SensEye: An architecture for ubiquitous, real-time visual context sensing and inference

Continuous real-time tracking of the eye and field-of-view of an individual is profoundly important to understanding how humans perceive and interact with cyber-physical systems. Such continuous monitoring can enable detection of hazardous behaviors such as drowsiness while driving, mental health issues such as schizophrenia, addictive behavior and substance abuse, neurological disease progression, head injuries and others. This proposal seeks to perform fundamental cross-disciplinary research into a vertically-integrated architecture for designing novel ultra-low power and affordable real-time visual context sensing systems, thereby advancing the state-of-art in on-body context sensing. Our work will advance both the *technology and engineering of cyber-physical systems* by designing an innovative paradigm involving next-generation wearable spectacles with capability for real-time visual context sensing and inference in conjunction with on-body and infrastructure sensors, and real-time interventions in health and transportation.

Intellectual Merit: The proposed effort integrates novel research into low-power embedded systems, image representation, image processing and machine learning, and on-body sensing and inference, to advance the state-of-art in body sensing for CPS applications. The resulting cross-layer ideas will address several fundamental research challenges in next-generation on-body sensing and inference including: 1) design of a novel, highly integrated, wearable eye tracker, SensEye, for tracking eye movements, visual field of a user, and head movement patterns, all in real-time; 2) a unified compressive signal processing framework that optimizes sensing and estimation, while enable re-targeting of the device to perform a broad range of tasks depending on the needs of an application; 3) design of a novel real-time visual context sensing system that extracts high-level contexts of interest while being robust to shifts in sensor placement; and 4) a layer of intelligence that combines contexts extracted from SensEye together with contexts obtained from other sources (chestbands, wristbands, smartphone, vehicles, etc) to learn higher-level patterns of behavior, and trigger application-specific interventions. Our research will result in novel hardware designs, signal processing techniques, machine learning algorithms, context inference techniques, and on-body sensing and intervention systems.

Broader Impact: Our technology can revolutionize a spectrum of disciplines including transportation, healthcare, behavioral science and market research. In this project, we plan to demonstrate the utility of our work by demonstrating its use in transportation and mental health. We propose to release a family of open-source SensEye platforms and software systems with different choices of cameras, computation, communication, storage, and tethering choices. Some of these platforms will be made available through hobby stores such as Sparkfun for broader use. The PIs will enable outreach by distributing some of the devices in high-schools in under-represented areas, and demonstrating the device at science fairs for high-schoolers. Our project also integrates SensEye into undergraduate and graduate courses.

Keywords: visual context, eye tracker, body sensors, Science of Cyber-Physical Systems, Engineering for Cyber-Physical Systems.

1 Vision

Mobile on-body sensor devices provide a remarkable new set of sensors and effectors for enhancing a variety of application domains. One of the tantalizing possibilities of on-body sensing technology is that we can combine behavior sensing together with infrastructure sensing to understand how humans interact with cyber-physical systems in their daily life. These behavioral models can, in turn, facilitate the design of novel real-time systems that provide personalized and timely interventions, enabling safer and lower cost cyber-physical systems.

An important aspect of on-body sensing is tracking the eye and visual field of an individual. Continuous real-time tracking of the state of the eye (pupil dilation, eye closures, eye movements, gaze direction) in conjunction with the field of view of a user is profoundly important to understanding how humans perceive and interact with the physical world. Real-time tracking of the eye can help detect a range of unsafe behaviors, including drowsiness while driving, lack of attention while using a medical device or distractions while crossing a road, The benefits extend to personal health, where the state of the eye provides a continuous window into Parkinson's disease progression, psychiatric disorders such as schizophrenia, drug use, head injuries and concussions, and others. Real-time tracking of gaze direction can also introduce new "hands-free" ways of interacting with computers or displays, and provide an accurate method of gauging user intent to drive context-aware advertising.

While our understanding of the human eye and gaze has grown through decades of research on the topic [**?**, **?**], wearable eye tracking remains limited to controlled user studies and clinical trials, and has not crossed the gap to daily consumer use. This is, perhaps, not surprising — eye trackers require continuous operation of several cameras, high data volume, compute-intensive image processing and calibration, all of which make such devices extremely complex from a hardware/software perspective. Thus, state-of-art eye trackers are bulky, expensive data loggers that are unsuitable for long-term wear.

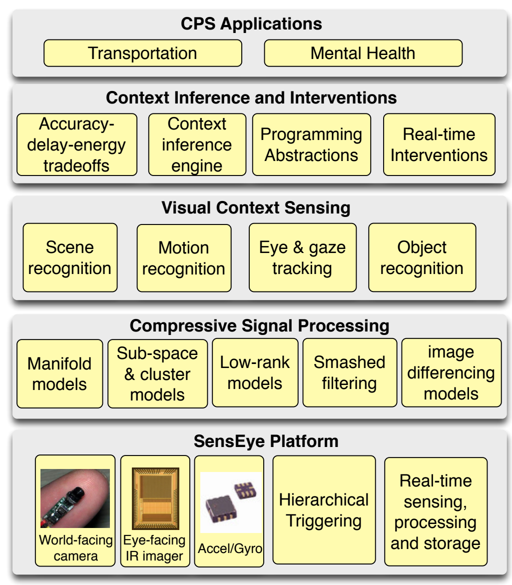

Figure 1: System overview.

But technology trends in ultra-low power hardware and ubiquitous use of smartphones promise a radical transformation in visual context sensing. Consider, for example, the M^3 sensor co-designed by PI Dutta [**?**], which integrates an imager, battery, computation, communication, and energy harvesting within a tiny *cubic-mm* package that consumes a few tens of nW of power. In addition to the obvious benefits of easy integration within wearables such as spectacles, the ability to perform on-board processing and inference, and communicate with a phone or cloud, can enable continuous visual context sensing and real-time interventions.

While the possibilities are enormous, there are fundamental challenges at various levels that need to be addressed. The first challenge is dealing with the stringent restrictions on sensing fidelity, energy budget, computation capability, and other resources that arise from the need to package on-body sensors into lightweight, aesthetically pleasing units. The second challenge is dealing with a variety of real-world artifacts such as changes in sensor placement, variability in lighting conditions, and others that can reduce inference accuracy. The third challenge is designing systems that can combine visual contexts together with contexts from other on-body or infrastructure sensors to enable real-time understanding of human behavior in cyber-physical systems, as well as to close the loop to design novel interventions.

