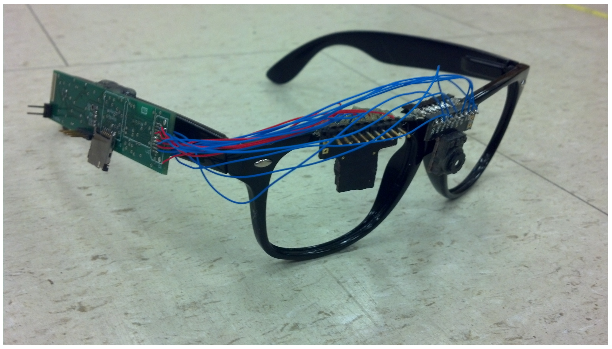
iShadow: The Computational Eyeglass System

Addison Mayberry, Pan Hu, Benjamin Marlin, Christopher Salthouse,* Deepak Ganesan†
University of Massachusetts

Figure 1: *The iShadow platform prototype.*



Abstract

Continuous, real-time tracking of eye gaze is valuable in a variety of scenarios including hands-free interaction with the physical world, detection of unsafe behaviors, leveraging visual context for advertising, life logging, and others. While eye tracking is commonly used in clinical trials and user studies, it has not bridged the gap to everyday consumer use. The challenge is that a real-time eye tracker is a power-hungry and computation-intensive device which requires continuous sensing of the eye using an imager running at many tens of frames per second, and continuous processing of the image stream using sophisticated gaze estimation algorithms. Our key contribution is the design of an eye tracker that dramatically reduces the sensing and computation needs for eye tracking, thereby achieving orders of magnitude reductions in power consumption and form-factor. The key idea is that eye images are extremely redundant, therefore we can estimate gaze by using a small subset of carefully chosen pixels per frame. We use a sparse pixel-based gaze estimation algorithm that is a multi-layer neural network learned using a state-of-the-art sparsity-inducing regularization function which minimizes the gaze prediction error while simultaneously minimizing the number of pixels used. Our results show that we can operate at roughly 70mW of power, while continuously estimating eye gaze at the rate of 30 Hz with errors of roughly 4 degrees.

CR Categories: J.3 [Computer Applications]: Life and Medical Sciences—Health;

Keywords: neural network, eye tracking, gaze estimation, mHealth

*salthouse@ecs.umass.edu
†{amayberr,panhu,marlin,dganesan}@cs.umass.edu

ETRA 2014, March 26 – 28, 2014, Safety Harbor, Florida, USA.

ACM 978-1-4503-2751-0/14/03

1 Introduction

We present iShadow, a novel ultra-low-power "computational eyeglass" system for performing eye tracking-related research in the wild. Many existing eye tracking systems are designed for use in a controlled testing environment [Hansen and Ji 2010]. These tools provide extremely high accuracy and image resolution, at the cost of requiring subjects to be brought into a controlled environment for collecting data. The results generated by researchers in the growing field of mobile-health-related research, or mHealth, clearly demonstrate the benefits of collecting physiological and environmental data on subjects in the wild. Given the well-known research opportunities afforded by eye and gaze tracking, it is very apparent that leveraging mHealth techniques to facilitate collecting of gaze data on subjects outside of a lab environment would open up new avenues of research that were previously infeasible.

There are already tools in existence that are designed to meet this need, however, existing tools suffer from several limitations. The largest of these issues are mobility and obtrusiveness. Current mobile eye trackers require the use of auxiliary equipment for data transmission or recording, for energy supply, and for control. In addition, even with these extended modules they can only run for a few hours at a time [Tobii 2013]. These tools represent a huge step forward in eye tracking for mHealth, since they bring eye research out of the lab and into the wild, but their limitations will continue to restrict researchers. Thus, we have designed a mobile eye tracking tool with an emphasis on ultra-low power consumption. In addition, the iShadow platform is designed to be programmable, providing computational resources that enable real-time processing and fusion of sensor data. This is, to our knowledge, a novel feature in the space of mobile eye tracking tools.

The key feature that enables our system design is the use low-power embedded imagers that provide a random-access pixel interface [Stonyman 2013]. This allows the processor to read the values of any individual pixel, which facilitates subsampling and the use of adaptive sampling methods. Our platform also includes a number of additional sensors for gathering data on personal and environmental context. Together, all of these features enable the iShadow platform to provide a unique power-accuracy tradeoff so that the most energy and computation can be used for collecting and processing more data during the most relevant periods of time. This can be accomplished using adaptive sampling methods and data triggering. In addition, iShadow operates strictly in the visible spectrum using passive illumination, removing the need for a power-intensive light source of any kind.

2 Design Overview

In this section, we provide a brief overview of the working of iShadow. The first step in using iShadow is calibration, where a user looks at a few points on a monitor while keeping their head relatively steady, in a manner similar to commercial eye trackers. During this calibration phase, iShadow captures a full image stream from the eye-facing and outward-facing imager, and downloads this data to a computer either via USB or Bluetooth.

