InSight: Seeing the World Through Your Own Eyes

Russ Bielawski, Joe Romeo, Justin Paupore, Pat Pannuto, Prabal Dutta
Electrical Engineering and Computer Science Department
University of Michigan
Ann Arbor, MI 48109
{jbielaws,jromeo,jpaupore,ppannuto,prabal}@umich.edu

ABSTRACT

For a long time, cameras have been power hungry, and this was doubly so, compounded by their complexity, requiring power-hungry processors to capture incoming data. However, significant advances in low-power technology will soon make these concerns a thing of the past. We lay out the low-power landscape and argue that conditions are ripe for low power, wearable cameras today. Glasses. With glasses we can see the eye and we can see what the eye can see. We highlight state of the art and trends in low-power hardware showing that low-power context-aware eyeglasses will be viable soon, and present the challenges remaining. Finally, we demonstrate a prototype, a pair of wireless eyeglasses containing two cameras. Combining photo data from a front-facing scene camera and an inward-facing eye camera, our system uses machine learning to predict the users gaze. We show that our glasses prototype, which was built using low-power COTS parts, can be used to track the wearer's gaze with reasonable accuracy.

Categories and Subject Descriptors

B.4.2 [**HARDWARE**]: Input/Output and Data Communications—*Input/Output Devices*; C.3 [**COMPUTER-COMMUNICATION NETWORKS**]: Special-Purpose and Application-Based Systems

General Terms

Design, Experimentation, Measurement, Performance

Keywords

Mobile phones, Energy harvesting, Phone peripherals, Audio communications, Participatory sensing

1. INTRODUCTION

It has been said since biblical times that "the eye is the window to the soul". In modern computing, however, the eye remains a one-way mirror, accepting input from the world but providing no insight to the world interacting with it. With the development of the InSight platform, we seek to rectify this imbalance, building an efficient, wireless, full-day, comfortable gaze tracking system.

Gaze tracking enables a diverse array of applications. It allows for targeted advertisements that can change content depending on who is looking at them as well as establishing an authoritative count of impressions. Gaze tracking provides a powerful diagnostic medium for the detection of diseases such as schizophrenia or more immediate measures such as intoxication. Advanced, high-precision gaze tracking even allows for instantaneous dynamic lens adaptation. Combined with advances in materials science, glasses that automatically adapt to the current wearer enter the realm of possibility.

Efficient. Leveraging previous work [2, **?**], we perform a short (order of 5 minute) training session to develop a neural net that is loaded onto the glasses. This enables real-time gaze tracking on a low-power microcontroller, with a reasonable degree of accuracy.

Wireless. Our system includes a bluetooth communication component that enables real-time communication of user focus. Currently this is communicated only to a prototype Android application, however the system could easily be extended to communicate with local advertisements. For advertisements with a specialized embedded QR-code, a look at the advertisement can generate a positive viewed impression to the advertiser, including duration of focus and number of impressions.

Comfortable. To fully realize our vision, the gaze-tracking system needs to be as natural to the wearer as a regular pair of glasses. This necessitates a small, compact form factor both for the sense hardware (2x cameras) and the energy storage (lightweight battery). Our complete prototype system (cameras, PCB, plastic enclosure, frames, wires, battery) weighs in at 85 g, barely 3.5 times the weight of traditional glasses[1].

Full-Day. In addition to comfort, an always-on gaze tracking system must be sufficiently power efficient to run for an entire day. The InSight glasses are duty cycled, evaluating gaze every XXX s at a cost of XXX mJ per measurement. Powered off of our 8.8 g, 600 mAh battery this allows the InSight system to run for XXX hours continuously.

2. BACKGROUND

Cameras are becoming near ubiquitous in modern society. Today, they are found in shops, on traffic lights, in schools, in the pocket of almost every cellphone carrying person, and even, increasingly, in our homes. Thanks in large part to advancing technologies and falling costs, cameras are everywhere. And, while the average person on the street is aware that cameras are all around capturing pictures and video, what they are not aware of is that there are a whole host of smart cameras, which turn visual data about some knowledge of the world.

