

SPIDERS+: A Light-Weight, Wireless, and Low-Cost Glasses-Based Wearable Platform for Emotion Sensing and Bio-Signal Acquisition

Jingping Nie, Yanchen Liu, Yigong Hu, Yuanyuting Wang, Stephen Xia,
Matthias Preindl, and Xiaofan Jiang

Department of Electrical Engineering, Columbia University

Abstract

We present a System for Processing In-situ Bio-signal Data for Emotion Recognition and Sensing (SPIDERS+) – a low-cost, wireless, glasses-based platform for continuous in-situ monitoring of user’s facial expressions (apparent emotions) and real emotions. We present algorithms to provide four core functions to detect eye shape and eyebrow movements, pupillometry, zygomaticus muscle movements, and head movements, using the bio-signals acquired from three non-contact sensors (IR camera, proximity sensor, and IMU). SPIDERS+ distinguishes between different classes of apparent and real emotion states based on the aforementioned four bio-signals. We prototype advanced functionalities including facial expression detection and real emotion classification with a facial expression detector based on landmarks and optical flow that leverages changes in a user’s eyebrows and eye shapes to achieve the average true positive rate up to 83.87%, as well as a pupillometry-based real emotion classifier with higher accuracy than existing low-cost wearable platforms that use sensors requiring skin contact. In addition, SPIDERS+ also includes optional contact-based sensors such as photoplethysmography (PPG) and multi-channel electroencephalogram (EEG) to monitor other bio-signals. SPIDERS+ costs less than \$20 to assemble and can continuously run for up to 9 hours before recharging. We design, implement and evaluate accompanying desktop and smartphone applications to demonstrate that SPIDERS+ is a truly wireless and portable platform that can impact a wide range of applications, where the knowledge of the user’s emotional state is critical.

*Parts of this paper have been previously published in the Proceedings of the 2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI) and a preliminary version of the system was described in a demo abstract at the 2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN) [1, 2].

Email address: {jn2551, yl4189, yh3104, yw3241, sx2194, mp3501}@columbia.edu, jiang@ee.columbia.edu (Jingping Nie, Yanchen Liu, Yigong Hu, Yuanyuting Wang, Stephen Xia, Matthias Preindl, and Xiaofan Jiang)

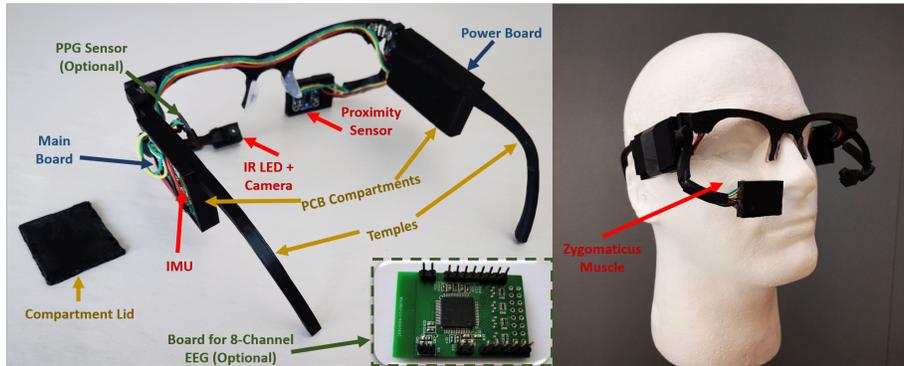


Figure 1: Left: The SPIDERS+ glasses platform including hardware sensors (red), PCB boards (blue), frame components (yellow), and optional attachments (green). Right: SPIDERS+ glasses on a mannequin.

1. Introduction

Intelligent emotion monitoring systems play an important role in improving the mental health conditions of the general public. They help inform the general population of their own mental states. They also provide psychiatrists and psychotherapists a means to monitor and diagnose their patients' mental health problems remotely. This is particularly critical during the current COVID-19 global pandemic, as it is difficult to schedule in-person and mask-off doctor visits. Emotion monitoring platforms also have immense impacts on businesses, providing companies with another measure of the effectiveness of their products and advertisements. Additionally, a low-cost and portable solution would enhance the lives of the general population by bringing, for example, enhanced entertainment experiences when movies/games/tv shows present dynamic content to users based on emotions experienced by the user in real-time and affordable mental health services to people who have access and cost barriers to mental health care [3, 4, 5, 6].