1.1 Proposed Work — Research Objectives

The goal of our research is to develop an end-to-end system that includes novel ultra-low power eye trackers with real-time sensing, processing, and inference capability, as well as the fundamental knowledge-base necessary to discover patterns of human behavior and leverage it for improving transportation and healthcare. Our architecture, SensEye enables long-term, continuous tracking of visual contexts in real-world settings for a variety of CPS applications, and involves fundamental inter-disciplinary advances in both engineering and computing. Our architecture has four key layers (Figure **??**):

- **The SensEye platform:** At the lowest layer of our stack is a novel, highly integrated, wearable computational glass, SensEye, for tracking eye movements, visual field of a user, and head movement patterns. This work answers several core scientific questions including: How can we strike a balance between the sense/compute activities and their impact on energy consumption? How can we support evolving computational requirements of both expected and unforeseen data processing workloads? How can we engineer a wearable sunglass form-factor device with sensors, processing, storage, communication, and charging elements?
- **Compressive Signal Processing:** The layer above the SensEye platform is a unified compressive sensing framework that is adaptive to diverse applications and yet highly optimized in terms of resource usage. This work answers several fundamental signal processing challenges including: how can we design highly optimized yet general image representations that can be re-targeted to different CPS applications? How can we achieve high-levels of compression through such representations to minimize the sensing and communication overheads? How can we perform signal processing over such highly compressed representations to minimize the computation overhead?
- **Visual Context Sensing:** Inferring visual contexts is often the most demanding part of the processing pipeline, and needs to extract low-level features from the compressively sampled signals, and perform inference over the continuous signal to extract eye, gaze and scene contexts. This work answers several fundamental questions including: How can we design a highly optimized context sensing pipeline for SensEye that integrates scene recognition, motion recognition, eye and gaze tracking, and object recognition? How can we reduce or eliminate the need for re-calibration to deal with device shifts as people move around?
- **Context Inference and Interventions:** This layer of intelligence is a learning engine that combines contexts extracted from SensEye together with contexts obtained from other on-body and infrastructure sensors (chestbands, wristbands, smartphone, vehicles, etc). This work addresses fundamental scientific questions including: How can we learn relations between visual, activity, location, physiological, and other contexts in real time, and utilize them for improving performance and learning personalized behavioral patterns? How can we design a real-time intervention engine that can enable timely interventions? What are the right abstractions for a high-level language and execution framework to enable diverse CPS applications to specify their sensing, inference, and intervention needs using SensEye and other sensors?

Overview of Experimentation and Validation: We will validate the SensEye system in two real-world CPS applications, leveraging existing relations with transportation and medical researchers. One of our applications will involve detecting hazardous driving conditions with the Transportation department at UMass, and another one will be for detecting mental health issues in collaboration with Yale Psychiatry. These real-world deployments will help understand the generality of SensEye in several application contexts, and uncover novel challenges.

Overview of the Broader Impacts: The broader impact of such technology cuts across a broad swath of disciplines including medicine, transportation, marketing, and behavioral science. Recent years have seen advances in several wearable hardware platforms such as physiology-monitoring chestbands, activity-monitoring wristbands, and other mHealth devices such as glucose/alcohol monitors. Each of these devices has revolutionized body sensing, leading to hundreds of mHealth apps for mobile phones and tremendous excitement in behavioral and health sciences. SensEye promises to be just as disruptive, if not even more.

Credentials: Our team is well qualified in carrying out the research proposed in this project. We have assembled a multi-disciplinary team with expertise spanning various relevant CS and EE disciplines: low-power embedded systems (Dutta, Ganesan), image representation, compression and estimation theory (Duarte), machine learning and image processing (Marlin), and camera sensor networks (Dutta, Gaensan). Our team has extensive experience with developing open-source hardware and software artifacts that are widely used. In addition, we have a domain expert from transportation science (Knodler) and a letter of collaboration from a mental health expert (Malison), who will help refine the specific application scenarios where SensEye will be used, and help validate our system under realistic application settings.

2 The SensEye Platform

Eye tracking research started over a century ago but only recently has it become possible to combine all of the key elements into a unified system – an everyday pair of eyeglasses. Today, state-of-art eye trackers such as the Tobii Glasses provide an existence proof that a camera and its supporting electronics can fit into an eyeglass frame [**?**]. Although the form factor improvements greatly aid usability, the Tobii system does not operate in real time, and requires a mobile phone-sized companion data recorder and calibration assistant. These factors limit size, generality, and utility in real-time CPS applications.

We argue that sensing, computing, communications, storage, and energy-harvesting technologies are on a trajectory that will soon make it possible to integrate all of the elements of an eye tracker into a self-contained, unobtrusive, and natural real-time visual context sensing system. Such an integrated system must address a number of other important concerns and tradeoffs. Perhaps the largest unknown is the level of computing possible for a particular power budget. Computation is not the limitation — electronics can be miniaturized to fit in eyeglass form-factor devices and execute a range of vision processing algorithms. Rather, the issue is that the systems are power-limited; the power cost of sufficient CPU cycles is beyond the limits of what one could reasonably expect to house inside a pair of glasses.

This tension leads to the following question: *how should we design a research-enabling platform that supports a rich exploration of the algorithmic design space while ensuring high platform utility?* Any viable platform must offer both low power and high performance, but not necessarily at the same time since these functions are somewhat exclusive [**?**]. Our goal, therefore, is to improve the range of possible designs by integrating as many power-performance tradeoffs/options into this platform as possible to support the widest diversity of workloads. We first describe the elements in a basic SensEye platform, and then present an evolution of the platform from a research-enabler to a high-performance integrated real-time sensing device.

SensEye Platform Basics The key elements of the platform include: 1) Two cameras – one eye-facing and one world-facing. The world-facing camera can be integrated into either the middle of the eyeglass frame or one on each side for stereo vision. The eye-facing camera placement will require experimentation and a balancing act between location, field-of-view, and depth-of-view. Infra-red filters may be needed on the eye-facing cameras to deal with different lighting conditions. 2) A flexible hardware (FPGA) and software (CPU) boundary by employing recently emerging flash-based, mixed-signal FPGAs. The key benefit of this approach is that is supports moving timing-critical or computationally-intensive operations to parallel hardware implementations, if needed. The approach also affords greater flexibility since the hardware/software boundary can changed dynamically in support of evolving analysis, algorithms, or workloads. 3) Several communication interfaces including a USB-OTG interface for high-rate data capture using modern mobile phones, which will be particularly helpful during early algorithm analysis and system development; a Bluetooth interface for compute offload to a phone or cloud; and an optional low-power 802.15.4 interface to directly interface with other body area sensors. 4) High-density flash memory storage for logging sensor data, and to store key index structures for visual context inference (§**??**). 5) Actuators such as a vibrating buzzer that provide the ability to close the loop and trigger interventions, for example, when an unsafe driving scenario is detected in a car. 6) A small, onboard, rechargeable battery that is inductively recharged using wirewound coils that are integrated into the spectacle frames, thus allowing the glasses to be recharged by simply placing them on "charging mat" at night.

2.1 Research Plan

Our platform research plan aims to understand and characterize the design space of computational eyeglasses. The key challenge lies in understanding and striking a balance between the sense/compute activities and their impact on energy consumption. To support the evolving computational requirements of both expected and unforeseen data processing workloads, we plan to experiment with a range of computing systems.

