CIDER: Enhancing the Performance of Computational Eyeglasses

Addison Mayberry Yamin Tun Pan Hu Duncan Smith-Freedman
Benjamin Marlin Christopher Salthouse Deepak Ganesan

University of Massachusetts, Amherst
{amayberr, ytun, panhu, dganesan, marlin}@cs.umass.edu
salthouse@ecs.umass.edu

Abstract

The human eye offers a fascinating window into an individual's health, cognitive attention, and decision making, but we lack the ability to continually measure these parameters in the natural environment. We demonstrate CIDER, a system that operates in a highly optimized low-power mode under indoor settings by using a fast Search-Refine controller to track the eye, but detects when the environment switches to more challenging outdoor sunlight and switches models to operate robustly under this condition. Our design is holistic and tackles a) power consumption in digitizing pixels, estimating pupillary parameters, and illuminating the eye via near-infrared and b) error in estimating pupil center and pupil dilation. We demonstrate that CIDER can estimate pupil center with error less than two pixels (0.6°), and pupil diameter with error of one pixel (0.22mm). Our end-to-end results show that we can operate at power levels of roughly 7mW at a 4Hz eye tracking rate, or roughly 32mW at rates upwards of 250Hz.

Keywords: eye tracking, pupilometry, mHealth, low-power sensing

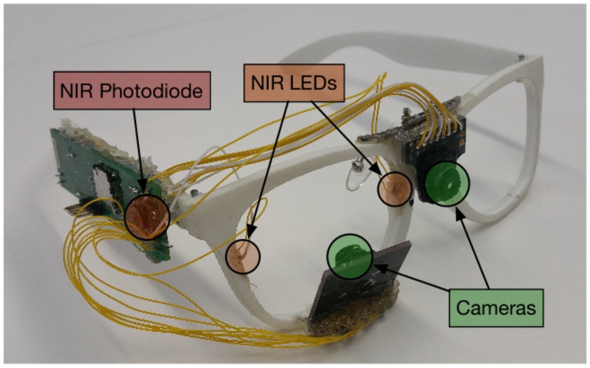
Concepts: •**Human-centered computing** → **Mobile devices;** •**Applied computing** → **Consumer health;** •**Computer systems organization** → *Sensors and actuators;* •**Computing methodologies** → Shape inference;

1 Introduction

Several initial efforts have been made to design low-power wearable eye trackers (e.g. iShadow [Mayberry et al. 2014], iGaze [Zhang et al. 2014]), but many challenges remain. We tackle two in this work — power and robustness. Power consumption is a major avenue for improvement in eye trackers. The iGaze eye tracker consumes 1.5W, and a more optimized eye tracker, iShadow [Mayberry et al. 2014] has a power budget at around 70mW. These numbers are still much higher than typical wearables which only consume a few milliwatts of power, so there is a significant gap that we need to bridge to enable long-term operation of eye trackers on small wearable batteries.

Our fundamental contribution is the design of a staged architecture for computational eyeglasses that can trade off between power and robustness to illumination conditions. The principle underlying our architecture is well-known to systems researchers — we optimize heavily for the common case but provide more power-hungry features to deal with the more difficult but uncommon scenarios that occur.

Figure 1: *The CIDER platform prototype.*



The common case is that a) we spend a substantial fraction of time indoors (homes, shopping malls, etc), and b) we spend 80% of the time fixating on points, during which time the eye moves only a small amount (referred to as microsaccades, which are typically less than 0.4°). We optimize for this regime by using a small amount of near-infrared illumination, a few tens of pixels sampled per estimate, and a few dozen instructions executed per pixel to estimate eye gaze and pupil dilation parameters. The power consumption for the common case is, therefore, only about 7mW — in contrast, iGaze [Zhang et al. 2014] consumes 1.5W (three orders of magnitude difference), and iShadow [Mayberry et al. 2014] consumes 70mW (order of magnitude difference).

One of the interesting auxiliary benefits of our staged processing pipeline is that it can operate at very high frame rates during typical operation. Our optimized pipeline can operate at rates exceeding 100 fps, which is at the the high end of tethered remote eye trackers. This capability is particularly useful for detecting small fine-grained saccadic movements which happen while reading or when fatigued, providing further window into an individual's neural activities. Our algorithm, CIDER (CIrcle Detection of Edges with Reinforcement), is the first wearable eye tracker to achieve such high frame rates.

Our experiments show that

- CIDER can track pupil center with accuracy of roughly 1 pixel (0.3°) and pupil dilation with accuracy of approximately 1 pixel (0.22mm) in indoor lighting conditions.
- We operate end-to-end at a total power budget of 7.5mW when running at 4Hz, which is 10× less than previous state-of-art in this area [Mayberry et al. 2014]. Alternatively, we can achieve eye tracking rates of upwards of 250 frames/second by scaling power consumption up to 32mW.