Smart cameras, perhaps better called vision sensors than camreas, are cameras which extract some information fro visual data. Long at the heart of cutting edge manufacturing techniques, these so called smart cameras using vision processing and machine learning techniques to extract context from or apply labels to visial data. Increasingly, smart cameras are being employed in more dynamic environments in tasks such as counting bodies or vehicles and even facial recognition.

Another purpose of smart cameras is to glean information from peoples' eyes. By tracking the movement of the eyes, cameras can glean information about the person who the eyes belong to, such as that he or she is sleepy or not paying attention or drunk. Eye tracking can even be used to detect and treat mental disorders such as Schizophrenia [1]. In addition to eye tracking, cameras can peform gaze tracking, which combines information from an image of the eye with an image of the scene within the field of vision to identify where the eye is looking.

Commercial eye and gaze tracking headgear exists, but such systems are expensive, can be bulky and are often inelegant. Examples of current eye tracking headgead are the Ergoneers Dikablis and the ASL Mobile Eye-XG. The Mobile Eye-XG is a wired system and the Dikablis glasses use a proprietary wireless interface which requires running special software for interaction with the glasses.

2.1 Related Work

3. SYSTEM OVERVIEW

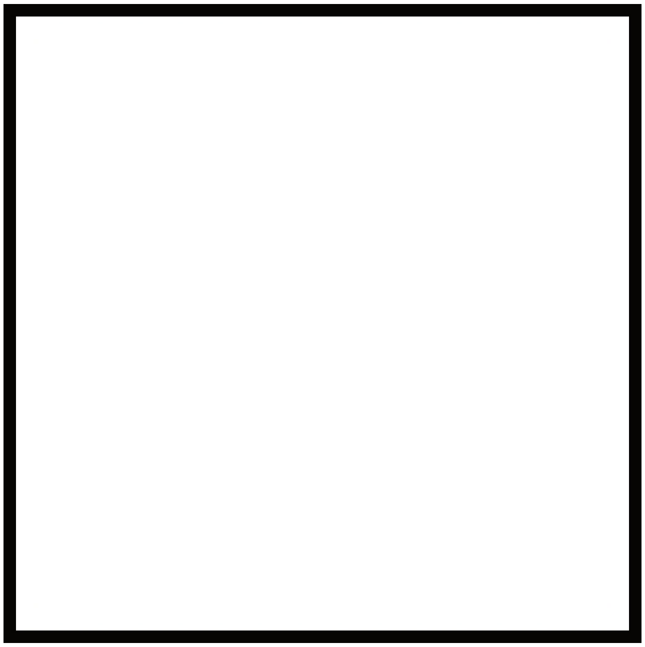

Figure 1: ((Picture of Glasses)) The complete InSight system. The camera board features two cameras, placed back-to-back observing both the eye and the scene the eye is observing. The battery is placed opposite the image processor and cameras, balancing the weight of the glasses.

The InSight system is composed of four discrete components: Bi-Directional Image Capture, Processing, Communication, and Energy Storage.

3.1 Bi-Directional Image Capture

–¿ This is 'what is it', design is 'why is it'

Idea of all the pieces, where are cameras, where is computation done, etc

4. SYSTEM DESIGN

In this section, we present the design of our wireless eyeglasses. Many important constraints must be considered including weight, battery life (power draw), data aquisition rate and the on glasses processing capabilities.

4.1 Weight

When developing a wearable computing device, one of the first constraints to consider is weight. Our eyeglasses must be worn on the user's head like traditional eyeglasses, and therefore weight is an important constraint. In a brief survey of collegues' spectacles, we found weights ranging from 15-50g. In order for a pair of eyeglasses to be feasible for all day use, they must weight something close to this range.