Design space exploration: Initially, we propose to create a tethered device, much like a real-time version of the Tobii Glasses, which will allow us to explore the platform design space while letting truly modern algorithm development proceed in parallel. The eyeglasses will have sufficient computing power to stream video from two or three cameras and multiple low-rate sensors to an Android phone connected over the USB On-the-Go interface. The glasses will also include energy metering support so that real-time energy consumption data are available to the software runtime using our previously developed iCount energy meter [**?**]. The Android phone will be augmented with our previously developed BattOr system, allowing us to capture fine-grained energy consumption data from the mobile phone [**?**]. The research goal at this stage will be to understand the computation and energy costs of the various subsystems and algorithms, and identify candidates that might best be implemented in hardware or software.

Integrated platform: With better models of the sensing/computing energy costs, we propose a second phase to design and characterize an evolved platform that integrates greater computing resources, onboard storage, additional low-power sensors, and wireless battery recharging. For computation, we propose to use the Microsemi SmartFusion customizable system-on-chip (cSoC) which integrates an ARM Cortex-M3 microcontroller, a flash-based FPGA, and an analog compute engine [**?**, **?**]. This platform will be augmented with dense flash memory storage, low-power sensors, and power support. Our research will address two key questions: 1) Which algorithms should be parallelized on the fast but finite FPGA fabric and which ones to time-multiplex on the slow but plentiful CPU cycles? The key benefit of our platform for studying this question is that it tightly couples the processor and FPGA into the same memory space, allowing us to explore hardware (FPGA) and software (CPU) realizations of the various algorithms, and 2) What low-power

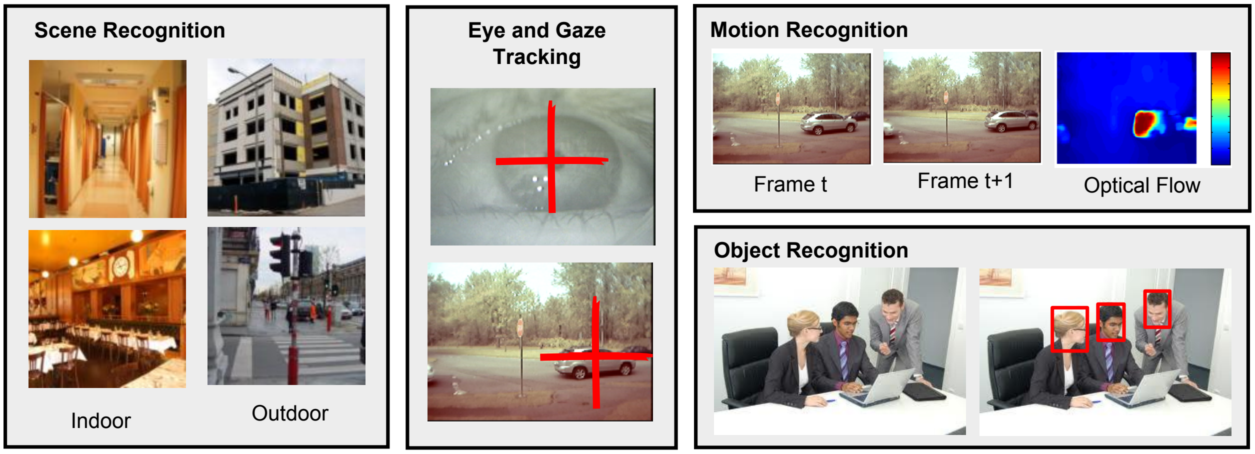

Figure 2: Illustration of the four visual context sensing task categories with examples: scene recognition (indoor/outdoor classification), eye and gaze tracking, motion recognition (magnitude of optical flow), and object recognition (face detection).

sensors should be used to minimize energy consumption. One option is nanopower imagers that integrate optical flow-based motion tracking. Specifically, we will explore how our prior work in developing cameras that are just 1 mm x 2 mm in size **??** can be used to design continuously-running micropower cameras to trigger more powerful cameras to support low-power visual context sensing. Another option is leveraging low-power sensors such as accelerometer, gyroscope, magnetometer, or light intensity sensors, e.g., to trigger the camera upon motion or light changes.

3 Visual Context Sensing

At the core of the SensEye architecture is the ability to extract valuable visual context information regarding the state of the user and the environment in which the user is located. In this work, we consider four broad categories of context features derived from scene recognition algorithms, motion recognition algorithms, eye and gaze tracking algorithms, and object recognition algorithms. The detected scene categories will provide location context information in terms of generic place types to complement other mobile sensors like GPS that provide location coordinates [**?**]. By leveraging motion features like optical flow in the world-facing video stream [**?**], we will be able to get macro-level contexts such as traffic flow information [**?**] and differentiate between activity contexts like driving and walking. The ability to sense motion is also very useful in safety-oriented applications like detection of possible collisions while driving. The eye-facing imager will provide context information about the user's state of alertness and attention. By analyzing eye movements in the eye imager video stream, we can measure correlates of alertness and attention such as frequency of eye closures, length of fixation times and dilation of the pupil [**?**]. We can also track the movements of the user's eye to perform gaze-point estimation and determine where the user is currently looking in the scene [**?**]. Object recognition provides the ability to introduce application-specific context features, such as a car or face, which can be computed using any object classification approach.

A fundamental challenge in the design of a wearable sensor is dealing with resource constraints. The need to package sensors into lightweight, small form factor, and aesthetically pleasing units places restrictions on sensing fidelity, energy budget, and computation capability. The smart spectacle sensor that we propose to design is particularly challenging since, in addition to form factor constraints, visual sensing and image processing in real time is far more energy and compute intensive than other physiological or activity sensors. To deal with these challenges, we will focus on implementing algorithms for the four core visual context sensing tasks directly on top of a common compressive sensing representation. This common representation will both aid in minimizing the energy cost of communicating sensed inputs to external devices and in minimizing the energy cost of computation using local processing resources. We begin by describing background work on compressive sensing and the fundamental visual context inference tasks. We then present our research plan combining these two topics into a single visual context sensing pipeline.

3.1 Background

Compressive Sensing: In compressive sensing (CS), a signal $x \in \mathbb{R}^N$ (e.g., an N-pixel image) is represented in terms of a reduced-dimensionality representation $y = \Phi x \in \mathbb{R}^M$, $M \ll N$, where M is a randomized measurement matrix that represents a linear acquisition operator [**?**, **?**, **?**, **?**]. The large majority of prior work in CS has focused on full-length signal reconstruction, i.e., obtaining an accurate estimate $\widehat{x}$ of the original signal from the measurements y. In this setting, CS provides several properties that are well matched to the constraints of our problem of interest [**?**]. In particular: (i) randomized CS measurement matrices are *universal*, in the sense that they can capture relevant signal information, regardless of its particular structure, with very high probability — and the structure need only be known at the time of signal recovery; (ii) the linear dimensionality reduction in CS provides a *fidelity-scalable data representation*, in the sense that modifying the dimensionality M of the projection provides a gradual increase or decrease in the quality of the estimation $\widehat{x}$; (iii) the simple dimensionality reduction performed in CS makes it possible to *shift the computational complexity of signal compression away from the sensor*: the dimensionality reduction is achieved by a simple matrix-vector product that is computationally light-weight, and the large majority of the computational burden can be performed at a central processing unit.

