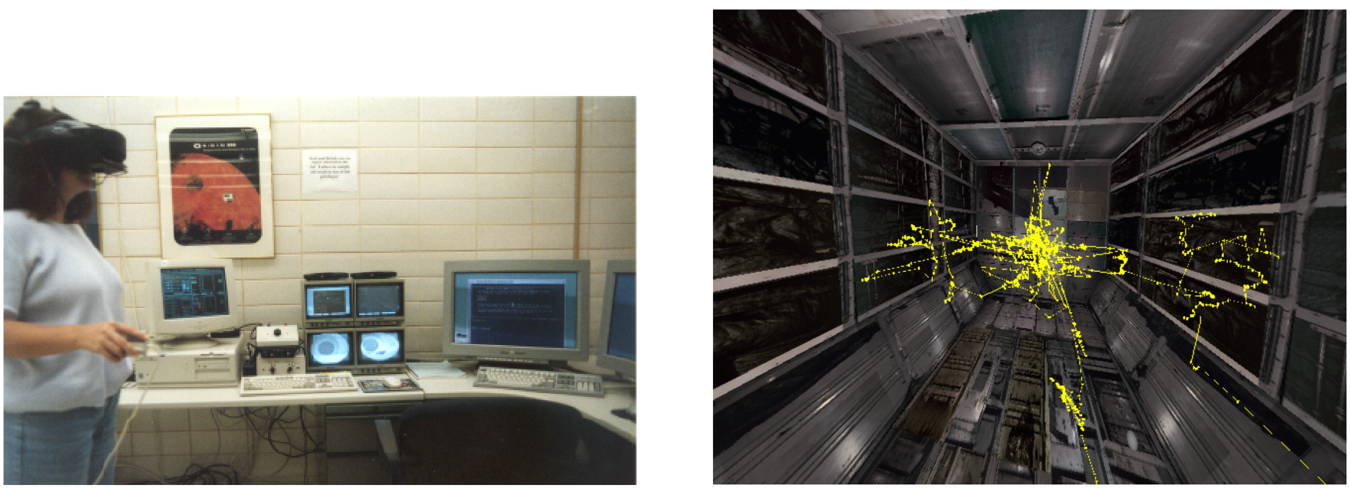
Besides usability issues, eye movements can be used to elucidate visual search patterns of experts during inspection tasks in an effort to



(a) Aesthetic group. (b) Semantic group.

Figure 48: Graph of 60 fixations of two subjects viewing Rembrandt's *Anatomy Lesson*. From [Mol81] (as cited in [Sol99]).

improve training techniques [SGM73, MR79]. In our lab at Clemson University, we are following this line of research within a computer-generated, virtual inspection environment. We use a binocular eye tracker mounted within an HMD to calculate the viewer's gaze in three-space. The environment is a relatively simple geometrical model of the interior of an aircraft's cargo bay in a state ready for physical visual inspection. In the physical environment, the bay's panels are removed to expose the fuselage. Our environment uses polygonal texture maps of images of the fuselage to simplify the underlying geometry. Three-dimensional scanpaths recorded in the simulator are shown in Figure 49(b). Scanpaths in the figure are rendered by joining consecutive gaze/wall intersection points with straight line segments.



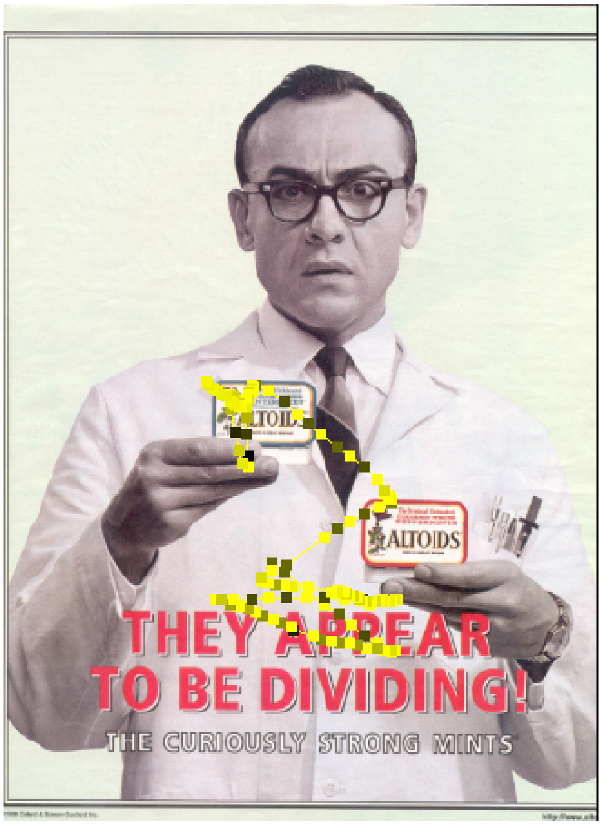
(a) Subject immersed in binocularly eye-tracked virtual environment. (b) 3D scanpaths during simulated inspection. Courtesy of Vinay Shivashankaraiah.

