

The Wearable Eyetracker: A Tool for the Study of High-level Visual Tasks

February 2003

Jason S. Babcock, Jeff B. Pelz
Institute of Technology
Rochester, NY 14623

Joseph Peak
Naval Research Laboratories
Washington, DC 20375

ABSTRACT

Even as the sophistication and power of computer-based vision systems is growing, the human visual system remains unsurpassed in many visual tasks. Vision delivers a rich representation of the environment without conscious effort, but the perception of a high resolution, wide field-of-view scene is largely an illusion made possible by the concentration of visual acuity near the center of gaze, coupled with a large, low-acuity periphery. Human observers are typically unaware of this extreme anisotropy because the visual system is equipped with a sophisticated oculomotor system that rapidly moves the eyes to sample the retinal image several times every second. The eye movements are programmed and executed at a level below conscious awareness, so self-report is an unreliable way to learn how trained observers perform complex visual tasks.

Eye movements in controlled laboratory conditions have been studied extensively, but their utility as a metric of visual performance in real world, complex tasks, offers a powerful, under-utilized tool for the study of high-level visual processes. Recorded gaze patterns provide externally-visible markers to the spatial and temporal deployment of attention to objects and actions. In order to study vision in the real world, we have developed a self-contained, wearable eyetracker for monitoring complex tasks. The eyetracker can be worn for an extended period of time, does not restrict natural movements or behavior, and preserves peripheral vision. The wearable eyetracker can be used to study performance in a range of visual tasks, from situational awareness to directed visual search.

1. Introduction

Historically, eye movement literature has concentrated on the mechanics of the eyes in motion. This has provided a rich understanding of the dynamics of the oculomotor system. The top-down (cognitive) and bottom-up (visual processing - starting at the retina) mechanisms responsible for saccadic selection in

scenes have also been studied, but with certain constraints (Fisher *et al.*, 1981; Rayner, 1992). Realistically, what we know about “scene” perception is based on studies involving how people look at two-dimensional images, and to a lesser extent, video sequences. One of the major limitations in these experiments is that subjects’ head movements have been confined by a bite bar and/or chinrest. While stabilizing the head allows for highly accurate eye movement records, in many cases the average fixation duration and saccade length reported from these studies may not be consistent or even comparable with realistic viewing conditions. Eye movements of subjects on a bite-bar and/or chin-rest are vastly different than the visual behavior observed when subjects are free to make both head and eye movements (Collewyn *et al.*, 1992; Kowler *et al.*, 1992). Only recently has the technology been available to study eye movements under more realistic conditions. Land *et al.* (1992, 1997, 1999), Pelz *et al.* (2000, 2001), Canosa (2000), and Babcock *et al.* (2002), have used portable video-based eye trackers to monitor subjects’ eye movements as they perform experiments outside of the laboratory.

Because portable eye-tracking systems have provided novel insight into observers’ visual behavior under more realistic situations, it is clear that the next generation of eye-trackers should be more robust, less obtrusive, and easier to use. The following sections provide some background on physiology of the eye and how the eye is tracked. The remainder of the paper will describe components of RIT’s portable eye tracking system and will discuss its use in research and applications.

2. The Foveal Compromise

Unlike a uniform CCD sensor in a digital camera, the eye’s retina is composed of two types of sensors called rods and cones. These receptors have independent thresholds of detection and allow humans to see over a wide range of conditions. In the periphery of the retina, the rods greatly outnumber the cone photoreceptors. The large rod distribution allows observers to see under low illumination conditions such as those experienced at twilight. Despite the high sampling density, visual acuity in the periphery is quite poor.