In contrast to this existing literature, our goal in this project is to adapt the signal processing tasks required by the SensEye system to share the aforementioned CS properties. Importantly, we aim to perform such signal processing tasks *without requiring signal reconstruction prior to processing*; in other words, we aim for the signal processing and computer vision tasks to be performed directly in the compressed representation, allowing us to extend the computational and communications savings afforded by CS acquisition. PI Duarte's research focuses on this compressive signal processing (CSP) setup. CSP shares the same theoretical underpinnings as CS recovery: the formulation of novel signal models that are amenable to signal processing and compatible with the randomized dimensionality reduction performed in CS.

Two CS models of particular interest are manifold models and low-rank matrix models (subspace models). When an image generations process is governed by a small number of parameters, the resulting images form a low-dimensional manifold in the space of all images. This type of manifold structure provides a powerful model for highly structured signals [**?**]. Since the manifold structure is geometrical and is captured succinctly by the Euclidean distance between points in local neighborhoods, the Johnson-Lindenstrauss lemma [**?**, **?**] guarantees that we can perform manifold learning, navigation, and processing directly on the random projections of the acquired images with high probability of success [**?**, **?**]. An alternative to manifold structure is provided by low-rank matrix (or subspace) models designed to capture differences in the correlation structure of a set of signals (e.g., principal component analysis, subspace clustering, etc.) [**?**, **?**].

Visual Context Sensing: The problem of generic scene recognition was first posed in the work of Torralba where global image features including the gist descriptor were developed and validated [**?**]. The problem of estimating motion from a video stream can be addressed using optical flow estimation [**?**]. The problem of activity or motion recognition on top of optical flow is accomplished using optical flow-based motion descriptors such as the histograms of oriented optical flow (HOOF) descriptor [?]. These techniques have previously been applied to the problem of activity recognition in videos of people, although their use with video taken from the perspective of the person performing the activity is rare.

Eye tracking by video-oculography is the problem of tracking the movement of the eye from video. While video-based eye tracking in static laboratory settings has a long history [?, ?], mobile eye tracking is a much more recent development [?, ?]. There are still many issues with the robustness and ease of use of mobile eye trackers, which our work will directly address. Eye tracking will form the basis for the collection of several key problems including recognizing fixations, estimating dilation of the pupil, and performing gaze-point estimation and tracking. The problem of detecting the number and duration of eye closure (blink) events from the eye imager stream is more straightforward and does not require eye movement tracking.

The final component of our visual context sensing framework is object recognition. Object recognition is a long studied problem in computer vision and our research will build on all of this past work by integrating gaze detection and object recognition for the specific purpose of supporting application specific contextual inference. Cascaded classifiers like the Viola-Jones approach to face recognition are particularly appropriate for use in the mobile, resource constrained setting [?].

Scene recognition is the problem of classifying an image into scene categories like office, home, restaurant, park, etc [?]. Motion recognition is the problem of detecting motion in the scene from video. By leveraging motion features like optical flow in the world-facing video stream [?], we will be able to get macro-level contexts such as traffic flow information [?] and differentiate between activity contexts like driving and walking. The ability to sense motion is also very useful in safety oriented applications like detecting possible collisions while driving. The eye-facing imager will provide context information about the user's state of alertness and attention. By analyzing eye movements in the eye imager video stream, we can measure correlates of alertness and attention such as frequency of eye closures, length of fixation times and dilation of the pupil [?]. We can also track the movements of the user's eye to perform gaze-point estimation to determine where the user is currently looking in the scene [?]. Object recognition provides the ability to introduce application specific context features such as a car or face, which can be computed using any object classification approach.

3.2 Research Plan

Scene and Object Recognition: Current state-of-the-art approaches to scene and object recognition both rely on the extraction of basic visual features from raw image data. Image descriptors like the *gist* descriptor [?] have been widely used for scene recognition while descriptors like histograms of oriented gradients [?] and integral image-based features [?] have been widely used for object detection. Our proposed research will extend this work to general scene and object recognition algorithms that operate directly on the compressive sensing representation of the scene. We leverage the fact that many of the descriptors above correspond to the outputs of carefully designed filters applied on the image. This process can be represented geometrically as the computation of an inner product in a high-dimensional vector space $\mathbb{R}^N$, whose values are preserved by the random dimensionality reduction used in CS as long as the number of random measurements M is sufficiently large [?]. We will initially develop scene descriptors and scene recognition models using the Scene Understanding (SUN) database recently made available by Torralba's research group [?]. We will focus on face detection for an object detection task and learn initial models using the Labeled Faces in the Wild data set [?]. In both cases, we will convert the existing images into their CS representations, estimate descriptor values from the reduced dimensionality data, learn classification models using data sets containing descriptor/label pairs, and evaluate accuracy using existing benchmark algorithms for these data sets.

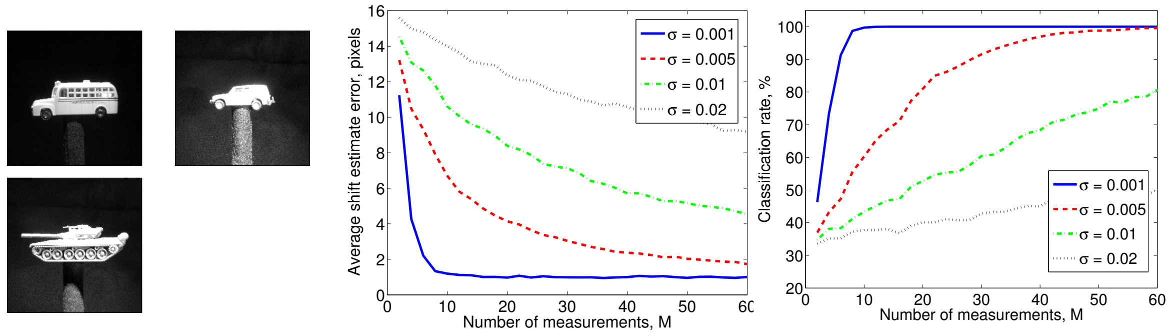

Figure 3: *Results for image classification experiments. Left: examples from three classes (bus, car, tank) that undergo articulation by spatial shift. Right: shift estimation errors and classification rates for the smashed filter from varying number of CS measurements M taken with PI Duarte's single pixel camera [**?**] and noise levels σ. As M increases, the performance improves due to higher-fidelity geometrical information in the compressed domain. (Taken from [**?**]).*