Figure 49: Virtual aircraft cargo-bay inspection environment.

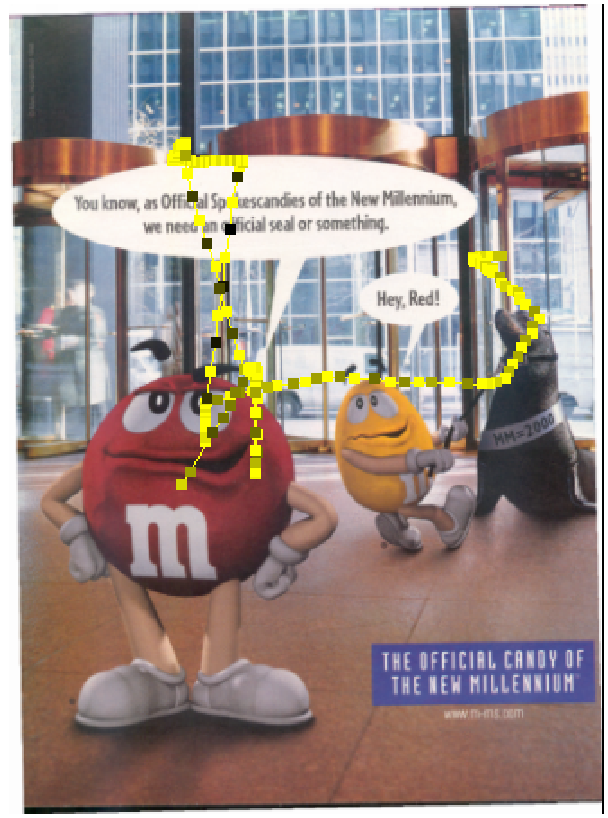
J Marketing/Advertising

Eye tracking can aid in the assessment of ad effectiveness in such applications as copy testing utilizing images, video, or graphics, and in disclosure research involving perception of fine print within print media and within available television and emerging High Definition TV (HDTV) displays. For example, analyses of eye movements over advertisements in the yellow pages disclosed that quarter-page ad displays were much more noticed than text listings, and color ads were scanned more quickly, more often, and longer than black and white ads [Loh97].

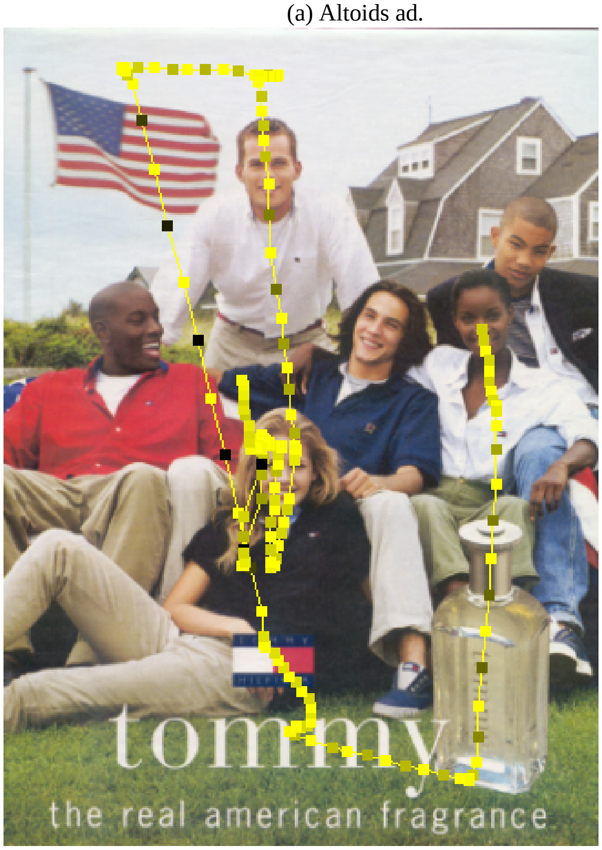
Eye movements recorded over advertisements are particularly informative since scanpaths immediately provide a visual depiction of whether the intended text or object was fixated (or at least scanned over). This was a topic of study of a pair of senior undergraduate students, one from Computer Science, the other from Marketing. Examples of scanpaths recorded over advertisement images are shown in Figure 50.