However, predicting emotions is challenging, let alone detecting emotions on a resource-constrained and wearable device. There are commercial products and research platforms developed for monitoring specific signals that provide insight into a user's real emotion. However, many of the existing commercial products and sensors such as EEG caps are bulky, expensive, wired, and require constant skin contact, limiting their use to clinical or controlled environments. All of the existing vision-based approaches and products for detecting facial expressions utilize static features found within a single image to detect whether a person is smiling, sad, disgusted, etc. Whereas, each person's eye and facial features, such as skin tone and eye shape, vary widely. This makes it difficult to create an algorithm or train a classifier that can generalize to unseen examples well.

Additionally, it is even more challenging to implement semi-real-time vision-based facial expression detectors on a resource-constrained, wearable, embedded

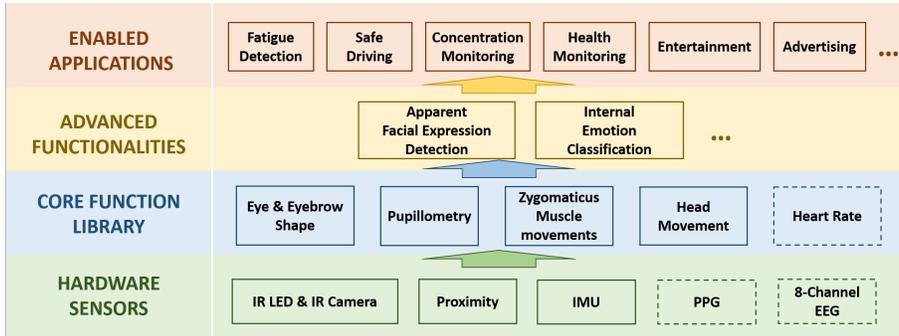


Figure 2: The four-layer system architecture of SPIDERS+. We designed the architecture in a top-down approach based on the types of applications that SPIDERS+ aims to impact. All components and layers in SPIDERS+ work towards enabling or improving the top layer: *Enabled Applications*.

platform. We observe that the facial and eye movements made by a person
 30 as he/she transitions into a new expression is more consistent across different
 demographics, and we leverage this information to develop our facial expression
 (apparent emotion) algorithm.

Pupillometry has been shown to be a strong indicator of real emotions [7,
 8, 9]. However, even the state-of-art algorithms and systems that utilize pupil-
 35 lommetry, such as the *Tobii Pro Glasses 2*, cannot reliably predict real emotions
 due to the inherent measurement noise caused by lighting changes, reflections,
 rapid eye movements, blinking, etc. As a result, in addition to real emotions,
 we provide facial expression (apparent emotion) as a direct output to support
 applications where apparent emotions are sufficient or preferred.

40 Our contributions in this paper are summarized below:

- We present SPIDERS+, a low-cost, energy-efficient, and portable wearable
 platform (see Figure 1) that collects data from a variety of non-contact
 sensors to detect: 1) eye & eyebrow shape, 2) pupillometry, 3) zygomaticus
 45 muscle (smile muscle) movements and 4) head movements. SPIDERS+
 can also be extended to include functionalities provided by contact-based
 sensors, such as 5) heart rate and 6) multi-channel EEG monitoring. We
 demonstrate the robustness of our sensing modalities by comparing against
 commercial products.
- We present novel and robust vision-based techniques to perform pupillom-
 50 etry and eye & eyebrow shape detection using infrared (IR) band gray-
 scale images, from an IR camera positioned at a low angle from the eye
 as to not block the user’s field of view.
- We develop and implement a data processing pipeline that intelligently
 55 partitions tasks and computation across the glasses wearable and more
 powerful processing units (e.g., computer, cloud, smartphone) to maintain

high accuracy, long battery life, and to ensure that our emotion recognition algorithms run in semi-real-time. We demonstrate that our wearable is able to last up to 9 hours.

- 60 • We propose a novel optical-flow based algorithm for estimating facial expressions (apparent emotions) leveraging movements in the user’s eyes as he/she transitions into a new facial expression. We experimentally show that our algorithm achieves an average classification accuracy of 83.87%, which outperforms the state-of-the-art approach (e.g., [6]). We also demonstrate SPIDERS+’s capability of estimating a user’s affective state (real emotions), achieving a classification accuracy of 49.32%, which
65 outperforms the results presented in [10].
- We demonstrate that SPIDERS+ can estimate user facial expressions and stream the results in semi-real-time to user-facing desktop and smartphone applications via WiFi. This end-to-end pipeline can enable a wide range of
70 user applications such as fatigue detection and mental health monitoring.
- We provide the community with 1) an anonymized dataset of biometrics measured from subjects during our controlled experiments; 2) the circuit design and 3D model design for our eyeglasses wearable; and 3) a suite of algorithms and libraries that enable everyone to easily build their own
75 emotion-based applications based on SPIDERS+.