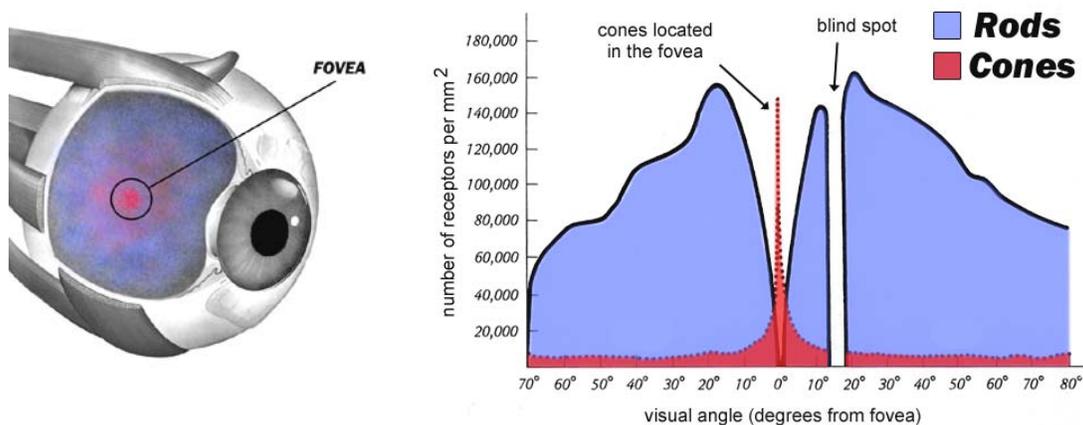


Figure 1- Left, region in the retina called the fovea. Right, number of receptors as a function of visual angle from the fovea. The light shaded region represents rods and the dark shaded region represents cones (right figure adapted from Falk, Brill, and Stork, 1986, pg. 153).

At the center of the eye the cone photoreceptors are distributed in the region of the retina referred to as the fovea (dark shading in Figure 1 - right). Here, high-resolution cone photoreceptors, responsible for color vision, are packed tightly together near the optical axis. From the center outward, the distribution of cones substantially decreases past one degree of visual angle. Unlike the rods, each cone photoreceptor in the fovea reports information in a nearly direct path to the visual cortex. In this region of the brain, the fovea occupies a much greater proportion of neural tissue than the rods (Palmer, 1999, pg. 38). Given these characteristics, detailed spatial information from the scene is acquired through the high-resolution fovea. Since the oculomotor system allows us to orient our eyes to areas of interest very quickly with little effort, observers are typically unaware that spatial acuity is not uniform across the visual field.

At a macro-level the temporal nature of eye movements can be described as a combination of *fixations* and *saccades*¹. Fixations occur when the point-of-gaze has paused on a particular spatial location in the scene. To re-orient the high-resolution fovea to other locations, the eyes make rapid angular rotations called saccades. On average, a person will execute more than 150,000 eye movements a day (Abrams, 1992). This active combination of head and eye positioning (referred to as gaze changes) provides us with a satisfactory illusion of high resolution vision, continuous in time and space. When performing everyday tasks, the point of gaze is often shifted toward task-relevant targets even when high spatial resolution from the fovea is not required. Since these ‘attentional’ eye movements are made without conscious intervention, monitoring them provides the experimenter with an objective window into cognition (Liversedge and Findlay, 2000). While eye movements do not expose the full cognitive processes underlying perception, they can provide an indication of where attention is deployed.

3. Bright Pupil Configuration –Theory of Operation

One common eye tracking technique uses *bright pupil* illumination in conjunction with an infrared video-based detector (Green, 1992; Williams and Hoekstra, 1994). This method is successful because the retina is highly reflective (but not sensitive) in the near infrared wavelengths. Light reflected from the retina is often exhibited in photographs where the camera’s flash is aimed at the subject’s line of sight. This produces the ill-favored “red eye.” Because the retina is a diffuse retro-reflector, long-wavelength light from the flash tends to reflect off the retina (and pigment epithelium), and, upon exit, back-illuminates the pupil. This property gives the eye a reddish cast (Palmer 1999).

Bright-pupil eye tracking purposely illuminates the eye with infrared and relies on the retro-reflective properties of the retina. This technique also takes advantage of the first-surface corneal reflection, which is commonly referred to as the first Purkinje reflection, or P1, as shown in Figure 2 (Green, 1992). The separation between pupil and corneal reflection varies with eye rotation, but does not vary significantly with eye translation caused by movement of the headgear. Because the infrared source and eye camera are attached to the headgear, P1 serves as a reference point with respect to the image of the pupil (see Figure 3). Line of gaze is calculated by measuring the separation between the center of the pupil and the center of P1. As the eye moves, the change in line of gaze is proportional to the vector difference between these points. The geometric relationship (in one-dimension) between line of gaze and the pupil-corneal reflection separation (PCR) is given in Equation 1:

¹ This excludes involuntary microsaccades and visual tremor. This includes various eye movements definitions such as smooth pursuit, nystagmus, VOR, OKN, which are considered to be mechanisms that allow humans to remain fixated on objects that are in motion. Details of these eye movement definitions can be found in Steinman *et. al.* (1990), and Becker (1991).

$$PCR = k \sin(\theta) \quad (1)$$

θ is the line of gaze angle with respect to the illumination source and camera; k is the distance between the iris and corneal center which is assumed to be spherical. In this configuration the eye can be tracked over 30-40 degrees (ASL manual, 1997).