[1] Estimated from a small survey of author's glasses

Battery	Capacity (mAh)	Weight, Manufacturer (g)	Energy Density (mAh/g)
3xAAA NiMH	800	36	22
Sparkfun 850mAh	850	18.5	46
Sparkfun 400mAh	400	9	44
Sparkfun 110mAh	110	2.65	41.5
Tenergy 3.7V 500mAh	500	12	42
Tenergy 3.7V 1000mAh	1000	13	76
Trustfire 10440	600	9.6	63
Trustfire 10440 (enclosed)	600	FIXME	FIXME

Table 1: Rechargeable Battery Roundup. This table shows a roundup of some off-the-shelf rechargeable batteries which were considered for our glasses prototype. All batteries in the chart are lithium except the NiMH AAAs. Their weights, capacities and an "energy density" attribute (capacity per weight, measured in mAh/g), are shown. The energy densities have been computed with the measured weight for each battery except the enclosed Trustfire 10440. All battery setups listed run at nominal voltages of 3.6-3.7V, and all lithium batteries listed contain protection circuitry.

In addition to a maximum weight constraint, a pair of glasses has a balance constraint. Accomplishing a balanced pair of glasses is as easy as adding weight to the lighter side until an acceptable balance is reached. However, since we want to minimize weight, the glasses should be designed to be close to balanced. Our primary sources of weight are the circuit board and battery. We simply decided to place the battery on one leg and the circuit board on the other.

The battery for the glasses is one of the heavier components, so we start our estimate of system weight by comparing several several candidate batteries. With our choice of battery, we have a trade-off between conflicting constraints, weight and energy capacity (which feeds directly into battery life). We capture this interaction by dividing capacity (in mAh) per unit weight (in g), resulting in an "energy density" with units mAh.g. A comparison of potential batteries and capacities is provided in table 1. Each COTS battery considered contains a protection circuit and runs at a nominal 3.6-3.7V.

Finally, we want a baseline weight for our circuit board. We have taken the Hermera board as an example of a similarly sized and weighted circuit board. The hermera board weighs Xg. This is heavier than some of the batteries in table 1 and lighter than others. Again, we will assume the system weight to be roughly double the larger of the two weights, plys the weight of the sunglasses chassis. Our sunglasses weight Xg.

4.2 Battery Life / Power

We want to take a look at the power constraints on a pair of wireless glasses. Wireless glasses have unique constraints. These batteries are just a selection of off the shelf parts which are reasonable capacities, voltages and weights for an embedded device running on a pair of glasses. It gives an idea of what kind of power draw our system really needs to achieve the goal of all day use.

Again, table 1 shows examples of COTS lithium rechargeable batteries. All of the batteries have protection built in and run at a nominal voltage of 3.6-3.7V. The heaviest (but not highest capacity) battery is

Because we intend our device to be worn all day potentially, we can come up with an initial baseline for battery life. We will assume that all day entails 20 hrs in the worst case. So, to get our energy needs, we merely need an estimate of our system's power draw. If, for example, our system draws 50mA on average, the glasses would require at least a 1000mAh battery to achieve 20hrs of battery life, assuming perfect efficiency.

4.2.1 Cameras

For a long time, cameras have been power hungry. In recent years we've seen proliferation of lower and lower power cameras, and they are now nearly ubiquitous within cell phones. Now, we are finally starting to see truely low power cameras enter the market. Typical CMOS low power cameras draw in the several 10s of mW range. Fr example the Aptina MT9V11 ultra low-power camera draws 80mW in active mode, and the Omnivision OV7670 draws 60mW, in both cases at 15 FPS and VGA resolution.

Another factor of energy cost in most CMOS cameras is the required complexity necessary when interfacing with them. Often, the camera runs at a set FPS and data must be captured according tight timing constraings. This requires more complexity from the embedded interfacing device, and this almost always translates to more power.

We are focusing low-power image sensing on truly low-power. In this usecase (a go-everywhere wearable computing device), we need a camera which can do significantly better thanthese VGA options, but we're willing to sacrifice image quality to get that. An exciting new camera, the CentEye Stonyman, is capable of running at any speed (completely asyncronous). We have taken some measurements of the CentEye Stonyman camera and reveal it's very modest energy consumption when active, capturing an image.