ETRA '16, March 14-17, 2016, Charleston, SC, USA
ISBN: 978-1-4503-4125-7/16/03
DOI: http://dx.doi.org/10.1145/2857491.2884063

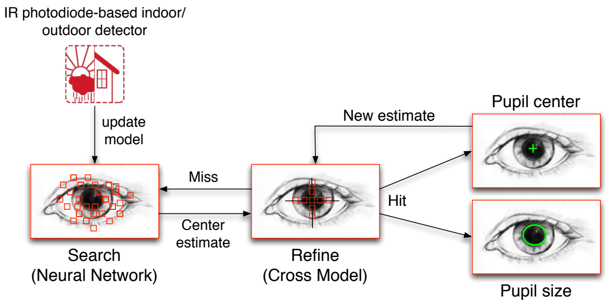


Figure 2: *The CIDER pipeline: a) search stage using a neural network to get an initial estimate of pupil location, b) refine stage to zone in on exact pupil center and perform rapid tracking unless the pupil is missed, and c) NIR-photodiode-based detection of indoor/outdoor mode to update neural network model.*

2 Design Overview

At a high level, CIDER uses two different approaches to trade off between robustness and power, as seen in figure 2 — the first is a two-stage rapid eye tracking controller, and the second is indoor-outdoor model switching to deal with different illumination conditions and noise. CIDER relies on a "Search–Refine" two-stage controller and a small amount of NIR illumination of the eye to estimate eye parameters in a fast, efficient, and accurate manner.

The search stage operates with no prior knowledge of pupil location, and uses a neural network to obtain an estimate of pupil center and size from a sub-sampling of pixels. The refine stage takes an estimate from the search stage, and uses a very fast and accurate procedure to locate and track the eye. When the refine stage loses track of the pupil due to specular reflections or other unpredictable variations, it reverts to the search stage to get another estimate of the eye location. The two stages differ in terms of the amount of sensing required (i.e. number of pixels acquired per frame from the imager) as well as the amount of computation performed to extract eye parameters from the pixels.

Overall our pipeline achieves a graceful tradeoff between robustness and power — under typical indoor illumination, CIDER spends most of its time in the fastest and most efficient stage while occasionally using the neural network to provide estimates. In outdoor illumination, CIDER spends all of its time in the slower but more robust neural network stage.

CIDER also addresses robustness issues by designing a model training pipeline that operates with no input from the user. Whenever the eyeglass is fully charged or has good connectivity, a block of images can be communicated to the phone, and a new model trained offline. The enabler is an offline image processing pipeline that generates accurate labels of pupil center and dilation from noisy image data.

3 Performance

We present a brief evaluation on data collected from 16 users, 12 male and 4 female. Each subject performed a video calibration routine where they looked at a high contrast dot moving on a computer monitor for several minutes, generating approximately 2500 eye images per user. We used this as our dataset for testing the CIDER algorithm.

Energy savings: Figure 3 shows the aggregate power consumption of CIDER and compares against two other baselines - using only the Search stage (ANN), and an idealized model that represents that best possible accuracy of the refine stage alone. We see similar trends as we saw earlier in that CIDER operates in between the idealized cross and ANN model with roughly a 3× reduction (compared to neural network models that have low error). The overall power budget for CIDER is roughly 7mW, which is a huge improvement over state-of-art (order of magnitude less power consumption than [Mayberry et al. 2014]), and a substantial achievement considering that the system is operating a camera, estimation algorithm, and NIR LED.

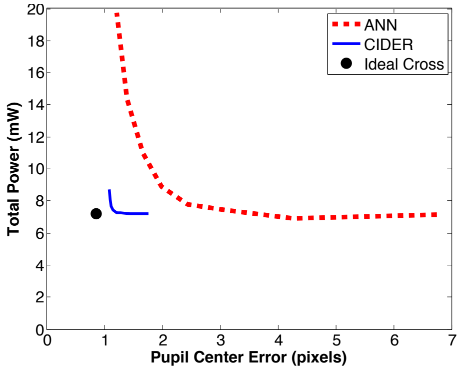


Figure 3: *Aggregate power vs accuracy*

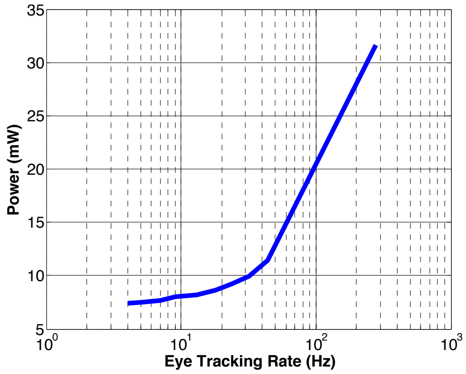


Figure 4: *Aggregate power vs eye tracking rate (log scale)*

Power vs tracking rate: Another benefit of CIDER is that it can achieve high tracking rates. We plot the power vs pupil tracking rate in Figure 4, which shows the total system power consumed as the tracking rate is varied. To generate this graph, we inserted sleep periods of variable length between each single execution of the CIDER pipeline. The measurements were taken using a DAQ sampling at 10kHz.

References

MAYBERRY, A., HU, P., MARLIN, B., SALTHOUSE, C., AND GANESAN, D. 2014. ishadow: design of a wearable, real-time mobile gaze tracker. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*, ACM, 82–94.

ZHANG, L., LI, X.-Y., HUANG, W., LIU, K., ZONG, S., JIAN, X., FENG, P., JUNG, T., AND LIU, Y. 2014. It starts with igaze: Visual attention driven networking with smart glasses. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, ACM, 91–102.