The rest of this paper is organized as follows. Section 2 presents related work and discusses limitations in current works and products that SPIDERS+ addresses. Section 3 details the sensing modalities included in SPIDERS+ and the hardware platform. Section 4 describes the library of core functions, including the eye and eyebrow shape detection, pupillometry, and zygomaticus
80 movements, that SPIDERS+ enables. Section 5 prototypes and evaluates the facial expressions (apparent emotion) detection and real emotion classification. Section 6 introduces the potential applications SPIDERS+ could enable. Section 7 illustrates basic smartphone and desktop applications for visualizing the semi-real-time results of SPIDERS+. Section 8 discusses limitations and future
85 work, and Section 9 concludes the paper.

2. Related Work

In this section, we first analyze related commercial products that track physiological data in order to identify issues in price, portability, and usefulness. We
90 then analyze different modalities of data that would be useful for emotion classification purposes and explore algorithms that derive the types of data that cannot be directly measured by minimal sensors.

2.1. Bio-signal Acquisition by Wearable Devices

There are a variety of commercial bio-signal acquisition devices on the market, but all have specific problems that make them unsuitable candidates for a
95

low-cost, energy-efficient, and portable platform for monitoring user emotion. For instance, *Fitbit* is a wireless wearable activity tracker, but only measures three-axis acceleration data, and photoplethysmography (PPG), both of which are not reliable measures of emotional states by themselves [11]. *Muse* is a head-
100 band that provides five-channel electroencephalogram (EEG) data, heart-rate, and gyroscope sensors [12]. Both heart-rate and EEG sensors require skin contact, which is difficult to maintain in a mobile wearable where the user may be on the move. Pupillometry is well-known as an indicator of affective states, and there are eye-tracking cameras on the market that measure these signals. For
105 example, the *Tobii Pro Glasses 2* leverages a wide-angle camera, microphone, gyroscope, and accelerometer to track the wearer’s eye [13]. However, *Tobii Pro Glasses 2* system is not truly wireless (requires an external recording unit) and very expensive (\$10,000+). To summarize, commercial products measure a limited amount of bio-signals that often require direct skin contact, or are very
110 expensive and not portable.

There are a number of research platforms and glasses developed to provide a wearable platform for emotion monitoring. However, most of these systems focus exclusively on pupillometry [14, 15, 16], but not apparent emotion. The second category of works leverages photo reflective IR sensors integrated within
115 the eyeglasses frames to detect facial expressions [17, 18]. However, using IR sensors provides limited resolution about the position of the cheeks/brows and no pupillometry measurements, making it difficult to perform robust facial expression recognition or real emotion classification across a varied population. [10] leverages sensor fusion, combining facial features extracted from a camera
120 pointed at the user’s eye and other physiological signals, such as skin conductance and PPG, to obtain a more accurate estimation of real emotions. However, we recognize that facial expressions are not reliable indicators of real emotions; for example, a user can easily fake a smile even if s/he is depressed. Additionally, measuring skin conductance and PPG requires direct contact between the
125 skin and sensor, which is not stable if the wearer is on-the-move.

2.2. Vision-Based Apparent Emotion Classification

Research about facial expression classification based on images of the whole face is abundant. Nevertheless, these approaches require continuous capturing user’s entire face with a camera looking from a distance, which is not portable
130 and can lead to privacy concerns. On the other hand, wearable devices are becoming more and more popular, and provide in-situ sensing capabilities [19, 20, 21, 22]. A camera mounted on the glasses can not only sense the pupil size and gaze angle, but can also clearly see the eyelids and eyebrows. Previous work directly uses raw images as an input to a machine learning classifier [10].
135 However, directly learning the features with raw images requires a large amount of labeled data and can perform well within the dataset and perform poorly in real-world cases.

For eyelid and eyebrow detection, many previous works utilize RGB images, conducting adaptive thresholding, luminance valley detection, and edge detection
140 to detect the shape of the eyelids and eyebrows [23, 24, 25]. However,

these methods are computationally expensive and not robust. Also, in order to perform pupillometry, we have to use IR cameras that produce gray-scale images, making the boundary between the sclera (white outer layer of the eye) and skin harder to distinguish. Recent works also focus on leveraging IR cameras to conduct deep-learning-network-based facial expression classification on resource-constrained platforms. However, the hardware form factors of many of these systems are bulky and not comfortable for daily wear [26].

Convolutional neural networks (CNN) and regression tree models have been commonly used for generating facial landmarks for facial expression recognition. These methods estimate facial landmark points that draw out the location and curvature of the eyelids and brows [27, 28], and some of these methods have been shown to run in real-time [29]. As such, we decide to adopt a similar approach in landmark generation. However, generating the shapes of the eyelids and brows and determining the user’s expression and apparent emotions based on this information are two different problems, the latter of which is not addressed by previous works.