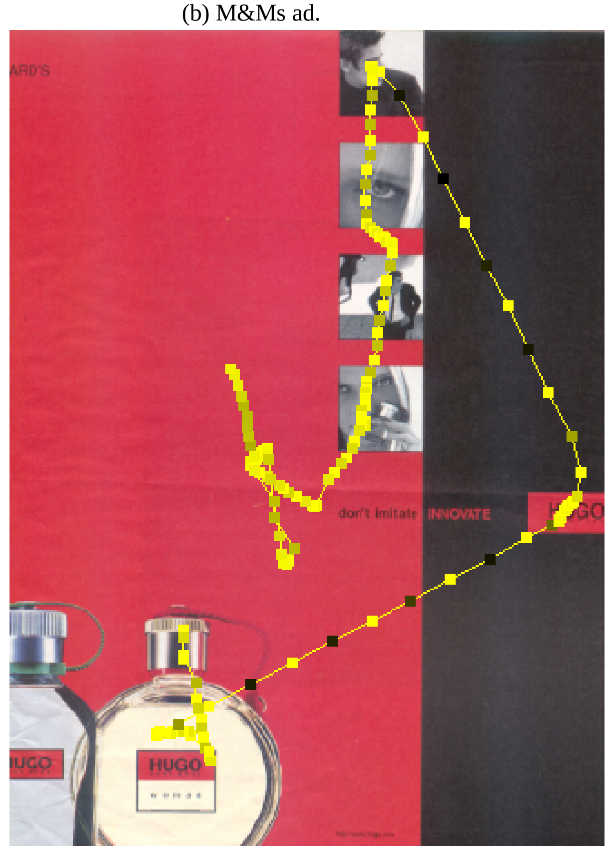
(a) Altoids ad.



(b) M&Ms ad.



(c) Tommy Hilfiger ad.



(d) Hugo ad.

Figure 50: Scanpaths over advertisements. Courtesy of Cristy Lander and Karen Kopp.

K Digital Displays

Human visual perception of digital imagery is an important contributing factor to the design of perceptually-based image and video display systems. Human observers have been used in various facets of digital display design, ranging from estimation of corrective display functions (e.g., gamma function) dependent on models of human color and luminance perception, color spaces (e.g., CIE Lab color space), and image and video codecs. JPEG and MPEG both use quantization tables based on the notion of Just Perceptible Differences to quantize colors of perceptually similar hue [Wal91]. Various gaze-contingent approaches have been proposed for foveal Region Of Interest (ROI)-based video coding [KG96, TEHM96, NLO94, ST94]. Often, however, these studies are based on automatically located image regions, which may or may not correspond to foveally viewed segments of the scene. That is, these studies do not necessarily employ an eye tracker to verify the ROI-based coding schemes. Instead, a figure-ground assumption is often used to argue for more or less obvious foveal candidates in the scene. This is a research area where either diagnostic eye movement studies can be used to corroborate the figure-ground assumption, or gaze may be used directly to display high-resolution ROIs at the point of regard in real-time (as in eye-based teleconferencing systems).

Instead of assuming a feature-based approach to foveal (high-resolution) encoding, an eye tracker can be used to directly establish the foveal ROI and a suitable image degradation scheme may be employed to render detail at the point of regard. This motivated research into finding a suitable image degradation scheme which would match foveal acuity [Duc00]. Using the Discrete Wavelet Transform for smooth image resolution degradation, images demonstrating three acuity mapping functions are shown in Figure 51. For demonstration purposes, the *cnn* image was processed with an artificially placed ROI over the anchor's right eye and another over the "timebox" found in the bottom right corner of the image. Haar wavelets were used to accentuate the visibility of resolution bands. Figure 51(b), (d), and (f) show the extent of wavelet coefficient scaling in frequency space. The middle row shows a reconstructed image where resolution drops off smoothly, matching visual acuity for a particular screen display at a particular viewing distance.