5. IMPLEMENTATION

We present the design of our sensing eyeglasses in this section. These eyeglasses represent a first prototype proving the feasability of all day, gaze tracking eyeglasses. To interface with our imagers and the send data back to a host PC or android phone, we connected our image sensors to a relatively low-power and low-processing-power microcontroller. We used the Teensy 3.0 for prototyping these glasses. The Teensy 3.0 prototyping board uses a Freescale Kinetis K20 ARM Cortex-M4 microcontroller running at 48MHz and housing 16KB of RAM, a modest enough processor in most respects. We have mounted our electronics on a sunglasses chassis, a pair of knock-off Ray Ban's. The evolution of the glasses prototype is show in Figure/Picture X. Finally, our InSight glasses prototype contain the Trustfilre 600mAh battery shown in table 1.

As detailed in the design section, we have chosen to use the CentEye Stonyman embedded vision chip. Actually, our glasses use breakout boards containing full cameras with lenses and all. These embedded cameras can be purchased directly from CentEye on their website and cost around fifty dollars. Our InSight glasses use two Stonyman cameras, one inward facing on the user's eye and one outward facing, capturing the scene in the wearer's gaze.

We have designed the camera configuration on the InSight glasses for the purpose of gaze tracking. Gaze tracking determines a user's gaze by taking pictures of the user's eye. With a camera in the front and a camera in the rear, we envision our system could potentially output a picture of what the wearer sees and a gaze location at regular intervals or on demand.

We have chosen bluetooth for the transmission of data off of the InSight glasses primarily due to it's ubiquity and ease of setup procedure. The Bluetooth module that we used, the ST SPBT2632, consumes roughly 30mW in active mode.

5.1 Gaze Tracking

In order to track the user's gaze with our InSight system, we have employed a machine learning algorithm on a host device. We have chosen to employ a k-Nearest Neighbor classification to determine the user's gaze in one of nine squares of the scene. Both width and height of the scene pictures are split into thirds giving nine squares in total. Machine learning is a promising technique for low-power devices because it works more effectively in natural lighting (i.e. without the addition of an IRED), and because after an initial training, often the cost of predicition is quite cheap.

Choosing features is an important part of designing a machine learning algorithm. The feature set we chose for each eye image was merely the raw pixels from the image. Feeding raw pixel values to a machine learning algorithm has been done by Baluja in [2]. This simplistic choice of features has the advantage of being easy to compute, in fact it costs nothing more thancapturing the image. This is promising for running on the embedded device, because it does not cost anything to compute. The disadvantage to this simplistic choice of features is robustness and flexibility.

6. EVALUATION

We now present the evaulation of our InSight wireless, gaze tracking glasses. We evaluate Insignt on three objectives: accurate gaze tracking, suitable battery life and acceptable weight and balance. Our experimental results show that with InSight, after a short training session (less than three minutes), we are able to track the wearer's gaze with over 95% accuracy in all trials. Measuring the power draw of our system, we show that we are able to achieve X hrs of continuous use. Finally, our glasses weight in at 85g including all circuity, a lithium battery, two cameras, sunglasses chassis and 3-D printed plastic enclosures.

6.1 Gaze Tracking

We evaluate the accuracy of gaze tracking using the InSight glasses with a data set of 28 trials comprising 23 unique individuals. Each trial contains a minimum of 225 sample points (eye image, gaze location pairs) with 18 of the 28 trials comprising 540 or more sample points. Our trials include varying eye colors and shapes and a few different indoor lighting environments. In a couple of trials, the user was also wearing corrective lenses underneath the InSight glasses. All eye images are captured at natural illumination.

Although the number of samples collected for each participant in our study is varied, the basic experiment was always the same. Users were instructed to place their focus on a particular point on a wall or computer screen while eye image and scene image matched pairs were recorded. In the case of using marks on a wall, a laser pointer was used so that we could later mark the user's gaze coordinates manually. The laser pointer shows up in the scene images, telling us which gaze location the user was holding. Users were instructed to blink normally.