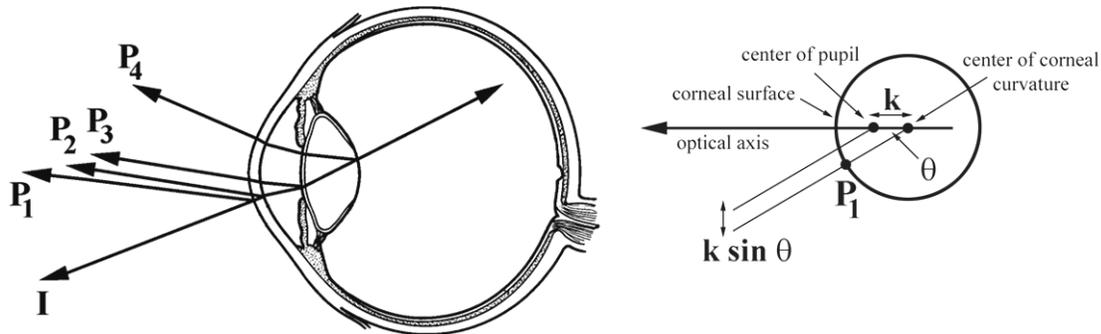


Figure 2 – Right, various Purkinje reflections within the eye. Left, geometry used to calculate the line of gaze using the separation from P1 and the center of the pupil. The cornea is assumed to be spherical (Green, 1992; ASL manual 1997).

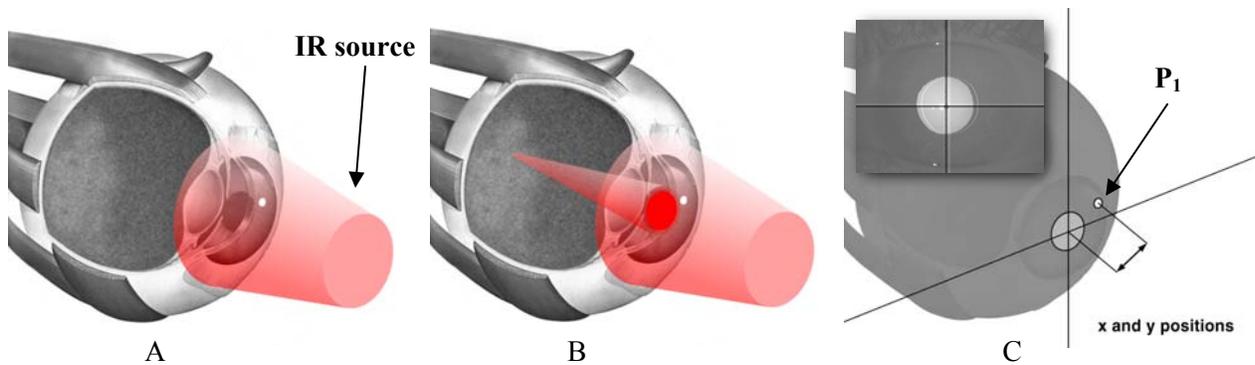


Figure 3 – A) An infrared source illuminates the eye. B) When aligned properly, the illumination beam enters the eye, retro-reflects off the retina and back-illuminates the pupil. C) The center of the pupil and corneal reflection are detected and the vector difference is computed using Equation 1.

4. Overview of the RIT Portable Eye Tracker

4.1 Goggles

The primary component of RIT's portable eye tracker is a pair of modified racquetball goggles as shown in Figure 4. The left side of the goggles supports an optics module with an infrared illuminator, miniature CMOS video camera (sensitive to IR only), and a beam splitter (used to align the camera so that it is coaxial with the illumination beam). An external first-surface mirror folds the optical path toward an infrared reflective mirror shown next to the nose bridge of the goggles. This mirror simultaneously directs

IR illumination toward the pupil and reflects an image of the eye back to the video camera. When aligned properly, the illumination beam enters the pupil, retro-reflects off the retina and back-illuminates the pupil. Figure 4 shows front and top views of the custom goggles.

