Addison Mayberry *University of Massachusetts Amherst*
Yamin Tun *Cisco* **Pan Hu** *University of Massachusetts Amherst*
Duncan Smith-Freedman *Boston Children's Hospital*
Deepak Ganesan and Benjamin M. Marlin *University of Massachusetts Amherst*
Christopher Salthouse *Draper Labs*

Editors: Robin Kravets and Nic Lane

THE "I" IN THE EYE

Excerpted from "CIDER: Enabling Robustness-Power Tradeoffs on a Computational Eyeglass," from *Proceedings of the 21st Annual ACM International Conference on Mobile Computing and Networking* with permission.
http://dl.acm.org/citation.cfm?id=2790096 © ACM 2015

The eye is a truly unique and fascinating piece of physiology. Through it, organisms take in massive amounts of sensory data; through it, they also communicate emotion and intent both consciously and unconsciously. It is so tightly coupled with the operation of the brain that even slight mental blocks like being tired are measurable through eye movements. Eye-related research has long worn out the cliché "the eyes are the window to the soul," and this is largely because there is a deeper truth behind this phrase that has come to light in the last several decades of psychology and neurology research. In short, we now understand that the eyes are the best window to the mind. (Kahneman 2011)

FIGURE 1. A team member wearing the iShadow glasses

It goes without saying that being able to emulate, understand, and even manipulate the way that humans behave and engage with the world is a worthy goal for computer science researchers of most flavors. This is obviously in large part due to species selfishness, since humans are the designers, operators, and the primary beneficiaries of computing machinery. Even when we set aside that self-interest and consider the issue from a purely scientific standpoint, the brain is a massively powerful computation engine that has been honed by millions of years of evolution. To be able to better understand its operation related to one of the most complex and useful tools in our species' possession – our sightedness – would be a significant step forward for science. As we discover that algorithms emulating the neurons of the brain are particularly effective for object recognition and other vision tasks (Szegedy, Toshev and Erhan 2013), one wonders how much we could accelerate those efforts by having a more complete understanding of the most effective vision engine known to man.

We shouldn't overlook the "selfish" aspect in all of this, though. While the complexities of the brain's operation have resisted a comprehensive understanding and will likely continue to do so for a long time to come, there is much lower-hanging fruit to be had. Specifically, the systematic observation and analysis of how we use our eyes in everyday situations could potentially lead to enormous benefits for public health. Even without understanding the visual cortex to its fullest, we can observe very consistent patterns in how the eyes move that allow us to infer a person's cognitive functions – i.e., what they're thinking about and how they're processing it. Armed with this knowledge, we can go a step further and engage with the user either by providing them useful information for their task or redirecting their attention elsewhere (e.g., reminding a distracted driver to look at the road). This kind of "augmented living" is one of the major goals of the mobile health (mHealth) research community, and we believe that a key step towards that goal is being able to unobtrusively and continuously instrument the human eye, a concept commonly referred to as "eye tracking."

Another principal goal of the mHealth movement is early diagnosis of illnesses of all forms. In this regard, eye tracking opens up new possibilities that have traditionally been difficult or impossible to measure directly in a non-lab setting. The most straightforward candidates are illnesses directly relating to the eyes – lazy eye, glaucoma, etc. However, the intimate connection between the eyes and the brain means that many health conditions related to mental functions can be diagnosed from the eye. It has been shown by medical researchers that a wide range of such conditions can be detected through eye tracking. The list includes relatively simple effects such as fatigue, but also includes more severe issues including ADHD (Fried 2014), autism (Schmitt 2014), and even Alzheimer's disease (Molitor 2015).

EYE TRACKING IN THE WILD

Nothing discussed so far is particularly eye-opening (pardon the pun) to those who are familiar with this area of work. The uses of eye tracking have been known or at least suspected for decades or more, and since the mid-1900s there have been many devices built for quantifying eye movements (see Hanson 2010 for an excellent overview). Contemporary devices boast near-perfect accuracy using conventional camera technology. Such tools have facilitated a huge step forward for our understanding of the human visual system.

The Achilles' heel of traditional eye tracking tools, however, is that they are restricted to a laboratory environment. They must be carefully deployed and calibrated in order to function properly; the subject must be brought indoors and seated for the duration of data collection. Obviously, such technologies are not of great utility to the mHealth effort since a tool cannot be used for mobile health if it has no mobility. Being forced to remain in the lab means that it is not possible to build systems for interventional or assistive purposes and it restricts researchers from doing any kind of longitudinal, in-situ studies that would be of great benefit for the purposes previously described. Lastly, even medical diagnosis is hindered by this restriction – like a heart arrhythmia, aberrant eye patterns indicative of mental health issues may arise only sporadically and not be caught during a brief examination or recording period. Continuous monitoring improves the chance of a prompt diagnosis.