2.3. Real Emotion Recognition

A variety of bio-signals have been explored for recognizing a person’s real emotional state. Pupillometry is one of the most popular signals used to classify real emotional states due to its non-invasive nature and its high discriminatory power between different classes of real emotion. To achieve high accuracy in classifying arousal and valence levels (two measures of real emotions), the majority of works in the literature acquire the pupillometry and gaze information from commercial eye trackers that are expensive and not truly wireless [30, 8, 31].

Other bio-signals that have been linked to real emotions are heart rate and EEG. Long-term heart rate variability (HRV) has been shown to be correlated to emotional patterns and shifts [32][33]. EEG head caps are also widely used in detecting emotion because it non-intrusively captures brain signals [34, 35, 36]. However both heart rate monitoring and EEG monitoring require sensors to be in direct contact with the skin; in a wearable system, where a user could be moving, maintaining contact is not guaranteed nor very comfortable for the wearer. Additionally, EEG-based methods have achieved high-performance rates because of the sheer number of channels, employed in these works (often 32 or more). It is difficult to incorporate even a few channels of EEG into a resource-constrained wearable.

3. System Design

SPIDERS+ is a low-cost, long-lasting, and mobile platform for monitoring apparent and real emotions of the wearer. In this section, we present our hardware design and data processing architecture that allows SPIDERS+ to achieve robust apparent and real emotional recognition while being energy-efficient and maintaining a suitable latency.

Figure 2 shows the functional system architecture of SPIDERS+. SPIDERS+ focuses on prototyping and enabling its functionalities and applications based on inputs from non-contact sensors, while also providing extended features and functionalities by enabling more intrusive, but optional, contact-based sensors. Each layer in the architecture directly enables the applications or functionalities in the above layers. We take a top-down approach in designing SPIDERS+ based on the types of applications we aim to address. In particular, the *Hardware Sensors* layer, consists of a wide range of contact-less and optional contact-based sensors that SPIDERS+ supports. SPIDERS+ uses outputs from its *Hardware Sensors* to extract a common set of bio-signal measurements, such as pupillometry and eye shape, that form the *Core Function Library*. The measurements that are extracted in the *Core Function Library* are all bio-signals that are correlated with a user’s emotional state. These features are then utilized by the *Advanced Functionalities* layer, that runs novel machine learning and signal processing algorithms (Section 5), to estimate aspects of user emotion, including facial expressions and internal emotion. Outputs from the *Advanced Functionalities* layer are also the outputs of SPIDERS+, but can be used to improve and enable a multitude of applications as illustrated in Section 6.

There are many applications and problems that untethered emotion monitoring could enable; our aim is to provide the community with a platform that enables emotion monitoring and allows researchers to develop their own algorithms to address a wide range of applications by leveraging the plethora of sensors and modalities that SPIDERS+ provides. In addition, we recognize that there are two facets of emotions that people exhibit, apparent and real emotion, provided at the *Advanced Functionalities* layer. By providing a mobile platform that is capable of measuring these two dimensions of emotion, we believe that we can impact some of the *Enabled Applications* shown in Figure 2. We also develop user-facing desktop and smartphone applications that leverage outputs from SPIDERS+’s *Apparent Facial Expression Detection* module to enable applications such as real-time apparent emotion monitoring and fatigue detection.

3.1. Sensing Modalities Overview

We now discuss the types of bio-signals incorporated in SPIDERS+ based on what is needed to achieve robust apparent and real emotion recognition. Pupillometry, EEG, HRV, skin conductance, and facial expressions are commonly used to determine a person’s emotional state, as discussed in Section 2. Previous works generally leverage all or a subset of these signals to classify emotional states. First, we recognize that there are two facets of emotion that people exhibit: apparent and real emotion, and that knowing one does not necessarily provide the other. For example, people may hide their internal feelings of sadness by smiling if he/she is celebrating or spending time with friends and family.

Common signals used to determine someone’s real emotions include pupillometry, skin conductance, EEG, and HRV. We omit facial expressions, as peo-

ple’s expressions may not accurately represent their real emotions. Skin conductance, EEG, and HRV are all commonly measured using sensors that require skin contact. However, in a mobile wearable system, the contact between the sensor and skin can often be unstable, making these signals difficult to acquire. Pupillometry has been shown to be powerful in discriminating between different emotional states and can be measured using a camera, which is non-contact. The IR-filtered camera coupled with an IR illuminator usually has better performance in conducting pupillometry than the commonly used RGB cameras. Hence, we integrate an IR-band gray-scale camera into SPIDERS+ to perform pupillometry to classify real emotion. Our proposed methods for obtaining robust pupillometry in face of these challenges are described in Section 4.2.