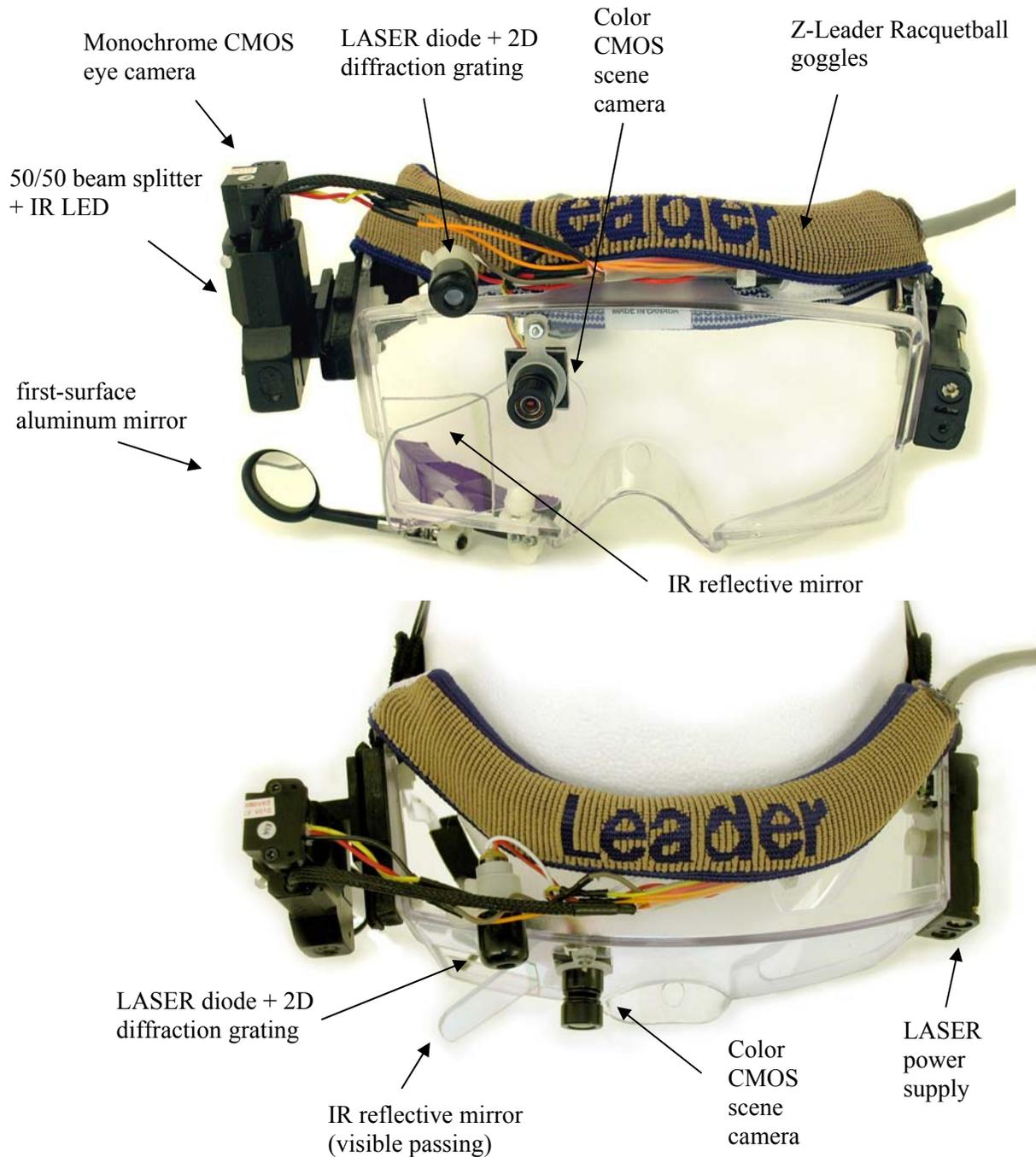


Figure 4 – Front and top views of the RIT portable eye tracking goggles.

5. Goggle Specifications

5.1 Z-Leader Racquetball Goggles

The Vision 2 (product no. RE1003b) racquetball goggles were used in prototype construction since they preserve peripheral vision and offer a sturdy support for mounting components such as the optics module, scene camera, and laser. The goggles are made of high-density polycarbonate plastic with an anti-scratch Silitec coating to prolong optical quality.

5.2 Applied Science Optics Module

Figure 5a shows a close-up of the coaxial LED infrared illuminator/imager sold by Applied Science Laboratories. ASL's monochrome camera has been substituted with a PC51XS CMOS camera for reasons described in the next section. Bright pupil illumination geometry produces the type of image shown in Figure 5b. In RIT's next generation of portable eye trackers, the optics module will be integrated into the goggle frame, allowing for a shorter optical path that will make optical alignment more robust and less sensitive to vibration.