The second step is the neural network based sparse pixel selection algorithm. In this stage, the learner divides the calibration dataset into training and testing sets, and sweeps through the regularization parameters to learn a set of models that correspond to different gaze prediction accuracies. Each model specifies both the set of pixels that need to be acquired and the weights on the pixels to use for the activation function that predicts gaze co-ordinates. Depending on the power constraints of the platform and the accuracy needs of the application, the appropriate model can be downloaded to the iShadow platform for real-time operation.

The third step is the run-time execution of the model that is downloaded onto the iShadow platform. The model is stored in an SD card when it is too large for the available memory, and the run-time system acquires the appropriate pixel set and executes the non-linear weighted sum to predict gaze co-ordinates in real time.

3 Performance

To test the effectiveness of iShadow at accurately and efficiently predicting the wearer's gaze location, we collected sample data from ten different subjects. We generated at least five minutes of labeled data for each subject in full-image-capture mode, and the resulting dataset includes at least 3000 images per user. We used this data to perform experiments testing iShadow's effectiveness.

Power-Accuracy Tradeoff. One of the key benefits of our algorithmic framework is that it is able to provide a variety of models that offer different tradeoffs between the overall complexity of the model (i.e. number of pixels sampled, and number of weights for computation) and the accuracy of the model (i.e. the precision in degrees). This tradeoff is enabled by using different choices of the regularization parameter, λ, which is set by the user during model generation and varies the penalty for model complexity.

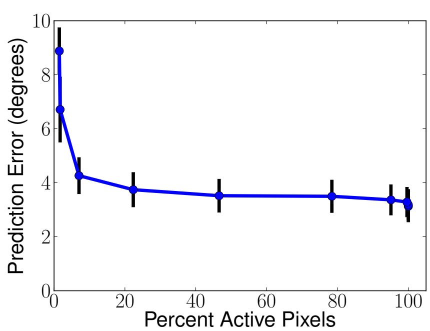


Figure 2: *The number of pixels acquired can be reduced dramatically (up to 10×) with minor effect on gaze prediction accuracy.*

Figure 2 shows the prediction accuracy vs model size — interestingly, we see that varying the percentage of activated pixels from 100% down to about 10% has little to no effect on the prediction accuracy. This shows that there is substantial redundancy in the eye image, and the neural network is able to predict gaze just as well with 10% of the pixels activated as 100% of the pixels. On our imager, this means that sampling 10K pixels per image vs sampling 1K pixels per image has roughly the same prediction error, which in turn can translate to substantial reduction in power consumption. This data is averaged over the results from all of the users in our study, and demonstrates that iShadow is able to predict gaze location with an accuracy of roughly 4 degrees at 10% of the pixels used (and therefore 10% of the power consumption).

Calibration.

We have shown that the ANN-based frame can learn accurate models with few pixels, but how much calibration data is needed to train these models? To evaluate this, we look at how quickly the gaze prediction converges as the amount of data that we use for training increases. Our goal is to minimize the calibration time so as to decrease the burden on the person undergoing calibration.

Figure 3 shows the results for a particular choice of the regularization parameter λ. We see that the convergence is very fast — even if there is less than 30 seconds of data used for training, that is more than sufficient for the algorithm to determine the appropriate parameters. Similar results were seen for other values of λ. Thus, the time for calibration is not a bottleneck in system operation.

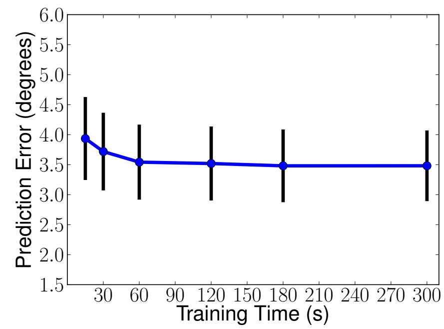


Figure 3: *Amount of training time for the system versus the resulting gaze prediction accuracy.*

4 Conclusion

We present a first-of-its-kind low power gaze tracker that is designed to predict gaze in real-time while operating with a power budget of a few tens of milliwatts. Our approach exploits the unique properties of random access pixel cameras to achieve a flexible energy-accuracy trade-off in the wearable/real-time setting. Our results show that we can dramatically reduce power consumption and resource needs by sampling only 10% of pixel values, without compromising accuracy of gaze prediction. These results are highly significant in that they offer a clear path toward ubiquitous gaze tracking for a variety of applications in computer vision, behavioral sensing, mobile health, and mobile advertising.

Acknowledgements

We would like to thank Russ Belawski and Prabal Dutta, who have been instrumental in getting earlier prototypes of our platform up and running and bootstrapping initial data collection. Also, we thank Jeremy Gummeson for much needed help with the final hardware prototype.

References

HANSEN, D. W., AND JI, Q. 2010. In the eye of the beholder: A survey of models for eyes and gaze. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 32*, 3, 478–500.

STONYMAN, 2013. Stonyman Vision Chip. `http://centeye.com/products/stonyman-vision-chip-breakout-board/`.

TOBII, 2013. Tobii Eye Tracker. `http://www.tobii.com/en/eye-tracking-research/global/products/hardware/tobii-glasses-eye-tracker/technical-specifications/`.