The natural response to this problem is to build a mobile device that subjects can carry with them. Major vendors of eye-tracking tools have been building wearable systems since the early 2000s. These systems are almost always head-mounted and normally take the shape of eyeglasses. However, these devices have been beset by a number of severe engineering challenges.

Scaling down the technology has proven challenging – high-end eye tracking systems rely on high-definition cameras with specialized, consistent lighting. Collecting and processing the data from such cameras requires a high amount of secondary resources – digital storage for potentially gigabytes of raw video data, a powerful computer for running the eye tracking algorithm, and a power supply for these devices and the cameras themselves. In short, to do effective mobile eye tracking traditionally requires the subject to be wearing the equivalent of a desktop computer.

While the industry has made great strides forward in the technical domain, even the most modern wearable eye trackers are relatively bulky headsets wired to a battery pack and a smartphone. Even with this setup, they typically run down the battery in two to four hours, which is not suitable for longitudinal studies of behavior over the course of a day. Power issues aside, such devices are obtrusive and very obvious when being worn, making it unreasonable to have people wear them in normal environments (home, work, etc.) for extended periods due to social considerations.

ISHADOW: LIGHTWEIGHT, LOW-POWER EYE TRACKING

We decided to tackle the problem of building a wearable eye-tracking device from the opposite direction. Instead of trying to miniaturize existing systems, we opted to design a new regime for doing eye tracking based on a platform built with the following key design goals:

1. **Unobtrusiveness:** In order to be useful as a practical wearable device for studies and other in-the-wild deployments, the tracker must have minimal impact on the normal behavior of the wearer. This generally means a small form factor and little or no extra equipment to be carried.

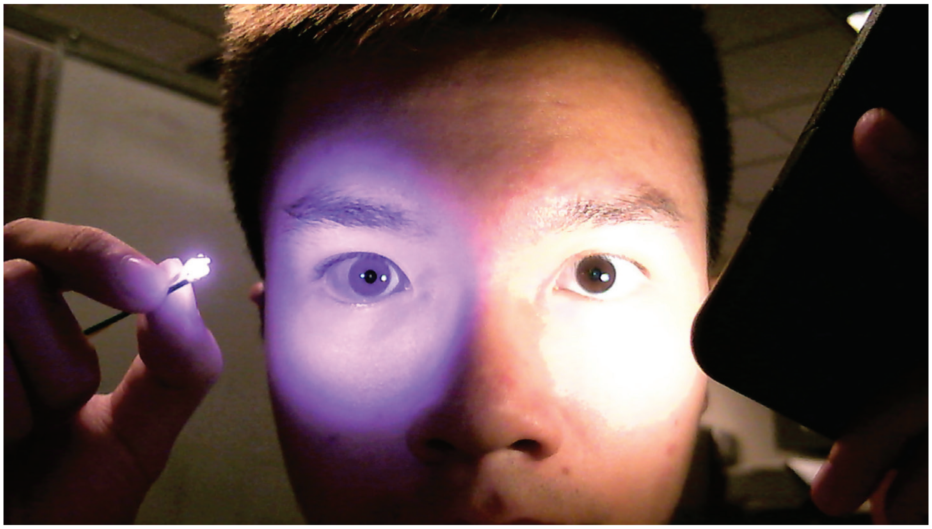

FIGURE 2. A near-infrared light (left), which is invisible to the human eye, highlights the pupil and the iris very clearly compared to a normal light source (right).

2. **Performance:** The device must be able to track the eyes accurately enough and at a high enough rate to be useful for the types of studies and interventions that researchers in the field care about.
3. **Longevity:** An ideal system would be able to operate for as long as the wearer is awake without needing to be recharged, so as to be able to record all waking activity while adding minimal burden to the user.

The major theme underlying all of these goals is power consumption – a device that requires less power will last longer and will have a smaller footprint. These goals led us to build the iShadow "computational eyeglass" platform, pictured in Figure 1. Our evaluation of the system on 10 subjects showed that the system had an error rate of 3° for gaze prediction and a max power consumption of 70 mW (A. P. Mayberry 2014).

Two key innovations work hand-in-hand to make this platform possible. Both stem from the insight that the major bottleneck of eye-tracking systems is the pixel acquisition – the cost of reading and digitizing the voltage output by the camera's individual pixels. The time and power cost of acquiring a pixel is defined by the hardware being used, so we chose to focus our efforts on reducing the total number of pixels acquired.