Facial expressions are representative of a person’s apparent emotion, or the emotion that people express through visible body language. Facial expressions, head movements, and zygomaticus muscle movements can be adequately captured using a camera, inertial measurement units (IMU), and proximity sensors pointed at a person’s face. Both are non-contact sensors, and proximity sensors have negligible power consumption when compared to that of a camera. Our novel algorithms for determining facial expressions and apparent emotion using signals from these sensors are detailed in Section 4.1 and Section 5.

A summary of the different bio-signals that SPIDERS+ provides is shown in the *Core Function Library* layer in Figure 2. A summary of the different sensors integrated into SPIDERS+ to enable the measurement of these bio-signals can be found in the *Hardware Sensors* layer.

3.2. Hardware Platform

In this section, we describe the hardware platform implementation of SPIDERS+ that allows us to perform robust apparent emotion detection and real emotion classification efficiently and with low latency. Additionally, we have open-sourced our glasses designs, code, and assembly instructions to the community to allow anyone to build upon or use SPIDERS+ for their own projects.¹

3.2.1. Processing and Wireless Communication Components

The system on a chip (SoC) module is responsible for sampling and transmitting sensor data from the glasses wearable to the client application. A microcontroller with proper performance is required to sample and transmit data from all the sensors with enough temporal resolution. To minimize the PCB layout area, we also prefer wireless SoC modules with an integrated radio module. Based on these requirements, we chose the Espressif ESP32 as our processor and wireless transmission module, which supports both WiFi and Bluetooth communication protocols, computational power at low power consumption, and direct memory access (DMA).

¹<https://github.com/Columbia-ICSL/SPIDERS>

3.2.2. Hardware Sensors

The camera sensor captures eye & eyebrow shape as well as pupillometry. We chose the OV2640 camera module because of its high sampling rate and moderate pixel resolution (320×240). Additionally, the OV2640 module performs
270 on-chip JPEG encoding, converting pixels into a machine-readable format without computation required from the MCU. We also install an 850 nm band-pass IR filter and two IR LEDs in front of the camera lens to illuminate the areas directly in front of the camera and to eliminate the influence of different iris colors and ambient light from our images.

275 Compared to the camera, the proximity sensor and IMU are relatively compact and low-power. We decided to use a VL6180X proximity sensor to measure the distance between the sensor and zygomaticus muscle (smile muscle) and an MPU6050 gyroscope and accelerator sensor to detect the head movements.

In addition to the non-contact sensors that are used to extract features to
280 estimate facial expressions and internal emotions in the *Advanced Functionalities* layer, SPIDERS+'s hardware platform also supports several optional contact-based sensors to measure heart rate and EEG. In particular, SPIDERS+ uses the MAX30102 photoplethysmography (PPG) sensor for heart rate detection and also incorporates the TI ADS1298, a low-power, high accuracy 8-channel ADC
285 for sampling EEG signals. SPIDERS+ is capable of maintaining a sampling rate of 250 Hz across all 8 EEG channels, even while sampling from all other sensors on the platform. This sampling rate is comparable to the Muse EEG headband [12]. The sampling rate is further increased to more than 2,400 Hz if other sensors are not being used. Due to the high sampling granularity of
290 our sensors, SPIDERS+ is capable of supporting numerous advanced research projects in neuroscience.

3.2.3. Customized PCB Board

To ensure that all of our circuitry could fit onto a single glasses frame, we separate the whole system into three PCBs: mainboard, biopotential measuring
295 board, and power board. Each PCB is fitted onto a different part of the glasses. The mainboard features an ESP32 module and an MPU6050 gyroscope-accelerometer sensor. The other sensors listed in Figure 1 are not mounted on the mainboard as they measure signals from specific areas of the face (e.g. the IR camera is placed just under the eye to monitor eye and eyebrow movements).
300 The EEG board features the ADS 1298 sensor and its supporting circuitry and can be easily stacked with the mainboard. Figure 3 shows the layout of the mainboard as well as the optional biopotential measuring board.

3.2.4. Eyeglasses Frame Design and Sensor Placement

In order to make the circuit board compact to fit onto the glass frame, we
305 separate the whole system into two miniature PCBs: the mainboard and the power board, as illustrated in Figure 1. The mainboard features our processing and wireless transmission SoC (ESP32-Wrover), as well as the gyroscope and accelerometer sensor (MPU6050). The power board houses a 2000 mAh lithium polymer battery to power the system.