Trial	Samples	Correctly Predicted (%)
1	266	98.11
2	685	99.27
3	995	97.91
4	1385	97.47
5	1095	97.26
6	965	97.41
7	1074	95.33
8	942	97.34
9	1130	95.58
10	540	97.22
11	540	97.22
12	540	98.15
13	540	95.37
14	540	98.15
15	540	98.15
16	540	97.78
17	225	100.0
18	225	95.56
19	225	100.0
20	225	97.78
21	315	100.0
22	315	98.41
23	315	96.83
24	315	96.83
25	315	96.83
26	540	100.0
27	315	100.0
28	540	99.07

Table 2: Gaze Tracking Results: Intra-trial Evalation. In this table, we list each of the 28 trial data sets we captured along with the number of data points (eye image and gaze label pairs) and the percent accuracy with which InSight is able track the wearer's gaze. We see that in the worst trials, InSight still achieves an accuracy of over 95%, and in the best trials, InSight correctly predicts the wearer's gaze in all of the test data points. Averaging over all trials, InSight attains a 97.83% accuracy. After initial training, InSight is able to coarsely rack the wearer's gaze with very high accuracy.

For evaluation, in all gaze tracking experiments, we followed the same methodology. First, we randomized the sample points within the data set under experimentation. Next, the data set was split 80/20 into training and test data. Finally, we ran our k-Nearest Neighbors machine learning algorithm over the data set using the training data for training and the test data for testing the classifier's accuracy. Tracking the gaze position to 9 disctinct gaze "bins", InSight achieves over 95% accuracy in all experiments.

6.1.1 Intra-trial Results

Initially, we evaluated InSight's gaze tracking accuracy by testing the classifier results against test data using only data from within the same trial. For each of the 28 trials, we randomized and split the data set 80/20 into training and test data. We the measured the performance as a percent accuracy by dividing the number of test labels returned which predicted the correct gaze location bin by the total number of test labels. The correct gaze location is predicted when the label prediction agrees with the ground truth label.

In table **??**, we show each of the 28 trial data sets we captured along with the number of data points (eye image and gaze label pairs) and the percent accuracy with which InSight is able predict the wearer's gaze. We see that in the worst performing trials InSight still achieves an accuracy of over 95%, and in the best trials, InSight predicts the wearer's gaze with 100% accuracy. Averaging over all trials, InSight attains a 97.83% accuracy. We can see from these results, after initial training, InSight is successfully able to coarsely rack the wearer's gaze with very high accuracy.

6.1.2 Results of All Trials Together

6.2 Battery Life

7. DISCUSSION

7.1 Battery Life

7.2 Performance

In speaking of performance in this section, we will constrain the discussion to the framerate performance of the InSight glasses. The low framerate of the InSight glasses is an obstacle to seamless gaze tracking. At 4.5 FPS per camera, 2.25 FPS when both camera's images together are taken as a frame, the image capture rate of the InSight glasses is quite low. There are 444ms between each frame, which is not an imperceptable time. In addition the human eye can initiate motion almost 5x faster than that [CITATION NEEDED]. We have teased apart the causes of the low frame rate and attacked them one at a time.

We have calculated the theoretical maximum framerate of one CentEye Stonyman camera. The camera requires that images are sampled pixel-by-pixel with an analog-to-digital conversion performed outside of the imager. This teqnique is allows for low-power operation, especially when the possibility of subsampling is on the table. But, it comes with a certain tradeoff, namely speed.

For low voltage operation (operation at voltages roughly ¡4V), the camera requires the use of an internal preamplifier. This preamplifier must be operated for each pixel captured. Operation of this preamplifier adds a constant overhead of 2us to *each* pixel capture. Given the camera's full resolution of 112x112 pixels, merely operating the amplifier adds 25ms to each frame. In addition capturing a frame without the preamplifier circuit requires on the order of 20-30ms. Therefore, operating at low voltage requires 45-55ms per full Stonyman frame. This leads to a theoretical maximum framerate of 18.18 - 22.22 FPS. This value is respectable, but not great.