Eye Tracking: Our research on eye tracking will begin by developing methods for recognizing eye closures. This problem is amenable to CS since a sequence of eye images exhibits different correlation structure for alert and drowsy subjects and this structure becomes apparent by frame differencing. PI Duarte has previously shown that such frame differencing can be applied directly on the CS representation, once again thanks to the linear quality of the dimensionality reduction [**?**]. Additionally, PI Duarte has previously applied manifold models to compressive parameter estimation and event classification using a technique known as the *smashed filter*, an extension of matched filtering to the compressive domain [**?**, **?**, **?**]. Smashed filtering significantly reduces the dimensionality of the data to the order of the information rate (i.e., the dimensionality of the parameter space) while still achieving successful parameter estimation and classification (cf. Fig. **??**). We will apply manifold models to the problems of eye movement and pupil tracking. Eye tracking is an excellent match for manifold models because the images of the eye actually are governed by a small number of parameters including the rotation of the eye relative to the eye socket, the dilation of the pupil and the lighting conditions in the environment. Manifold models often require significant conditioning of the incoming data to eliminate confounding signals due to sources such as lighting. We will investigate techniques for overcoming these limitations including background subtraction, masking, registration, scaling, etc. Such processing requires additional side information on the event of interest, which may itself be obtained from initial data available. We will pursue the extraction and leveraging of such adaptation feedback between cameras in the SensEye system. We expect to be able to increase the number of tasks in computer vision that can be performed directly on the compressed data representations by leveraging such adaptivity.

Gaze Tracking: The key challenge in gaze tracking is the calibration of the eye-facing and world-facing imagers to robustly determine where a user's gaze is actually positioned in terms of world-image co-ordinates. This often requires an a priori calibration phase where the user must gaze at multiple specific points in a displayed image [**?**]. Static eye trackers go to great lengths to ensure that the calibration parameters don't change while a user is wearing the device, for example using head-mounts. Our setting is far more challenging since the accuracy of gaze-point detection can decay over time due to motion of the eye-facing imager relative to the eye in real-world use. This necessitates periodic re-calibration, which can be a significant barrier to adoption. We propose to address this issue by developing a novel self-calibration application for a smart phone connected to SensEye. For first-time calibration, the user will gaze at a location in the natural environment, click a button on the phone to record an image of the scene, and then click on the image displayed on the phone to register a calibration point. A number of calibration points can be registered and when the true gaze point and predicted gaze point are aligned, the calibration procedure ends. To reduce the need for periodic re-calibration, we will develop an automatic visual saliency-based calibration correction procedure. This algorithm will use the fact that people are more likely to fixate on salient points in images to correct the drift in the calibration by adjusting the mapping parameters according to the gradient of a saliency map [**?**, **?**] computed from the CS representation of the world-facing imager.

Motion Recognition: In the case of motion recognition, we will develop image differencing models that favor localized activity in the image. Such activity can also be succinctly represented using emerging low-rank matrix models for signal ensembles, which have been shown to efficiently recover and distinguish between the background and the activity in compressively acquired video sequences [**?**, **?**]. We will also consider the development of optical flow-based descriptors similar to HOOF implemented directly on top of the CS representation. To support the development of motion and activity recognition algorithms, we will use our hardware to collect a novel database of video sequences and annotate the sequences with standard activity labels like walking and driving, as well as labels indicating extrinsic motion in the scene.

Resource Constraints: While there has been a considerable body of prior work on all of these tasks, SensEye presents several new challenges. First, implementing a complex classification pipeline on SensEye is extraordinarily challenging due to the tight resource constraints and real-time requirements. This requires that we carefully optimize the pipeline to operate within the budget and timeliness constraints, while minimizing its impact on context sensing accuracy. For example, we will explore the use of cascaded classifiers (e.g. [**?**]) or triggered computation pipelines (e.g. drowsiness detector triggering a scene classifier to detect unsafe situations). The design of such a resource-bounded visual context sensing framework is unique to our platform and has not been addressed in prior work. Second, we will explore how to split computation across SensEye, smartphone, and the cloud. We will consider approaches to scheduling the components of the visual context sensing pipeline across these tiers to minimize energy use, as well as off-loading energy intensive but delay-tolerant tasks to remote cloud computing resources. In addition, cloud resources play an important role in the ability to re-train and refine the context sensing pipeline during periods when SensEye is charging and connected to the Internet. We will explore how our system can take advantage of daily periods of unlimited energy and connectivity to refine the models and algorithms in our pipeline. Third, we will look at how to optimize the speed-accuracy-resource trade-off of the algorithms underlying each of the four main visual context sensing tasks. As discussed in §**??**, SensEye platforms can differ in sensor and resource capabilities. We will design a library of techniques that offer the best performance for different design choices.

4 Context Inference from On-body Sensors

While our discussion so far has focused on visual contexts provided by SensEye, a myriad of other contexts are available from on-body sensors (e.g. smartphones, chestbands, wristbands, etc) as well as sensors in the infrastructure (e.g. vehicular sensors). Other on-body sensors can be used for obtaining contexts such as activity [**?**, **?**], location [**?**], stress and anxiety [**?**, **?**, **?**, **?**], social interactions [**?**], health disorders [], sleep [], an others. Infrastructure sensors such as car sensors or cameras can provide several contexts as well — for examples, cars can provide information about sharp turns, acceleration, and others [].

The ability to utilize contexts from diverse sensors sources can be valuable in several applications. For example, hazardous driving can be caused by driver stress and anxiety, or due to a user speaking on a phone or turning back to look at a co-passenger. In addition to visual contexts such as drowsiness, symptoms of hazardous driving behavior can also be reflected in anomalous driving patterns such as sharp turns. There are also other longer term interactions that influence driving behavior, for example a stressful day or poor sleep prior to driving. Thus, a combining contexts can provide richer understanding of human behavior in CPS settings, enabling more targeted and anticipatory interventions.

Despite the possibilities, there is still a significant gap in the design of an intelligence layer that combines these streams towards enabling smarter CPS applications. Prior work focuses on improving individual context classifiers, or systems-level optimizations (e.g. energy-optimization [**?**], hierarchical triggering [], cloud offload [**?**, **?**, **?**]), but there has been limited work on understanding relations from a variety of on-body and infrastructure sensing context in CPS settings.

In this section, we lay out a research agenda for combining contexts derived from SensEye with information from diverse on-body and infrastructure-mounted sensors sources. We tackle several research questions in this work: How can we leverage contexts obtained from multiple sensors to improve accuracy of detecting events of interest? How can we automatically and in real-time learn predictive behaviors specific to the individual, which can enable early feedback or intervention? What programming model can we provide application designers who wish to develop new interventions using our analytic tools?