The first key innovation was the change from the high-rate, high-definition cameras used by traditional eye trackers to a small, low-resolution grayscale camera called the Stonyman, produced by Centeye, Inc. In addition to being small and operating at very low power, the Stonyman cameras allow for pixels to be accessed individually instead of row-by-row as in most traditional cameras. This feature facilitates algorithms that only require a subset of the camera's total pixel grid.

Designing such an algorithm was the second key innovation. We designed a small artificial neural network model that we could train to predict a user's gaze location based on images of the eye. In order to reduce the number of input pixels (features) needed and thus realize a power savings, we applied L1 regularization to the neural net. In brief, this technique forces the model to discard those inputs that it deems least useful for the classification task in favor of those that are the most valuable for performing the task at hand.

This neural network model is small enough to be run on the embedded processor onboard the glasses. We take advantage of the Stonyman's random-access pixel feature and only sample the pixels specified by the neural network, saving power proportional to the number of pixels left unsampled. In our experiments we found that we discard 80% or more of the imager pixels, realizing an approximate power savings of the same magnitude.

Getting started with the iShadow

The iShadow glasses have to be calibrated for each new user one time, as the on-board neural network needs training data for the specific individual. This is a relatively fast process and, once completed, does not need to be done again for that subject. Here's a brief overview of the steps involved.

1. Collect training video
The subject is fitted with the glasses and they are set to constantly record the wearer's eye motions to a video. In order to ensure that the video has good coverage of the entire range of eye positions, we provide a tool that displays a rapidly moving target on a monitor. We recommend that subjects be seated in front of the monitor for two to three minutes while recording their eye motions.

2. Label video for training
Once the video is collected, it must be labeled for the supervised neural network training. We have built tools that facilitate rapid labeling of the position of the pupil in the video frames, in order to make the process straightforward.

3. Train the neural network parameters
After all of the labels have been generated, the video data and labels are passed into the neural network trainer. This can be located on a server or local desktop machine for speed.

4. Upload parameters
As soon as the neural network training is finished, the researcher can upload the parameters of the trained network to the onboard memory of the iShadow unit. As soon as this is done, the glasses are ready to begin eye tracking for that subject.

CIDER: ENHANCING THE PERFORMANCE OF COMPUTATIONAL EYEGLASSES

While our initial results showed promise, they still did not measure up to the performance of traditional eye-tracking devices, nor the ultra-low power requirements of true wearable devices. In order to push the performance of the iShadow platform even further, we revisited the problem of pixel acquisition.

Near-infrared (NIR) illumination can be used to make the distinction between separate regions of the eye – especially the pupil and iris – much clear to a camera device, as shown in Figure 2. By

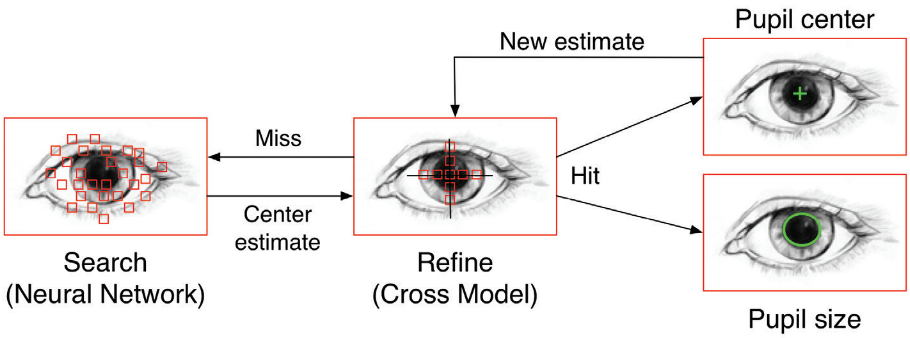

FIGURE 3. The CIDER algorithm pipeline: The neural network and cross model together form a two-stage controller for tracking the eye.

including NIR illuminators in our device we significantly improved the quality of the images at minimal power cost, which opened up new possibilities in terms of efficient tracking algorithms requiring even fewer pixels than our original technique.

We implemented a multi-stage tracking algorithm as shown in Figure 3. The first stage is the familiar sparse neural network model, which calculates an approximate location for the center of the pupil with reasonable accuracy. This approximation is passed to the next stage, a computer-vision-based technique that samples a single row and column from the Stonyman imager. It localizes the edges of the pupil within these regions and uses those to estimate the pupil position and dilation more accurately than the neural network. By passing the pupil position forward into the next iteration, this "cross model" is able to naturally track the movements of the eye at a fraction of the cost of the neural network, requiring only a few hundred pixels to be sampled.