The Stonyman's theoretical framerate of 20 frames per second is still a way off from the InSight glasses' framerate of 4.5 FPS per camera. We examined the causes of the low frame rate in the InSight system. The two major causes of our low frame rate were the relatively low bandwidth bluetooth link and the lack of multiplexing in the microprocessor, which we will discuss further later. First let's start with the wireless bandwidth.

d

7.2.1 Wireless

Our choice of bluetooth as the wireless technology of the InSight gaze tracking glasses was primarily for the ease of interfacing with reasonable devices like PCs and cell phones. That choice, however, constrained the speed of the insight system.

The maximum theoretical datarate of a bluetooth link in one direction is 1.5Mbps. Our bluetooth module interfaced with our microcontroller via UART ran at a maximum of 921.6kbps. A quick and dirty calculation of maximum framerate as constrained by the bluetooth module is to divide our maximum datarate by the size of a frame in pixels, which we'll assume can be represented as a single byte. 921600bps / (112*112*8)bits per frame = 9.18 FPS, which barely half of the theoretical framerate of the stonyman camera. This is an extremely crude calculation, neglecting all overhead both timing wise and data wise, meaning the actual theoretical framerate should be a bit lower.

WiFi represents a promising technology for glasses such as these. Since operation of the radio represends the largest energy cost, we ought to send data as little as possible. Using wifi would allow better something or other...

FIXME: finish this writeup!

7.2.2 Lack of Multitasking

Perhaps a more fundamental performance limitation of the InSight wireless glasses design was the lack of simple multitasking. The embedded system basically has 2 goals:

* Capture data from the image sensor * Transmit that data back to the device (PC, phone, tablet)

With 2 threads of execution, this problem is a simple one. Put one thread to work capturing data and the other thread to work transmitting captured image data over the Bluetooth link. In our InSight system, we don't handle this problem. When imager data is being captured, we are not transmitting data, and when we are transmitting data, we are doing nothing with this camera.

For example, digitizing pixels is a large chunk of the time it takes to capture a single frame, accounting for roughly 25-33% of the entire capture, and during that time, the processor is idle, waiting on the ADC. The same is true when the aplifier is operated, except that pulsing the preamplifier represents an even larger chunk of the pie at 45-55% of the total time to capture a single-full-frame. At minimum, 60% of the time spent capturing images the processor is idle, waiting.

There are a few ways we can meet the need for multitasking in an embedded system. The traditional solution is to use interrupts. Interrupts would let control flow back and forth between the camera control and capture code and the data transmission code. Interrupts could be triggered on events (such as an ADC conversion completition), or external timers for fixed duration delays. Given that our problem is basically that we meed to move data from one buffer to another, DMA could potentially provide a workable source of multitasking.

Interrupts? DMA? Are these good solutions? Fundamentally, the microcontroller is a poor fit for the task of capturing this image data and tramsitting it again. An FPGA, however, can give us natural parallelism and strict and efficient timing of short delays as well. These are the solutions we need to in crease performance.

We have recreated the dual camera configuration of the InSight glasses on an FPGA and evaluated its performace as a benchmark of possible performance the system could achieve.

8. CONCLUSIONS

In this work, we have laid out the design space for low-power wireless vision enabled systems. We have argued that now is the time for such etcetcetc.

We have demonstrated this with our proof of concept, the InSight wireless gaze tracking glasses. Our glasses are capable of coarsely tracking the wearer's gaze with startling accuracy.

9. ACKNOWLEDGMENTS

10. REFERENCES

[1] L. Abel, S. Levin, and P. Holzman. Abnormalities of smooth pursuit and saccadic control in schizophrenia and affective disorders. *Vision Research*, 32(6):1009 – 1014, 1992.

[2] S. Baluja and D. Pomerleau. Non-intrusive gaze tracking using artificial neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, 1994.