Computers & Graphics xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on Serious Games and Virtual Environments

Gaze-based interaction: A 30 year retrospective☆

Q1 Andrew T. Duchowski

School of Computing, Clemson University, 100 McAdams Hall, Clemson, SC 29634, USA

ARTICLE INFO

Article history:
Received 2 January 2018
Revised 6 April 2018
Accepted 9 April 2018
Available online xxx

Keywords:
Eye tracking
Eye movement analysis & synthesis

ABSTRACT

Gaze-based interaction is reviewed, categorized within a taxonomy that splits interaction into four forms, namely diagnostic (off-line measurement), active (selection, look to shoot), passive (foveated rendering, a.k.a. gaze-contingent displays), and expressive (gaze synthesis). Diagnostic interaction is the mainstay of eye-tracked applications, including training or assessment of expertise, and is possibly the longest standing use of gaze due to its mainly offline requirements. Diagnostic analysis of gaze is still very much in demand, especially in training situations such as flight or surgery training. Active interaction is rooted in the desire to use the eyes to point and click, with gaze gestures growing in popularity. Passive interaction is the manipulation of scene elements in response to gaze direction, e.g., to improve frame rate. Expressive eye movement is drawn from its synthesis, which can make use of a procedural (stochastic) model of eye motion driven by goal-oriented tasks such as reading. In discussing each form of interaction, seminal results and recent advancements are reviewed, highlighting outstanding research problems. The survey paper extends an invited proceedings contribution to VS-Games 2017.

© 2018 Elsevier Ltd. All rights reserved.

UNCORRECTED PROOF

1. Introduction

Motivation for this paper was found in the recent inclusion of eye tracking technology in virtual reality headsets. Acquisitions of eye tracking companies Eye Tribe, Eyefluence and SMI by Facebook (Oculus), Google, and Apple, respectively, were notable events. Other eye tracking developments in helmet-mounted displays (HMDs) include the FOVE, and SMI or Pupil Labs add-ons to the HTC Vive. Interestingly, these HMDs are affordable (~$600) compared to what was available some 15 years ago (~$60,000) [1]. Most of these systems, including the one used by the author in 2002, feature binocular eye tracking sampling at 60 Hz or better. New systems sport a larger number of infra-red LEDs, e.g., surrounding each eye, and are more comfortable than the author's 2002 HMD custom-built by Virtual Research and ISCAN.

Head-Mounted Displays only constitute one type of eye-tracked display, typically suggestive of immersive interaction in virtual reality. Currently most of these displays make use of 60–120 Hz eye trackers. While being worn on the head but with the immersive display removed, so-called mobile eye trackers can be used for various augmented reality applications such as examination of navigation in public spaces such as evaluating the utility of signage as aids to wayfinding, e.g., in an airport. These are also typically 60–120 Hz devices. More traditional devices, so-called remote or table-mounted, can offer very fast sampling rates, currently up to 2000 Hz when combined with a chin rest. Generally speaking, eye trackers are usually evaluated in terms of their sampling speed and accuracy, measured in terms of degrees visual angle. Current eye-tracking devices typically boast about 1° visual angle accuracy.

Why has eye tracking suddenly become so popular, or, perhaps more importantly, how is tracked gaze being exploited in virtual reality and other applications? A useful taxonomy for reviewing these applications is shown in Fig. 1, which splits gaze-based interaction into four forms, namely diagnostic (off-line measurement), active (selection, look to shoot), passive (foveated rendering, a.k.a. gaze-contingent displays), and expressive (gaze synthesis).

Diagnostic analysis of gaze, e.g., for assessment of proficiency or training, is mainly performed offline following its recording during performance of some task, often under controlled conditions. Active use of gaze makes use of the real-time (x, y, t) data that eye trackers provide as a streaming signal, similar to the mouse although the eye movement signal is continuous and more noisy that the mouse, which can often show no movement, e.g., when "parked". Active gaze is often meant to effect selection or some kind of command. Passive gaze usually does not imply any specific user action, however, it implies a change to the display in response to gaze movement. Finally, expressive eye movement implies movement of the eyes that is in turn observed by the user, e.g., movement of the eyes of an avatar or virtual character. This type of eye movement can be produced from processed recorded gaze, i.e., data-driven, or it can be synthesized by procedural (e.g.,

---

☆ Originally published in the proceedings of IEEE VS-Games 2017.

E-mail address: duchowski@clemson.edu

https://doi.org/10.1016/j.cag.2018.04.002
0097-8493/© 2018 Elsevier Ltd. All rights reserved.

Please cite this article as: A.T. Duchowski, Gaze-based interaction: A 30 year retrospective, Computers & Graphics (2018), https://doi.org/10.1016/j.cag.2018.04.002

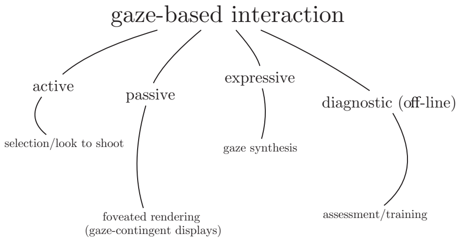

Fig. 1. Gaze interaction taxonomy.