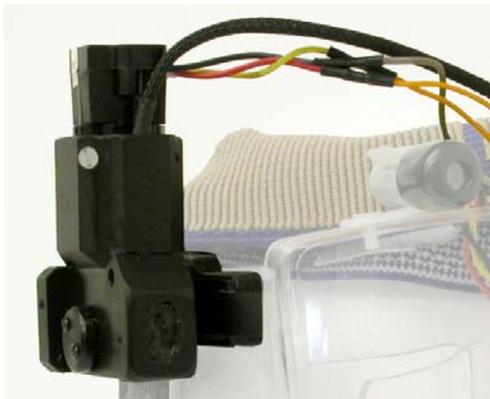


Figure 5a – Close-up of the ASL optics module

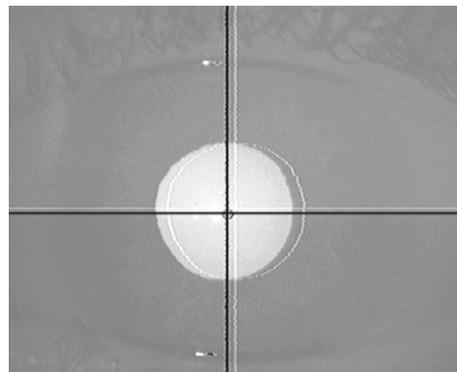


Figure 5b – Typical bright pupil image.

5.3 Monochrome CMOS Eye-camera

Monochrome cameras such as the PC51XS CMOS sensor are ideal for building eye tracking systems because of their small size, low power consumption, and low cost. The PC51XS weighs 1/3 of an ounce, and uses from 6-18 volts DC supply, drawing a current of 30 milliamps at 12 VDC. The sensor size is 1/4" with a resolution of 320 x 240 pixels. These low-resolution CMOS sensors provide an image of the eye that is adequate for threshold and edge detection algorithms employed by ASL's control unit.

5.4 Folding Mirror

A first-surface mirror was attached to a ball and joint arm to allow alignment of the eye image and ASL's optics module (Figure 6a). While the mirror can be translated over a large range, this flexibility can make it difficult to align the optical path quickly and easily for novice users. The problem is analogous to finding an object under a microscope when the magnification of the objective lens is too large. Small translations of the microscope slide result in large shifts of the image. A new design was fabricated to simplify the eye image alignment and is discussed in the next section.



Figure 6a – Ball and joint arm allows for a large range of motion.



Figure 6b – Folding mirror with controlled image alignment.

5.5 Improved Folding Mirror

Because small movements of the folding mirror can result in large displacements of the eye image, it is important to start the adjustment process with the mirror roughly aligned with the optical path. It is necessary to allow for some mirror adjustment so that the optical path can be adjusted to fit observers with different facial features. The ball and joint arm was replaced with a pivoting tension-mounted mirror as shown in Figure 6b.

5.6 Color CMOS Scene-camera

The color PC53XS CMOS color camera was selected because it is one of the smallest commercially available color cameras. It weighs 1/3 of an ounce and consumes 50 milliamps at 12 volts DC. The base of the camera is 0.64" square, with a lens extending to 1.05". The camera is mounted to the face of the goggles as shown in the Figure 4.

5.7 Infrared Reflective Mirror

The infrared reflective (visible transmissive) mirror is placed at a near 45 degree angle normal to the center of the eye. The mirror is made of clear Plexiglas (0.0625") coated with a thin film. The mirror reflects short wave infrared (~700-800 nm) and passes visible wavelengths.

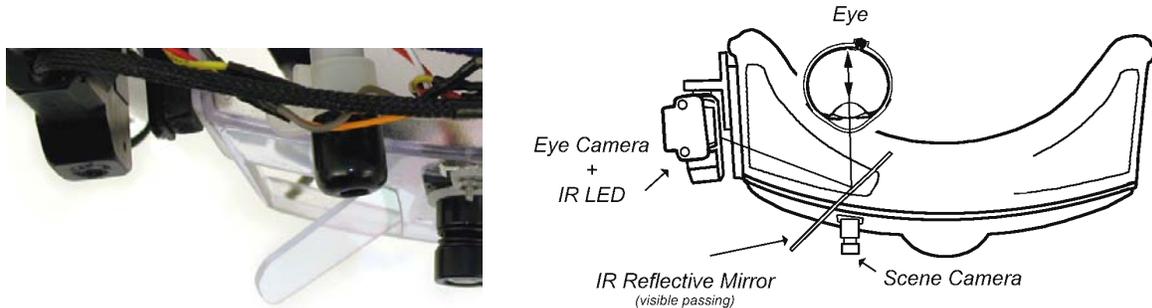


Figure 7 – Illustrates how the infrared reflective mirror is positioned in the goggles.

5.8 Calibrator: LASER and 2-D Diffraction Grating

The laser diode and 2D diffraction grating are used to project a grid of 9-points in front of the person wearing the goggles. The 9-point grid serves as a reference for calibrating the eye position with the camera's scene image. Currently the system uses a fixed-focus laser diode coupled with a 13,500 lines per inch double axis diffraction grating. The figures below show a close-up of the laser module on the goggles and a conceptual projection of the 9-point target. This feature is important because the calibration target always moves with respect to the scene image and can be used to quickly check the accuracy of the track.