L Human-Computer Interaction (HCI) and Collaborative Systems

Eye-based interactive systems have been presented at several SIGCHI conferences with a significant increase in the number of papers in recent years. One of the first eye-based interactive systems demonstrated an intelligent gaze-based informational display—a text window would scroll to show information on visually selected items—and described an important consideration in eye-based interactive systems: the Midas Touch problem [Jac90]. This is an important problem in systems where gaze is used as an interaction modality. Essentially, if the eyes are used in a manner similar to a mouse, the difficulty arises in determining intended activation of foveated features (the eyes do not register button clicks!). Several notable solutions to the Midas Touch problem were investigated including blinks and dwell times.

At the same SIGCHI meeting, a graphical "self-disclosing" display was presented where story world characters responded in interesting ways to the user's gaze, e.g., visually selected characters would blush [SB90]. At more recent SIGCHI meetings, several novel eye-based interactive systems were described, ranging from gaze-augmented mouse pointing to the use of hidden Markov models to analyze eye movements [ZMI99, Sal99].

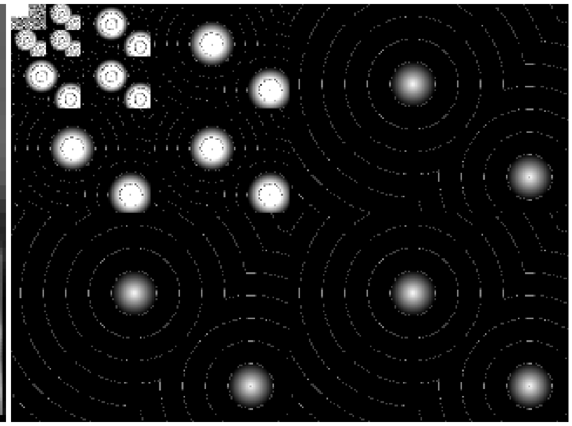
Beside interactive systems, eye tracking can be utilized to aid multiparty communication in collaborative systems [Ver99]. In the GAZE Groupware system, an eye tracker is used to convey gaze direction in a multiparty teleconferencing and document sharing system, solving two problems in multiparty mediated communication and collaboration: knowing who is talking to whom, and who is talking about what. The system displays 2D images of remotely located participants in a VRML virtual world. These images rotate to depict gaze direction alleviating the problem of turn-taking in multiparty communication systems. Furthermore, a gaze-directed "lightspot" is shown over a shared document indicating the users' fixated regions and thereby providing a deictic ("look at this") reference. The system display is shown in Figure 52, with the the system interface shown in Figure 53. For further information, see: <*http://www.cs.queensu.ca/home/roel/gaze/home.html*>.

M Graphics and Virtual Reality

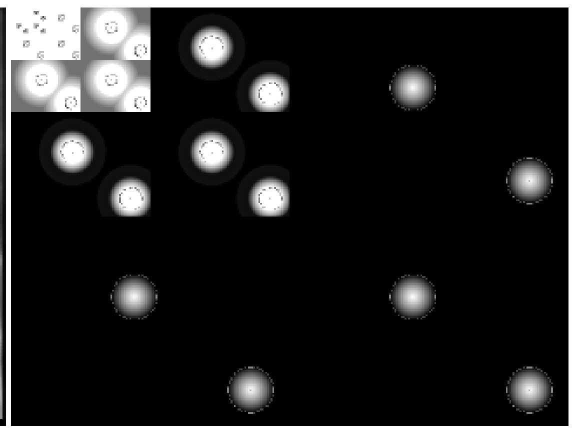
Gaze-contingent interactive graphical displays have been explored with limited success in early flight simulators where emphasis was placed on representing a detailed foveal Region Of Interest (ROI) while homogeneously degrading the periphery [Koc87, LTFW+89]. Early systems used a *display-based* strategy to effectively alter the resolution of the display by filtering peripheral regions. Today's simulators may once again benefit from gaze-contingent strategies where the foveo-peripheral rendering is *object-based*, that is, effectively altering the resolution of individual objects in the scene through Level-Of-Detail (LOD) modeling strategies. A fair amount of work has recently been devoted to LOD modeling [EDDH+95, LKRH+96, MK96, CPDD+96, Hop96, ZSS96, Hop97, ZSS97, SS97]. Although some of these authors address view dependent object representation, it does not appear that any of these techniques have been tested in a true gaze-contingent system (i.e., equipped with an HMD-fitted eye tracker).