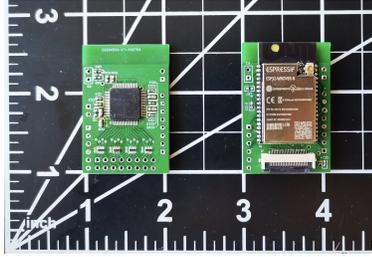


Figure 3: PCB layout of the mainboard (right) and biopotential measurement board (left)



Figure 4: Thermal image of SPIDERS+ in the active mode.

Table 1: Price Breakdown and Power Consumption

Module	Unit Price (Retail) [U\$]	Unit Price (Wholesale) [U\$]	Active Power	Sleep Power
MCU (ESP-32)	4.80	4.80	128.00 mW	2.64 mW
ESP32 WiFi Connection	N.A.	N.A.	286.40 mW	0
Proximity Sensor (VL6180X)	4.50	2.18	82.10 mW	3.30 μ W
IMU (MPU-6050)	7.28	3.52	16.30 mW	16.50 μ W
CAMERA (OV2640)	8.50	8.50	168.00 mW	1.98 mW
Voltage Regulators (TPS63031DSKR)	1.87	0.79	N.A.	N.A.
Glasses (PLA Filament)	0.01	0.01	N.A.	N.A.
PPG (MAX30102) (optional)	8.60	4.75	66.6 mW	2.31 μ W
EEG (ADS1298) (optional)	36.91	29.9	12.5 mW	20 μ W
Total (w/o optional contact-based sensors)	26.96	19.80	730.30 mW	4.6398 mW
Total (all)	72.47	54.45	809.4 mW	4.66121 mW

310 The assembled SPIDERS+ glasses platform is shown in Figure 1. The frame itself was 3D printed, and we use 6 mm wide metal hinges to join the glasses temples to the frontal frame to allow for foldability just like a normal pair of glasses. We include two compartments, which we labeled *PCB compartments*, on both of the temples to house the mainboard and power board.

315 The proximity sensor is placed below the right lens, pointing towards the left cheek (zygomaticus muscle) of the user. Placing the proximity sensor in this location allows SPIDERS+ to detect movements of the cheek muscle as a person changes facial expressions. The camera module + IR lights are placed in a similar position below the left lens and angled upwards to capture images of the user’s left eye and brow. The two extensions where the sensors are mounted have a certain degree of freedom to rotate, which enables customized sensor placement.

320 All boards and sensors are connected with wires running along with the frame of the glasses. The total weight of SPIDERS+, with all boards, sensors, and battery installed, is below 100 g, which is around a quarter of the weight of the Tobii Pro Glasses 2.

3.3. Price Breakdown

330 Table 1 summarizes the retail unit price and wholesale unit price for each major hardware component integrated into SPIDERS+. All software packages and programs used to develop SPIDERS+ is freely available at no cost. The total price to produce a single SPIDERS+ glasses hardware platform is as low

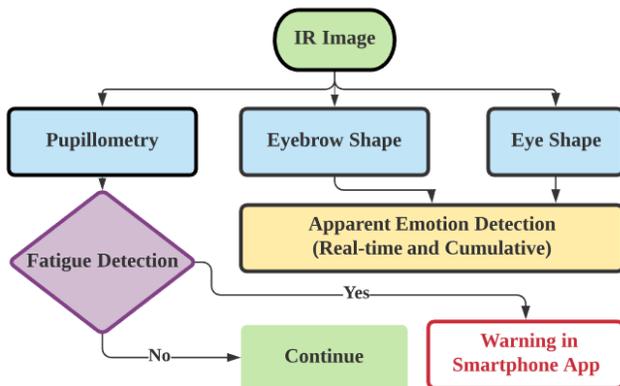


Figure 5: The flow diagram of the functions in SPIDERS+ client application and the user-facing smartphone/desktop application.

as \$19.80, which is low-cost. Adding in the optional contact-based sensors increases the price to around \$50 under mass production. Compared to existing commercial products like the Tobii Pro Glasses 2, SPIDERS+ is orders of magnitude less expensive.

3.4. Power Consumption

The active and idle power consumption for all major components of SPIDERS+ are also included in Table 1. When sampling and transmitting data from all sensors, SPIDERS+ consumes around 759.9 mW and 839 mW with and without the optional contact-based sensors, respectively. Powering the total system using a 2000 mAh battery allows SPIDERS+ to continuously run for more than 9 hours in a battery life test. Additionally, we allow users to turn off specific sensors when not needed to further reduce power consumption. Turning all sensors and the MCU processing unit to the sleep mode consumes less than 5 mW of energy.