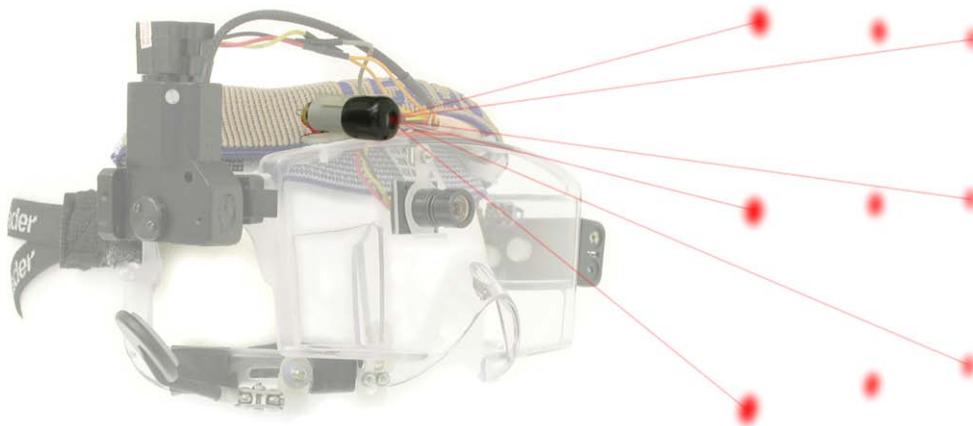


Figure 8 – A laser diode and 2D diffraction grating are used to project a grid of 9-points in front of the person wearing the goggles. This is used to calibrate the observer's eye position with respect to the video scene image.

6. Backpack Specifications

6.1 Backpack

A Camelbak hydration pack is used to carry a customized Applied Science Laboratory (ASL) 501 control unit. Eye and scene video-out from the ASL control unit is piped through an Emerson picture-in-picture unit so that the eye image can be superimposed onto the scene image. The combined video image is then recorded onto a Sony DCR-TRV mini DV camcorder. The camcorder is mounted inside the backpack so that the LCD of the camcorder is visible to the person setting up the eye tracking system.



Figure 9 – Top images show a person wearing the RIT portable eye tracking system. Bottom image shows a diagram of the components carried in the backpack.

6.2 The ASL Control Unit

The Applied Science Laboratories Model 5000 control unit provides the hardware necessary to process the video image of the eye and superimpose the observer's gaze onto the reference scene image. An external computer is used to perform calibration, control parameters such as pupil/corneal reflection thresholds, and adjust the control unit system parameters. Once the calibration has been performed the external computer can be disconnected. Video-out from eye and scene is sent to the picture-in-picture unit to superimpose a small image the eye over the scene (see Figure 10).

6.3 Picture-in-Picture Unit

An Emerson EPP 1800 picture-in-picture is used to superimpose a small image of the eye onto the scene image (shown in Figure 9). This is used to get precise timing information from the eye image and identify blinks and track losses during analysis of the video record.

6.4 Sony DCR-TRV Digital Video Camera

The Sony DCR-TRV mini DV digital video camcorder serves two purposes. The first purpose is to record the video data coming from the picture-in-picture unit. The second purpose is to use the LCD display to set-up the eye image and perform calibration. This particular model was chosen because it has a large (2.25 x 3 inches) LCD. Figure 10 shows the video image as recorded from the DV camera.



Figure 10 – Images show footage from the scene camera with an eye image superimposed in the upper right corner of the screen. Crosshairs indicate observer's point of gaze.

7. Research and Applications

Monitoring observers' eye movements in laboratory settings has proven a valuable method for understanding visual perception, cognition, and the way that perception and cognition guide actions. While those experiments have been very informative, laboratory instrumentation has limited the range of tasks and behaviors that can be studied. The ability to monitor observers' eye movements as they perform natural tasks opens up new fields of research, development, and applications of eyetracking systems.