It is possible that the cross model could lose track of the pupil, due to blinks or other occlusions. In this case the neural network is used to re-localize the pupil position and bootstrap the process once more. In this way the robust neural network model mitigates the brittleness of the faster and cheaper cross model to provide excellent tracking performance at low average power consumption.

In a new evaluation with 16 subjects, we demonstrated that the iShadow glasses running the CIDER algorithm can track a person's pupil with an error of only 0.6° at a power draw of as little as 7 mW (A. Y.-F. Mayberry 2015). Thus, our system provides performance close to that of industrial tracking devices while meeting the ultra low–power standards of modern wearable devices. We hope that the iShadow will pave the way to new ways to study the human mind and new opportunities for augmented health and living. ■

Addison Mayberry is a graduate student in the College of Information and Computer Sciences at the University of Massachusetts, Amherst. He received a B.S. in Electrical Engineering from Grove City College in 2011. His research interests include wearable medical devices, embedded sensing, and low-power computer vison.

Yamin Tun received her B.S. in Engineering Science from Smith College in 2014, and a Master's in Computer Science from the University of Massachusetts, Amherst in 2016. She works as a software engineer in the IoT team at Cisco.

Pan Hu is a graduate student in the College of Information and Computer Sciences at the University of Massachusetts Amherst. He received a bachelor's degree in Electrical Engineering from Wuhan University of Technology in 2012. His research interests include low-power wireless systems, mobile computing and general computer networks.

Duncan Smith-Freedman received his B.S. in Electrical and Computer Engineering from the University of Massachusetts, Amherst in 2015. He works as part of the Boston Children's Hospital Simulator Program in preparation for medical school.

Deepak Ganesan is a professor in the College of Information and Computer Sciences at the University of Massachusetts, Amherst. He received his Ph.D. in Computer Science from UCLA in 2004. His research focuses on ultra-low power wireless communication via backscatter, novel platforms and algorithms for mobile and wearable health sensing, learning and inference on multi-modal sensor data, and micro-powered sensors.

Benjamin M. Marlin is an assistant professor in the College of Information and Computer Sciences at the University of Massachusetts, Amherst, where he co-leads the Machine Learning for Data Science Laboratory. His research focuses on the development of machine learning models and algorithms motivated by applications in clinical and mobile health, including wearable sensing systems.

Christopher Salthouse received bachelor's, master's, and doctoral degrees from the Massachusetts Institute of Technology. He now works at Draper Labs as the lead for the Mixed Signal Integrated Circuit Group.

REFERENCES

[1] Fried, Moshe, Eteri Tsitsiashvili, Yoram S. Bonneh, Anna Sterkin, Tamara Wygnanski-Jaffe, Tamir Epstein, Uri Polat. "ADHD subjects fail to suppress eye blinks and microsaccades while anticipating visual stimuli but recover with medication." Vision research 101 (2014): 62-72.

[2] Hansen, Dan Witzner, and Qiang Ji. "In the eye of the beholder: A survey of models for eyes and gaze." IEEE Transactions on pattern analysis and machine intelligence 32, no. 3 (2010): 478-500.

[3] Kahneman, Daniel. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.

[4] Mayberry, Addison, Pan Hu, Benjamin Marlin, Christopher Salthouse, and Deepak Ganesan. "iShadow: design of a wearable, real-time mobile gaze tracker." *Proceedings of the 12th annual international conference on Mobile systems, applications, and services (MobiSys)*, 2014.

[5] Mayberry, Addison, Yamin Tun, Pan Hu, Duncan Smith-Freedman, Benjamin Marlin, Christopher Salthouse, and Deepak Ganesan. "CIDER: Enabling Robustness-Power Tradeoffs on a Computational Eyeglass." *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking (MobiCom)*, 2015.

[6] Molitor, Robert J., Philip C. Ko, and Brandon A. Ally. "Eye movements in Alzheimer's disease." *Journal of Alzheimer's Disease* 44, no. 1 (2015): 1-12.

[7] Schmitt, Lauren M., Edwin H. Cook, John A. Sweeney, and Matthew W. Mosconi. "Saccadic eye movement abnormalities in autism spectrum disorder indicate dysfunctions in cerebellum and brainstem." *Molecular Autism* 5, no. 1 (2014): 1.

[8] Szegedy, Christian, Alexander Toshev, and Dumitru Erhan. "Deep neural networks for object detection." *Advances in Neural Information Processing Systems (NIPS)*, 2013.