For environments containing significant topological detail, such as virtual terrains, rendering with multiple levels of detail, where the level is based on user position and gaze direction, is essential to provide an acceptable combination of surface detail and frame rate [DDGM00]. Recent work in this area has been extensive. Particularly impressive is Hoppe's view-dependent progressive mesh framework [Hop98], where spatial continuity is maintained through structure design, and temporal continuity is maintained by *geomorphs*.

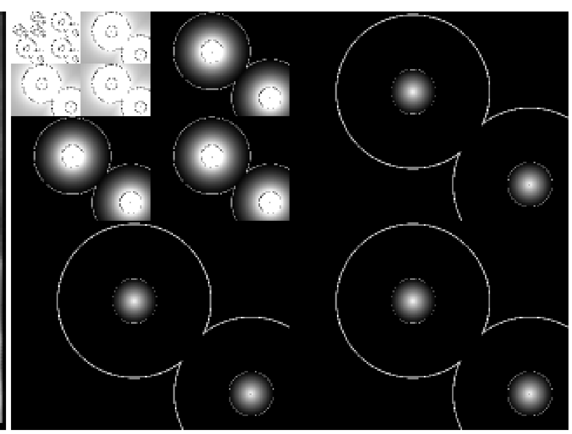
The approach pursued at Clemson is comparatively simple, but still reasonably effective. A surface with significant topological detail is represented as a quadrilateral mesh, which is divided into fixed-size (number of vertices) sub-blocks. Rendering for level of detail is then carried out on a per-sub-block basis. From a fully-detailed surface, lower levels of resolution are constructed by removing half of the vertices in each direction and assigning new vertex values. The new values are averages of the higher resolution values. Resolution level is chosen per sub-block, and is based on viewer distance. The resolution level is not discrete; it is interpolated between the pre-computed discrete



(a) Haar linear mapping. (b) Linear mapping.



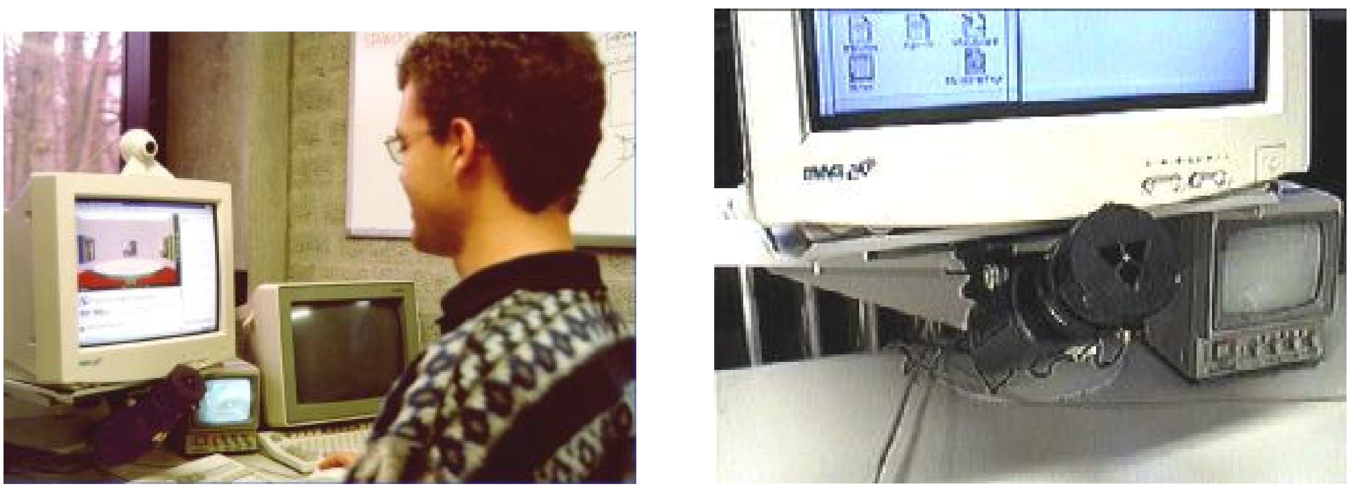
(c) Haar HVS mapping. (d) HVS mapping.