Portable, robust eyetracking instrumentation will see a range of applications from training to a mobile, flexible human-computer interface. A real-time system could be used to monitor an individual's situational awareness or fatigue state, or as a communication device allowing a group to distribute their pooled attentional resources across a given region. In addition to real-time systems, there are applications for systems that record images of the surrounding scene and the observer's eyes, and calculate gaze off-line in post processing. Advantages of such a system are compact size, ease of use, eliminating the need to calibrate the system before use, and allowing more processor-intensive algorithms to enhance the accuracy, precision, and temporal resolution of the system. The current system is monocular, and therefore can report only a single gaze vector. The distance from observer to the point of regard can only be inferred from the 2-D gaze vector and knowledge of objects and surfaces in the scene. Development of wearable binocular systems will extend eyetracking into three dimensions, allowing vergence eye movements to be monitored as well. Binocular tracking presents significant difficulties because even very small angular errors in monocular eye position extend to large errors in distance; an error of $\frac{1}{2}$ degree in each eye's gaze vector for an observer fixating on a central point 2 meters distant would lead to computed distances from 1.25 to 4.9 m. While some laboratory-based eyetrackers offer sufficient accuracy and precision to allow useful binocular eyetracking, the current generation of video-based eyetrackers does not. Ongoing work on the RIT wearable eyetracker includes efforts to extend the system to be daylight capable, wireless and/or off-line post-processing, and adding binocular capabilities.

In addition to developing applications, robust wearable eyetrackers also open up a new range of research opportunities for studying vision in the context of extended natural tasks – in other words in its natural state. While there are a number of ways to categorize perceptual tasks, a useful structure is to consider whether the scene and/or observer are static or dynamic and whether the subject views the display passively or interacts with the environment. While most eye movement research to date has taken place with static observers passively viewing static scenes, a body of research is emerging that examines more complex, natural behaviors. Experiments with static observers viewing dynamic scenes have shown that observers can make use of experience and expectations about the environment to guide eye movements. Kowler and McKee (1987) and Kowler (1989) demonstrated that gaze fixations on moving targets were influenced by experience with, and expectations about target motion, as well as visual and aural cues discovered to be relevant in previous trials. While such experiments demonstrate the sophistication of the oculomotor control system, the instructed tasks were the eye movements, not a realistic task.

Studying dynamic observers interacting with static scenes has also provided insight into performance. Epelboim *et al.* (1997) studied subjects performing a complex eye-hand coordination task. Observers' eye, head, and hand movements were recorded as they either looked at, or interacted with a three-dimensional pattern display. The results demonstrated that the pattern of eye movements and the coordination of eye, head, and hand interact with the high-level task and goals of ongoing behavior. This new understanding is critical because it calls into question much of our understanding of the oculomotor system – an understanding gained by considering the system in isolation without regard to ongoing tasks. New instrumentation, such as the RIT wearable eyetracker, now permits experiments to be performed

under natural conditions that will lead to a better understanding of how the visual system works when it is used as a tool serving perception rather than as a task itself.

Studies with natural tasks performed by mobile observers interacting with the environment have begun to reveal oculomotor behaviors used in the real world. Land and colleagues used a head-mounted video camera to capture concurrent images of observers' eyes and the surrounding scene. Eye position was later determined by using a trackball to manually locate the pupil centroid in each video frame. Despite the labor-intensive analysis, Land and colleagues have studied subjects' eye movements as they performed a range of complex tasks such as playing cricket and making a pot of tea (1997, 1999). They found that nearly all fixations were related to the immediate task, with eye movements made to reaching targets about ½ second before contact. Land, Mennie, and Rusted reported that only a small fraction (~5%) of the fixations were irrelevant to the task.

Because under natural conditions 'attentional' eye movements are planned and executed at a low level without conscious intervention, monitoring them can reveal visual behaviors that are not available via verbal report. This characteristic provides a powerful tool for understanding and evaluating task performance, and may offer a way to enhance training by revealing the characteristics and behaviors of skilled experts in a number of arenas.

Visual perception is typically considered at the level of behaviors, but when studying vision as a tool in support of natural tasks, emergent perceptual strategies become apparent. Pelz & Canosa (2001) used the RIT wearable eyetracker in a study of natural extended tasks. In most research to date, visual tasks were brief – typically measured in milliseconds. The RIT wearable eyetracker can be used for periods of up to two hours, making it possible to study ongoing behaviors on the order of seconds, minutes, or hours. Pelz & Canosa monitored observers' eye movements as they walked freely in a building, performing a number of active tasks in which they interacted with objects and people. The study revealed that observers typically made brief fixations on regions and objects that were not associated with the immediate task, but would become relevant in the near future. These look-ahead fixations are an example of perceptual strategies that may prove useful in enhanced or artificial vision systems.

As a research tool and an enabling technology, robust wearable eyetrackers will support research and applications development in a range of visual perception tasks from visual search to human-computer interface design.

8. References

Abrams, R.A. (1992). Planning and producing saccadic eye movements. In K. Rayner (Ed.), *Eye Movements and Visual Cognition: Scene Perception and Reading* (p.66). New York: Springer-Verlag.