Preliminary Approach

Our prior work (PIs Ganesan and Marlin) has looked at learning relations across contexts is by using a Dynamic Bayesian network (DBN), which is a learning framework that describes both temporal and static relationships among random variables at different time instances. This makes it particularly useful for correcting intermittent misclassifications of context inference algorithms as well as exploiting relations across contexts. Figure **??** shows a two-step DBN constructed by combining several contexts from a mobile phone including activity, social, and location contexts. The DBN framework can be used in several ways: 1) It can be used to improve accuracy of context inference by leveraging multiple related context information. For example, in the example, we improved accuracy of activity context sensing by XXX% by leveraging other context information, and 2) It can be used to reduce energy use by identifying the most energy efficient combination of contexts to execute to achieve a certain context sensing accuracy. For example, we showed that even if only two context algorithms execute out of a required set of eleven, the average accuracy is relatively good at 73.4% , and when six contexts can run, the accuracy is at 94.45%

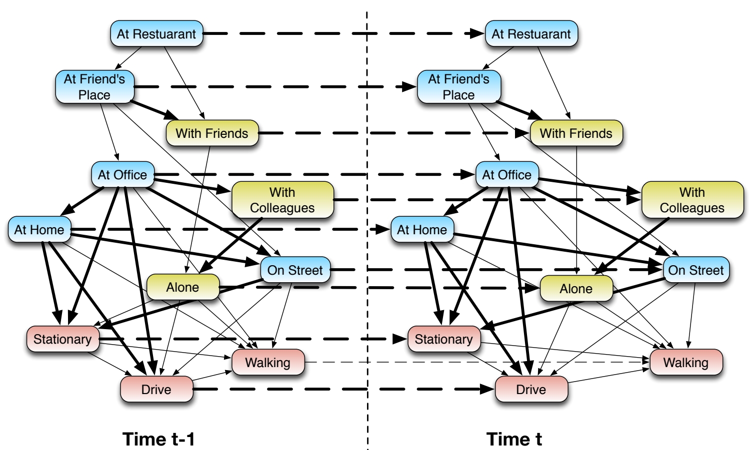

Figure 4: Figure shows a DBN constructed from real user traces — the thickness of each line shows the strength of relationship across the contexts. Dotted line corresponds to temporal relations and solid lines are relations across contexts at each timestep.

Proposed Research

While our preliminary work demonstrates the potential of leveraging context relationships to improve performance, there are significant challenges that remain, particularly when applying these ideas to contexts inferred from SensEye.

Complex accuracy-delay-energy tradeoffs: Performance tradeoffs in a system combining SensEye with chestbands, wristbands, phones, and infrastructure sensors are considerably more complex than the preliminary approach describes. Our initial model does not take into account the energy-accuracy tradeoffs by tuning various knobs provided by the compressive sensing engine and visual context inference algorithms, the energy cost of sensing on different devices, the energy availability on the device and diverse application requirements in terms of accuracy, precision and recall. In addition, delay is a factor that needs to be taken into account for real-time sensing. Delay is incurred at each stage of the sensing, processing, and inference stages — in fact, higher delay can improve accuracy by leveraging temporal models. In addition, accuracy can be traded off for delay and energy by tuning the number of contexts that are active at any given time. Our objective is to formalize these complex interactions between different performance metrics, and design algorithms that can enable tradeoffs across these dimensions.

Temporal multi-sensor patterns: Our example assumed that all contexts were continuously sensed, whereas CPS applications can look for sequential patterns of behavior. For example, in a transportation scenario, an application may pose a query that looks for a continued period of stress in the hours preceding a driving event, followed by aggressive driving behavior while in a car. Such a query needs to run on on-body sensors during the day, and transition to body and car sensors while driving. To save energy, the execution engine may only trigger sensing on the vehicle if stress is detected beforehand; in addition, an execution engine may want to predict likely driving times and only trigger the stress detector a few hours prior to this to further optimize energy. To improve accuracy, the execution engine may take advantage of other contexts that have been identified as relating to high stress or aggressive driving for the individual. Our research will understand how to efficiently execute complex queries that execute over a network of body and infrastructure sensors, and what kind of reliability, accuracy, and delay guarantees can be provided in such a system.

High-level programming abstractions From a programming perspective, our system presents a daunting challenge for a developer who has to take into account extremely challenging resource constraints, complex interactions between contexts, diverse behavioral patterns across individuals, and wide range of application needs. In this work, we propose a high-level programming and execution framework for application developers who wish to leverage SensEye together with other on-body and infrastructure sensors to develop CPS applications. Our goal is to provide an abstraction for applications that hides: a) accuracy-energy-delay tradeoffs in determining how to obtain accurate, real-time context while minimizing energy usage, b) complexity of fusing context information from several distributed on-body and infrastructure sensors, c) personalizing context inference to the individual, and d) dealing with real-world dynamics in network connectivity, sensor placement and availability, sensor uncertainty, and others.

Enabling real-time interventions: On-body sensors and mobile phones also provide a rich opportunity for real-time interventions. One of the key challenges that we face is understanding what type of interventions are most appropriate to change behavior, and how to personalize interventions to the application, individual, and context. For example, possible interventions for hazardous driving behavior may include phone alarms, or a buzzer on SensEye. However, if the cause for erratic driving is more likely to be driver anxiety, then a more appropriate intervention may be soothing music whenever high stress is detected. Our work will define an "intervention" programming framework that enables real-time detection, and a) which intervention is most suitable given the context of the user? b) when to provide the intervention to minimize annoyance while maximizing the likelihood of success, c) how to learn user preferences and personalize interventions, and d) what programming abstractions can be provided to domain experts who want to design interventions.

Handling disconnections: One challenge in our system is that it should be robust to intermittent connectivity across devices. On-body sensor devices are not always worn by a user, for example, SensEye may be worn only while in a car, a chestband might be worn only a few hours a day, and a mobile phone may not always be carried by the user. In addition, devices are likely to be interconnected via a range of technologies — low-power Bluetooth Low Energy (BLE) or ANT+ between on-body sensors and a phone, WiFi between a phone and car sensors, and 3G between the phone and cloud. The heterogeneous connections raise several questions: a) how to enable inference across contexts to proceed despite intermittent connectivity, b) how can we execute a large-scale learning framework (e.g. DBN) in a distributed manner across several partially connected devices, while take advantage of interconnected periods? This research will leverage our prior work on the design of bulk transfer protocols for Delay Tolerant Networks (DTN), Mesh networks, Sensor networks, and RFID-scale devices [].

5 Implementation and Validation

A crucial aspect of our work is real-world experimentation, prototyping, and validation of SensEye. We will target two applications: a) hazardous driving in transportation, with Senior Personnel, Knodler, and b) mental health monitoring, which will leverage ongoing collaborations between PI Ganesan and Dr. Malison at Yale Dept. of Psychiatry (see letter of commitment).

Hazardous driving: Human error is a causative factor in 85% or more of all crashes []. Many traffic accidents are caused by, or at least related to, a driver's mental workload, both when it is either too low due to fatigue, drowsiness or substance use, or too high due to stress, anxiety, or distractions [**?**]. When the driver mental workload is too low, it can lead to reduced alertness and lowered attention, whereas when it is too high, it can lead to distraction, diverted attention and insufficient capacity and time for adequate processing of sensory input [**?**, **?**, **?**, **?**]. High mental workload is also caused by distractions such as cellphone use, conversations, car navigation systems, etc. Mental workload is also known to be related to factors such as age, gender and experience of the driver, for example both advancing age and lack of experience has been found to be related to reduced visual attention, and reduced scanning of roadways.