(e) Haar nonlinear mapping. (f) Nonlinear mapping.

Figure 51: Image reconstruction and wavelet coefficient resolution mapping (assuming 50dpi screen resolution). From [Duc00].

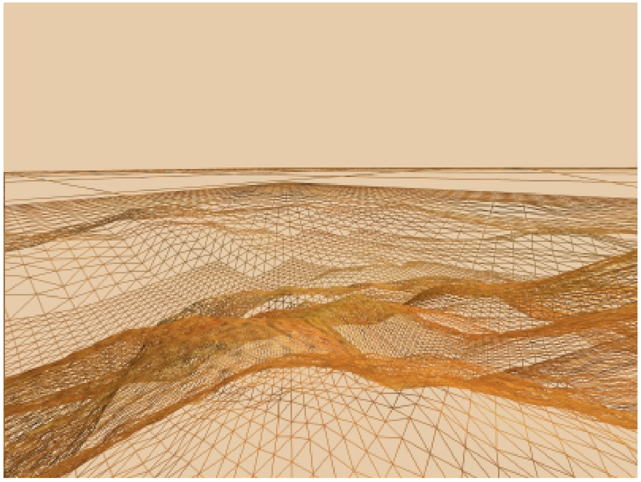
Figure 52: GAZE Groupware display. Courtesy of Roel Vertegaal.



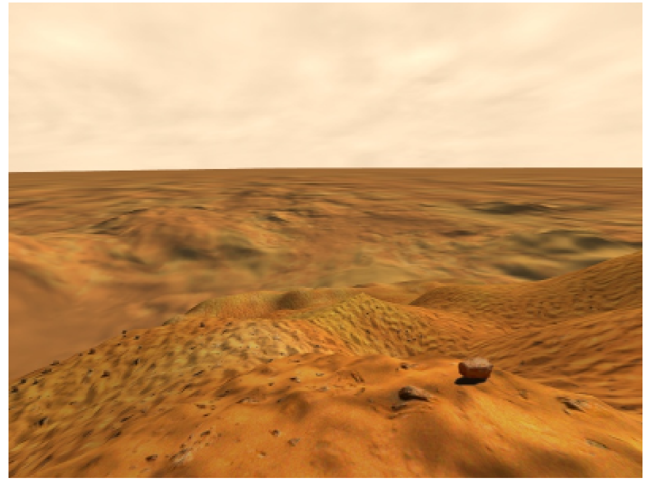
(a) User interface. (b) Eye tracking optics.

Figure 53: GAZE Groupware interface. Courtesy of Roel Vertegaal.

levels to avoid “popping” effects. The terrain, prior to gaze-contingent alteration, is shown in Figure 54. Rocks in the terrain are rendered by



(a) Wireframe.