Fig. 2. An update on Yarbus [2], replicating his classic demonstration of task dependency. The painting at left, photographed by the author, is Ilya Efimovich Repin's *Vsevolod Mikhailovich Garshin (1855–1888)*, 1884, Oil on canvas, Gift of the Humanities Fund, Inc., 1972, The Metropolitan Museum of Art, New York, NY. At upper right is raw (unprocessed) eye movement data recorded at 500 Hz by Nina Gehrer, when performing two visual tasks: gauging the emotion of the subject or free viewing. At lower right is the author's visualization of microsaccades depicted in bright yellow within fixations shown as orange discs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Q2

stochastic) algorithms of eye motion. Such models can be driven by goal-oriented tasks such as reading.

Before reviewing the four forms of gaze-based interaction, a short review of eye movement basics offers some nomenclature and characteristics of gaze.

2. Eye movement basics

Detailed human vision is limited to the central 2° visual angle, about the dimension of one's thumbnail at arm's length. Outside of this range, visual acuity drops sharply, e.g., about 50% during photopic (daytime) conditions. High visual acuity within the central 2° is due to the tight packing of cone photoreceptors in the central *foveal* region of the retina. Outside foveal vision, the visual field can be delineated further into *parafoveal* vision (out to about 5°), then *perifoveal* vision (10°), and then *peripheral* vision (all the way out to about 80° on either the temporal or nasal side of each eye). Sundstedt showed a nice depiction of the human visual field in her SIGGRAPH 2010 course notes [3] and subsequent book [4].

Because of the fovea's limited spatial extent (2°), in order to visually inspect the entire 160°–180° (horizontal) field of view, one needs to reposition the fovea along successive points of *fixation*. Most of viewing time (about 90%) is spent in fixations, which is why detection of these eye movements is of particular importance.

Fixations are characterized by tremor, drift, and *microsaccades* which are used to stabilize gaze on the point of interest on the one hand, but keep the eyes in constant motion on the other, so as to prevent adaptation [5]. This is a consequence of the directional selectivity of retinal and cortical neurons implicated in visual perception [6,7]. If the eyes were perfectly still, the visual image would fade from view.[1] Pritchard [9] illustrates the three eye movements carrying an image across the retinal photoreceptor mosaic by curved lines away from the center of vision (slow drift), high-frequency (150 Hz) tremor (superimposed on drift), and straight lines representing microsaccades, the fast *flick* movements back toward the center. The magnitude of all these movements is very small; the diameter of the foveal patch shown is 0.05 mm. Microsaccades have received a great deal of attention, as they have been identified as potential indicators of task difficulty (i.e., cognitive load) [10], mental fatigue [11], emotional attention [12], and perceived threat and anxiety [13], among others. For reviews, see Martinez-Conde et al. [14,15] and Kowler [16].

Note that from an analytical perspective of fixation (or in general event) detection, microsaccades are often seen as signal noise that may be undetectable within the measurement noise introduced by the eye tracker itself [17]. Indeed to detect microsaccades themselves, not only are fast sampling rates required (≥ 300 Hz), but also specialized detection algorithms, with Engbert and Kliegl's [18] being one of the more popular approaches that relies on examination of the median of the eye movement velocity to protect the analysis from noise [19]. An example visualization of detected microsaccades is shown in Fig. 2.

The fovea is repositioned by large jumps of the eyes known as *saccades*. Saccade amplitudes generally range between 1°–45° visual angle (but can be larger; at about 30°, the head starts to rotate [20]). Saccades and microsaccades show comparable spatiotemporal characteristics, suggesting a dynamic continuum, supporting the hypothesis of a common oculomotor generator [21].

When tracking an object, *smooth pursuits* are used to match the motion of the moving target. When fixating an object, the semicircular canals of the inner ear provide signals to counter-rotate the eyes when the head turns—this is known as *Vestibulo-Ocular Reflex*, or VOR. The eyes may also rotate in opposite directions during *vergence* movements; when looking close, the eyes converge, when looking far, they diverge. Vergence eye movements are used for depth perception and are tightly coupled to *accommodation*, the focusing of the eye's lens. Further details can be found in the author's monograph on eye tracking methodology [22].

3. Diagnostic applications

Diagnostic analysis of eye movements generally relies on detection of fixations in an effort to discern what elements of the visual scene attracted the viewer's attention. Note that fixations may themselves be detected by first finding saccades. There are generally two approaches to eye movement event detection: a position-variance approach meant to locate fixations vs. a velocity-based approach generally designed to identify saccades [22]. The sequential pattern of fixations is referred to as the *scanpath* [23]. What is perhaps most relevant is the observation made classically by Yarbus [2]: the pattern of fixations is task-dependent (see also Fig. 2). That is, vision is largely *top-down*, directed by viewing strategy and task demands. However, vision is also *bottom-up*, drawn often involuntarily by eye-catching elements in the scene [24]. Being able to visualize and analyze an expert's strategy, e.g., during inspection or monitoring, is of prime importance to the understanding of expertise. A cogent example lending insight into expertise was given by Law et al. [25] in a virtual laparoscopic training environment: eye movements clearly showed novices fixated on the laparoscope tip while experts, practiced in the tool's manipulation, focused on the target.

---

[1] An impressive simulation of this phenomenon was demonstrated by Mahowald and Mead [8] in the design of a silicon retina based on physiological principles—when held still the image faded.

UNCORRECTED PROOF

Please cite this article as: A.T. Duchowski, Gaze-based interaction: A 30 year retrospective, Computers & Graphics (2018), https://doi.org/10.1016/j.cag.2018.04.002

Download English Version:

https://daneshyari.com/en/article/6876772

Download Persian Version:

https://daneshyari.com/article/6876772

Daneshyari.com