Our goal is to study the use of SensEye in conjunction with other on-body sensors (chestband, wristband, phone), as well as sensors on the car (via the OBD interface) to detect hazardous driving conditions. Our primary evaluation of SensEye will be using a driving simulator at UMass, where we will use SensEye to monitor the frequency and degree of eyelid opening [**?**], as well as benefits of combining contexts from the different devices. We will compare our approach against existing dashboard-mounted camera-based solutions as well as ground truth.

Mental health: The ability to monitor eye movement patterns of a user of a user has considerable implications on developing targeted interventions (and non-medication) treatments for those with mental illnesses. Research on eye tracking has showed that the manner in which the eyes can follow a target and how well one can pay attention to a task together help to pinpoint risk factors related to schizophrenia [**?**]. They have also been used understand study gaze behaviors in people with Autism or Williams syndrome [**?**], and Alzheimers disease progression [**?**]. Monitoring pupil dilation has been shown to be effective for detecting intoxication and substance abuse, and the use of accelerometers/gyroscopes on SensEye can help detect instability of gait.

Our goal is to understand how SensEye can be used in mental health. PI Ganesan has on-going collaborations with Dr. Malison at Yale Psychiatry as part of a soon-to-be-funded NIH R01 grant on the use of body sensors for understanding addictive behavior. While a true randomized control trial to study the use of SensEye in mental health is outside the scope of this proposal, we will perform preliminary studies to assess how accurately SensEye can estimate metrics of interest to mental health studies (eye tracking, attention duration, pupil dilation, etc). Our preliminary non-clinical evaluation will be performed as part of this grant be a pre-cursor to an NIH proposal for a larger clinical trial.

6 Plan of Work: Team, Schedule and Milestones

Team strengths and synergy We have assembled a strong team, each with a unique complementary expertise, who together cover all aspects of the SensEye research. PI Ganesan has led and participated in several collaborative cross-disciplinary projects that combine hardware, signal processing, learning and systems [**?**]. To facilitate coordination, we propose to hold bi-weekly teleconferences. We will also hold bi-annual meeting of all investigators and students at UMass or at a conference venue (taken into account in the budget).

There is a excellent synergy among the team members. Co-PI Marlin and PI Ganesan have collaborated for the past several months on contexts sensing and inference for mobile sensor data. Co-PI Dutta and PI Ganesan have had extensive interactions over the past several years on various topics in wireless sensor networks. Co-PIs Duarte and Marlin are both new at UMass, and have discussed potential collaborations. In summary, our team although diverse and complementary in areas of expertise has good synergy, as evidenced by our joint effort in putting this proposal together.

Research Task Specific Collaborations Dutta and Ganesan will work on developing the SensEye hardware design, as well as the low-level embedded sensing, processing, and communication architecture. Duarte will work together with Dutta on the design of appropriate compressed sensing techniques for the hardware, and will work with Marlin on the boundary between compressive sensing-based image representation, and scene recognition algorithms. Marlin and Ganesan will work on image representation, context inference; Ganesan will collaborate with Dutta on a programming framework. Ganesan will also work with domain experts Knodler and Malison on gathering/refining domain requirements, and will be responsible for the implementation and real-life testing of SensEye for transportation and mental health applications.

Overall Annual Execution Plan The team will work on designing the SensEye hardware prototype in the first year, and distribute several samples of the prototype across the PIs. In parallel, we will start developing the sensing, estimation, and learning framework that is needed on the software side of our system. In year 2, we will do real-life measurements that integrate compressive signal processing, and start developing the visual context sensing pipeline with SensEye. We will also develop context inference techniques that combine visual contexts with other contexts. In year 3, we will focus on real-life testing at scale, results of which will help us validate our ideas and uncover issues that appear at large scale. Several students will be involved in these efforts across the two institutions — two students on developing SensEye hardware architecture, one on compressive signal processing, one on visual context sensing, one on context inference, one on the programming framework and systems optimizations, and one focused on applying SensEye to transportation and health.

7 Broader Impact

There is tremendous excitement surrounding the use of on-body sensing devices for a spectrum of applications ranging from behavioral science, fitness, remote healthcare, and others. But continuous monitoring of the eye and field-of-view remains a key missing piece of the sensing puzzle, and significant advances in health and safety-critical applications are possible through such technology.

Applications and Deliverables We propose to release a family of open-source SensEye platforms and software systems with different choices of cameras, computation, communication, storage, and tethering choices. Our software stack will include all the layers that we have discussed in this research, integrated for easy use. Our team has considerable track record in releasing such systems — Dutta has released the Epic Mote family of sensor platforms that is widely used in academia, Ganesan's group has released several software frameworks (e.g. Capsule [**?**], Hop [**?**]) for the broader research community, Duarte has embraced the concept of reproducible research, which refers to the idea that the product of algorithmic research to be shared with the community should include both research papers and the full computational resources (code and data) used to produce the results in the paper. Similarly, Marlin has released much of the code and data used in his research papers. Thus, all PIs have a strong record in dissemination of hardware, software, and datasets.

Outreach to under-represented minorities and women: The PIs will involve under-represented minorities and women in several ways. The PIs support several undergraduate students and women in their research. PI Ganesan recently ran a workshop on wearable sensor technology for female undergraduate students to introduce them to interesting ways in which computing and crafts can merge. Further, the research and teaching that will flow from the project will impact students across the Five Colleges, which includes two all-women's colleges, Smith and Mount Holyoke, whose students can enroll in courses taught at UMass.

The proposed work will broaden participation of underrepresented groups. In years 2 and 3, we will work with the STEM program at the two universities, and the Commonwealth Alliance for IT Education (CAITE) at UMass, to distribute SensEye to a randomized set of students in high school classrooms. The students will be asked to design an experiment with a clear hypothesis, plan an experimental methodology that incorporates SensEye and develop a simple web page that can display the results. The goal of this effort is to inculcate scientific, hypothesis-driven thinking as well as exposure to sensors, phones, and web programming. The program will be evaluated by comparing the self-reported post-secondary STEM educational plans of the students vs a non-participating control group at the time of high school graduation.

Interdisciplinary research: Students from Engineering and Computer Science will not only learn the hardware design, signal processing, computer vision, embedded systems approaches, and energy management techniques for on-body sensors, but also understand human behavior in real-world settings. The specific outcome of our research will be a family of SensEye hardware platforms, image and sensor data processing stack, a context learning and inference engine, and a programming framework. Since our prototypes and reference implementations will be made available to the community, researchers will directly benefit from them. We expect our work to lead to an easy-to-deploy visual context sensing system that will be widely used in academia and industry, and influence future designs.

Industry We also plan to work with several industry partners in furthering this technology. UMass works closely with Applied Science Laboratories, the leading eye tracking hardware manufacturer in the US, and we will have greater collaboration with them. Co-PI Dutta also has made his hardware platforms available through online electronic hobby stores (e.g. HiJack [**?**] on Seeed Studio [**?**]), and we will make SensEye available for the broader community through such mechanisms.