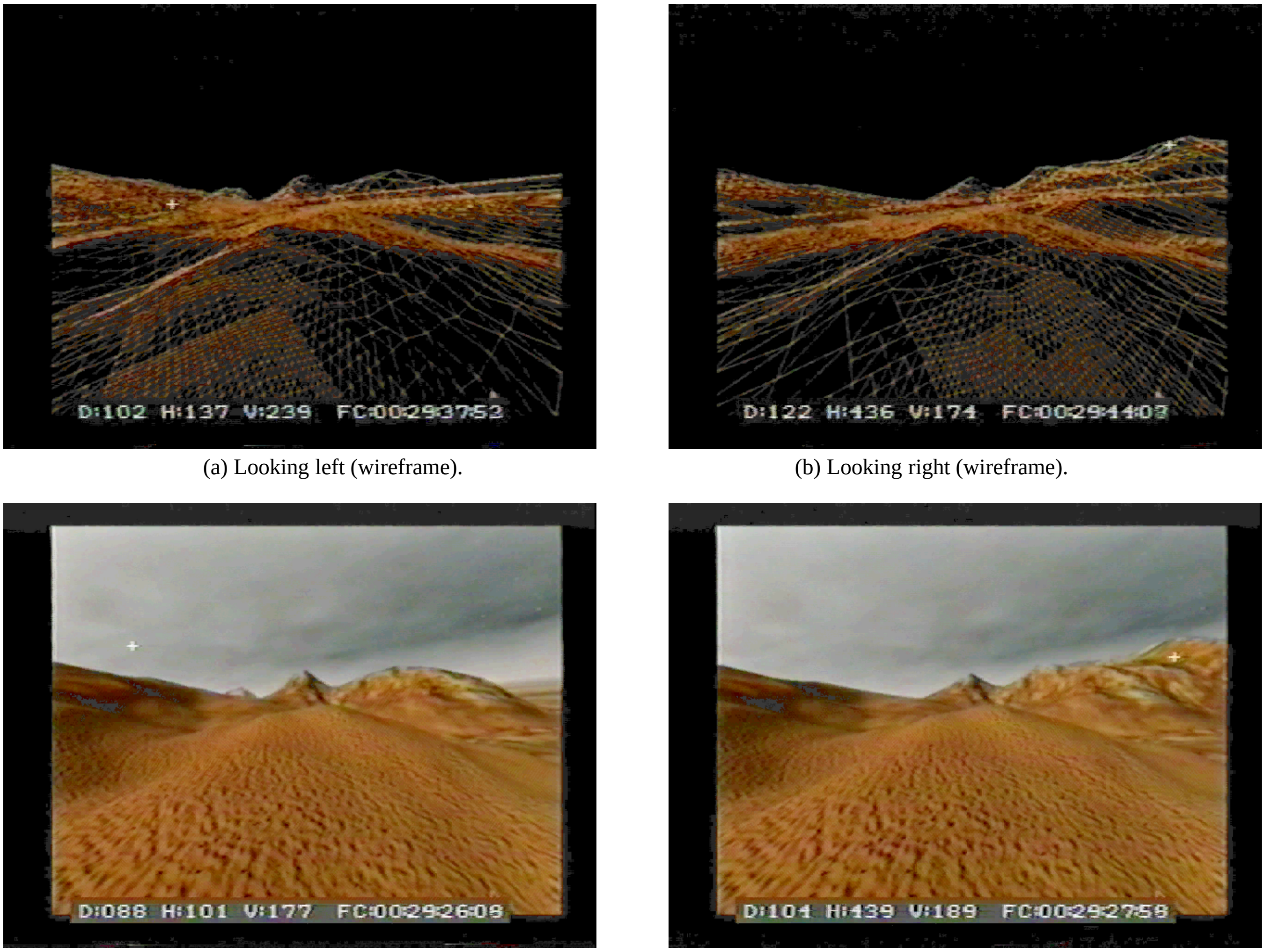
(b) Rendered.

Figure 54: Martian terrain: fractal terrain used for gaze-contingent virtual environment generation. From [DDGM00].

bill-boarding, i.e., images of rocks from the Pathfinder mission to Mars (see: <*http://mars.jpl.nasa.gov*>) were rendered onto 2D transparent planes that rotate to maintain an orientation orthogonal to the viewer. Two views of the gaze-contingent environment (shown rendered and in wireframe) are seen in Figure 55. To exaggerate the gaze-contingent effect, in this environment, fractal mountains appear and disappear from view, based on direction of gaze. Notice also, in Figures 55(a) and (b), the increasd resolution (number of quads) below the gaze vector. The images in the figure are snapshots of the scene images generated by the eye tracker, i.e., what is seen by the operator—the point of regard cross-hair, coordinates, and SMPTE timecode are not seen by the viewer immersed in the environment.

N Summary

A wide variety of (interdisciplinary) eye tracking applications are available, each potentially relying on the incorporation of increasingly sophisticated graphics and imagery. The examples described and shown here by no means constitute a complete survey of the field, but it is hoped that they provide sufficient motivation to spur further interest in this fascinating area.



(a) Looking left (wireframe).

(b) Looking right (wireframe).

(c) Looking left (rendered).

(d) Looking right (rendered).

Figure 55: Martian terrain: gaze-contingent rendering.