3.5. Heat Dissipation and Comfort

To verify that SPIDERS+'s temperature would not cause discomfort for the wearer, we measure the heat dissipation of the entire system for one hour. To measure the temperature of the mainboard (the area with the highest temperature on SPIDERS+) platform during one hour of continuous operation, we use the *Fluke TIS75 Thermal Imager*. In the steady-state, the highest temperature of the inward-facing side of the frame is approximately 35°C, which is close to the normal body temperature of a person. The thermal images shown in Figure 4 suggests that the heat is mostly generated by the wireless MCU and the camera module. Since the temperature of SPIDERS+ is no higher than the human body temperature, the user would not feel any discomfort from heat dissipation.

Table 2: Expressions and eye area features.

Platform	Eye & Eyebrow Shape	Pupillometry	Zygomaticus Muscle Movements	Heart Rate	Head Movements	EEG Acquisition	Price [US]
Fitbit Inspire 2				✓			99.00
Muse 2				✓	✓		249.00
Tobii Pro Glasses 2		✓				✓	> 10,000.00
SPIDERS+	✓	✓		✓	✓	✓	19.80

3.6. System Pipeline and Typical Usage

SPIDERS+ senses a wide range of bio-signals using multiple sensors, as summarized in Figure 2. Processing all of this data on a resource-constrained wearable to perform robust apparent and internal emotion classification, while maintaining semi-real-time and low power consumption, is not possible. Additionally, we propose novel algorithms for extracting pupillometry and facial expressions, which are detailed in Section 5. These algorithms, including a facial expression detector that captures eye sizes and movements in the eyebrows, as a person changes expressions to more robustly classify apparent emotion, are too expensive to compute on the SPIDERS+ glasses efficiently and in real-time. As such, we offload all of our computation, via WiFi, onto the SPIDERS+ client application running on an internet-connected desktop server.

The client application computes the functions listed in *Core Function Library* and *Advanced Functionalities* layer and stores bio-signal data extracted from all of the sensors on the SPIDERS+ glasses platform. We also deploy user-facing desktop/smartphone applications to further process outputs from the client application, depending on the the application. Specifically, the desktop/smartphone applications not only display the outputs from the client application, but also leverage these outputs to conduct fatigue detection. In addition, users are able to set alarms and view results from the fatigue detector in their smartphone and desktop applications. The client application to user-facing smartphone/desktop application flow is shown in Figure 5.

The user-facing desktop/smartphone application displays a user panel that streams a semi-real-time visualization of some of the outputs of the *Core Function Library* and *Advanced Functionalities* layers provided by the client application. The user panel also presents information on detected categories of facial expressions, mental health, and fatigue. The smartphone application will also send an alert to the user if it detects the user is fatigued. This is critical in scenarios where the users' attention is essential (e.g., safe driving). The details of fatigue detection pipeline and the desktop and smartphone applications will be presented in Sections 6 and 7, respectively.

Users are able to view their processed bio-signal and sensor readings on the desktop and smartphone applications in semi-real-time (on the order of seconds in delay). Although there is a noticeable delay between when the glasses platform samples from its sensors to when a user can see his/her sensor and bio-signal readings, we posit that most of the applications that utilize SPIDERS+ do not require sub-second-level latency. For example, knowing when a person becomes depressed 200 ms or 2 s after the fact would likely not change the outcome or quality of care the person receives.

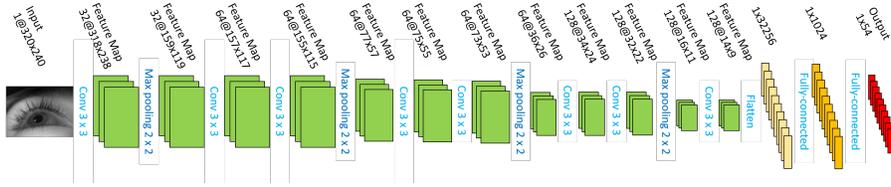


Figure 6: Breakdown of our proposed CNN for eye shape and eyebrow landmark generation.

4. Core Function Library

In this section, we discuss the algorithms we employ to extract bio-signals listed in the *Core Functionality Library* in the system architecture shown in Figure 2. SPIDERS+ is capable of measuring a wide range of bio-signals and factors correlated to a person’s emotional state. Table 2 compares the different functionalities between SPIDERS+ and existing emotion monitoring platforms. We can see from this table that SPIDERS+ is capable of much more functionalities at a fraction of the price of existing products. These signals can be freely used by developers building upon SPIDERS+ for their own applications. Additionally, SPIDERS+ leverages the measurements of these signals to detect facial expressions (apparent emotion) and real emotion.