8 Curriculum Development Activities

Our project integrates research and education of graduate and undergraduate students through the close interaction and mentoring of students by project faculty. Ganesan has presented on body sensors and smartphone-based sensing as part of short courses for high-schoolers as part of science fairs organized for high-schoolers. PI Duarte has also co-developed a K-12 educational program component with colleagues at UMass. The College of Engineering at UMass, in collaboration with local school districts from Springfield, Holyoke, Mahar, and Greenfield, is currently preparing an NSF Research Experience for Teachers (RET) proposal. High school teachers will spend four days a week for six to eight weeks at UMass, where they will assist with the ongoing research and create an educational module on compressive sensing (CS) concepts to engage high school math and science students. Finally, SensEye will be incorporated into undergraduate and graduate embedded systems courses at both institutions, collaborating on educational materials will an important part of improving the quality of those courses. As the proposed research stretches from hardware to algorithms to applications, this will enable an iterative synthesis. For example, students at Michigan can have projects on extending and designing the hardware platform based on efforts from signal processing and machine learning students at UMass who will design algorithms for extracting new visual contexts.

9 Most Relevant Prior NSF Support

A framework for the real-time sensing of visual context in natural environments has not been comprehensively addressed in past work, and none of our current or past projects overlap with the proposed one. We list selected past awards in the following that are most relevant to this project.

Ganesan is Co-PI on *NetSE: Large: Collaborative Research: FieldStream: Network Data Services for Exposure Biology Studies in Natural Environments* This project explores the design of a body sensing system to provide long term unattended collection of objective, continuous, and reliable physiological/behavioral data from natural environments that can be used for conducting population based scientific studies.

Dutta is a Co-PI on *CSR: Large: Collaborative Research: Integrating Circuits, Sensing, and Software to Realize the Cubic-mm Computing Class*. This project has resulted in a one cubic millimeter sensor node, that integrates a 96x96 pixel imager, UWB communication, processor, memory, thin-film battery, and energy harvester.

Marlin and Duarte are new faculty at UMass Amherst and do not yet have NSF support.

Data Management Plan

Our team has considerable track record in releasing open-source systems and datasets— Dutta has released the Epic Mote family of sensor platforms that is widely used in academia and industry [?, ?]. Ganesan's group has released several software frameworks (e.g. Capsule [], Hop []) and hardware platforms (design files for a NAND flash board for Motes [?]), and datasets (e.g. traces of solar harvesting-based sensors at [?]) for the broader research community. Duarte has created the Rice Compressive Sensing Repository and has embraced the concept of reproducible research, which refers to the idea that the product of algorithmic research to be shared with the community should include both research papers and the full computational resources (code and data) used to produce the results in the paper. Similarly, Marlin has released much of the code and data used in his research papers. Thus, all PIs have a strong record in dissemination of hardware, software, and datasets, and we will do the same with this project. We outline the main data management aspects in this section — all of our data sharing will be co-ordinated with the CPS Virtual Organization (VO).

Types of Data to be Collected

The data obtained during the proposed project will consist of measurement records of (i) sensor data captured by SensEye such as images, video, accelerometer traces, (ii) context information from other sensors (phone, wristbands, etc) in conjunction with data from SensEye, (iii) performance statistics of SensEye in terms of communication to a phone, real-time actuation, etc, as described in the main body of the proposal. These records will consist chiefly of time-series sensor datasets that will be recorded, with essential metadata present either as headers in the relevant electronic files, or included along with the laboratory notebook narrative or README file.

Data and Metadata Standards

The data and metadata will use open formats with free tools for viewing and manipulation whenever practical. Where free tools or open formats are either not customarily used (e.g. PCB design) or where no such tools exist, the data and metadata will use low-cost proprietary options. In general, however, the data and metadata will be shared as text files in CSV, TSV, or similar format. The data processing scripts will be written in a language with a freely-downloadable runtime (e.g. Bash, Perl, or Python). Using the scripts will require a person to have basic Unix literacy and programming skills. Figures will be generated using open tools whenever possible (e.g. gnuplot, R) but proprietary software may be used if models (e.g. SPICE) or algorithms (e.g. spectrogram) are only available in proprietary tools (e.g. LTSpice, MATLAB). Some contextual metadata will be shared as figures (e.g. sensor placement relative to a window, drawing of an experimental setup, etc.) These figures will be made available in open formats like PNG and JPG, or in some cases as PDF files. Design files will be shared as text files where possible (e.g. Assembly, C, nesC, PHP, Python, HTML, Javascript, etc. for software, SPICE netlists for circuit simulation, and GERBER photoplots for circuit board fabrication). Circuit schematics and circuit board designs will use proprietary formats and will require commercial software to view and modify them. To the extent possible, the project will use tools and formats for which inexpensive, free, or trial versions of software is available (e.g. Cadsoft EAGLE schematic capture and circuit board layout).

Policies for Access and Sharing

Data, metadata, and design files will be made publicly accessible by posting them to the project homepage and contributing them to public repositories where appropriate. Data and design files will be released contemporaneously with publications that reference them.

Privacy, Confidentiality, Security, and IP Protection

The project will involve human subjects data for our real-world validation studies in CPS application settings such as traffic simulators or real driving conditions. This data will be made available after careful anonymization. The data sets will be free of IP restrictions and they will not be covered by copyright. The hardware designs will be the property of the Regents of the University of Michigan or University of Massachusetts but the Universities offer PIs latitude in whether to file for patent rights or release their work under a permissive open source license. This project plans to release data and design files under an open source license. The data sets and designs will be released along with any papers that use or reference them on the project home page. Where results *using* proprietary models are necessary (e.g. thin-film Lithium batteries from Infinite Power Solutions), this will be noted, and the non-proprietary aspects will be released.

Policies and Provisions for Reuse and Redistribution

There are no restrictions on data distribution; the data may be distributed freely. The sensor data sets and hardware/software designs will be of interest to body sensing and platform designers in both academia and industry. They will use the design files to create derivative works that are tailored to their own project needs. There are no forseeable reasons to avoid sharing the data that this project generates.

Plans for Archiving and Preservation of Access

The long-term strategy (beyond the scope of this project) for maintaining, curating, and archiving the data, designs, and other artifacts is to contribute them to public repositories. The data and metadata will be contributed to the CRAWDAD data collection archive which is hosted at Dartmouth College (`http://www.crawdad.org`). The software and hardware designs will be contributed to the Google Code repository hosted by Google (`http://code.google.com`). In addition, we will consult the CPS Virtual Organization (VO) to see if alternate sharing mechanisms are in place. The data to be collected will not require transformation, cleaning, or anonymization. The data and designs will include metadata describing the experimental setup, methodology, data types, and other relevant metadata. The data will include a README file with links to any reports or papers generated using it. The data and designs will have a useful life of five years beyond the end of the project. The data and design files will be mirrored on the project homepage and backed up regularly for the duration of this project using a Departmental service available to faculty and students.