Applied Sciences Laboratories. (1997). *Eye tracking systems handbook*. Waltham, MA: Applied Science Laboratories.

Babcock, J.S., Lipps, M., Pelz, J.B. (2002). How people look at pictures before, during and after scene capture: Buswell revisited. In B.E. Rogowitz and T. N. Pappas (Eds.), *Human Vision and Electronic Imaging V, SPIE Proceedings*, **4662**, 34-47.

Becker, W., (1991) Saccades, in *Eye Movements*, In R.H.S. Carpenter (Ed.), *Vision and visual dysfunction* (Vol. **8**). Boca Raton: CRC Press.

Canosa, R.L. (2000). *Eye movements and natural tasks in an extended environment*, Master's Thesis. New York: Rochester Institute of Technology.

- Collewijn, H., Steinman, R.M., Erkelens, C.J., Pizlo, Z., van der Steen, J. (1992). Effect of freeing the head on eye movement characteristics during three dimensional shifts of gaze and tracking. In Berthoz, A., Graf, W., Vidal., P.P. (Eds.), *The Head-Neck Sensory Motor System* (Chapter 64). Oxford University Press.
- Epelboim J., Steinman R.M., Kowler E., Pizlo Z., Erkelens C.J., Collewijn H. (1997) Gaze-shift dynamics in two kinds of sequential looking tasks. *Vision Research* **37**: 2597-2607.
- Falk, D., Brill, D., & Stork, D. (1986). *Seeing the light*. New York: John Wiley & Sons.
- Fisher, D.F., Monty, R.A., Senders, J.W. (Eds.). (1981). *Eye Movements: Cognition and Visual Perception*. New Jersey: Lawrence Erlbaum Associates.
- Green, P. (1992). Review of Eye Fixation Recording Methods and Equipment, *Technical Report UTMTRI-92-28*. Ann Arbor, Michigan: The University of Michigan Transportation Research Institute.
- Kowler, E. & McKee, S.P. (1987) Sensitivity of smooth eye movement to small differences in target velocity. *Vision Research*, **27**, 993-1015.
- Kowler, E. (1989) Cognitive expectations, not habits, control anticipatory smooth oculomotor pursuit. *Vision Research*, **29**, 1049-1057.
- Kowler, E., Pizlo, E., Zhu, G., Erkelens, C.J., Steinmann, R.M., Collewijn, H. (1992). Coordination of head and eye during the performance of natural (and unnatural) visual tasks. In Berthoz, A., Graf, W., Vidal., P.P. (Eds.), *The Head-Neck Sensory Motor System* (Chapter 65). Oxford University Press.
- Land, M.F. (1992). Predictable head-eye coordination during driving. *Nature*, **359**, 318-320.
- Land, M.F. & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Phil Trans R Soc Lond*, **B 352**, 1231-1239.
- Land, M.F., Mennie, N., & Rusted, J. (1999) The roles of vision and eye movements in the control of activities of daily living. *Perception*, **28**: 1311-1328.
- Liversedge, S.P. & Findlay, J.M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, **4**(1), 6-14.
- Palmer, S.E. (1999). *Vision Science Photons to Phenomenology*. Cambridge, MA: MIT Press.
- Pelz J.B., Canosa, R.L., Kucharczyk, D., Babcock, J., Silver, A., Konno, D. (2000). Portable eyetracking: a study of natural eye movements. In B.E.Rogowitz and T. N. Pappas (Eds.), *Human Vision and Electronic Imaging V, SPIE Proceedings*, **3659**.
- Pelz, J.B., Canosa, R.L., Babcock, J.S. (2000). Extended Tasks Elicit Complex Eye Movement Patterns, *ETRA 2000: eye tracking research and applications symposium*, 37-43.
- Pelz, J.B., Canosa, R., Babcock, J., and Barber, J., (2001) "Visual Perception In Familiar, Complex Tasks," *ICIP 2001 Proceedings*.
- Pelz, J.B. & Canosa, R.L. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, **41**, 3587-3596.
- Rayner, K. (Ed.). (1992). *Eye Movements and Visual Cognition: Scene Perception and Reading*. New York: Springer-Verlag.
- Steinman, R.M., Kowler, E., and Collewijn, H. (1990). New directions for oculomotor research. *Vision Research*, **30**, 1845-1964.
- Williams, M. & Hoekstra, E. (1994). Comparison of Five On-Head, Eye-Movement Recording Systems, *Technical Report UTMTRI-94-11*. Ann Arbor, Michigan: The University of Michigan Transportation Research Institute.