4.1. Eye & Eyebrow Shape Detection

4.1.1. Shape Detection

The shapes of the eye and eyebrow are detected to extract features for facial expression classification. Additionally, determining the shape of the eye allows us to better detect and localize the pupil, which is an important preprocessing step in pupillometry.

To determine the shape of the wearer’s eye and eyebrow, we leverage a convolutional neural network (CNN) based approach to determine landmark points that outline the wearer’s eyelids and eyebrows from gray-scale and partial eye images. First, we construct a dataset of over 6,000 images using our IR camera taken from 10 human subjects of varying ages, genders, and ethnicities. During the experiment, subjects are instructed to express six facial expressions reflecting the six basic human emotions indicated in [37]: angry, disgusted, fearful, happy, sad, and surprised, as well as a neutral facial expression, while wearing SPIDERS+. The details of subject demographic distribution and experiment setup are illustrated in Section 5.1. We manually label each image with 27 landmark points per image, as shown in Figure 7. Generating 27 landmark points provides us with enough resolution to make out the shape of the eye and position of the brow, while not being too computationally expensive.

The implemented CNN consists of 8 convolutional layers, 4 max pooling layers, and 2 fully connected layers that output the predicted locations (x and y coordinates) of the 27 landmark points for each input image. The structure of the CNN model is shown in Figure 6, which takes a single 320x240 image

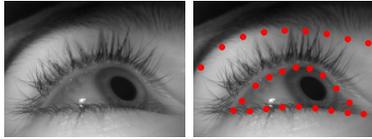


Figure 7: Left: Example raw image of the eye. Right: Eye image with 27 landmark points generated by the CNN (9 points on the upper eyelid, 9 points on the lower eyelid, and 9 points on the eyebrow).

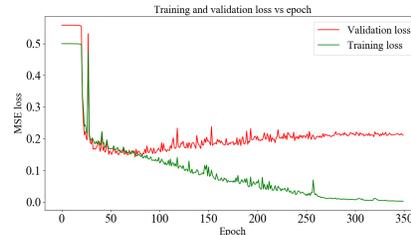


Figure 8: Training and validation loss during the training process of the landmark generation CNN, where the validation loss (red line) converges in around 300 epochs.

and outputs the coordinates the landmark points. Finally, we obtain the final contours of the eyelids and brows by fitting the generated landmark points onto a fifth-order polynomial curve.

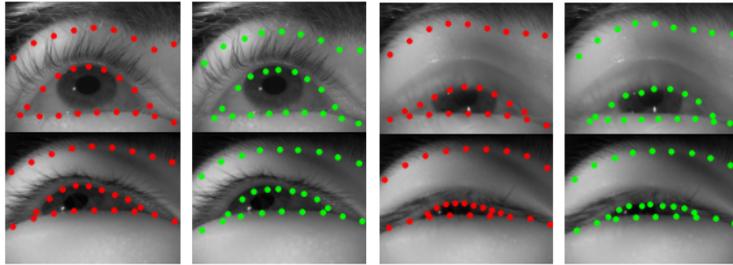
4.1.2. Training

435 We select 6,000 IR images as the dataset used for training and validation of
 the landmark generation CNN network (5,400 images as the training set and 600
 images as the validation set). To ensure that our training and validation sets are
 balanced, we randomly select images to include for both sets, but ensure that
 frames from each subject are present in both sets as well. We used the stochastic
 440 gradient descent optimizer (SGD) with 0.5 momenta and a fixed learning rate
 of 0.01. We terminated the training process when the validation loss becomes
 stable (the change in mean square error (MSE) loss is less than 0.02) over the
 last 10 epochs. As shown in Figure 8, validation and training loss decrease as
 the network is trained with more data over more epochs. This suggests that the
 445 CNN model for generating eyebrow and eye shape landmarks has become more
 generalized.

4.1.3. Evaluation

450 We test the eye and brow shape detector on approximately 200 gray-scale eye
 images outside of the training set. Since the CNN model is light, consisting of
 only 8 convolutional layers, we are able to generate the shape of the eyelids and
 brows in less than 5 ms per input image. The average root means square error
 (RMSE) of all the estimated landmarks in the test set is 4.90 pixel. Figure 9
 shows the decent performance of our CNN model, even in the case of squinted
 eyes or blurry pictures. Additionally, we tested our model on people of varying
 455 races, skin tone, and facial characteristics (e.g., different eyelash or hair lengths)
 and found similar performance across a wide range of demographics.

460 We show a set of eyelid and brow shape landmarks output by our model,
 compared to our hand-labeled ground truth, in Figure 9. Overall, our model is
 able to find the contours of eye and eyebrows with high accuracy, even in the
 case of as shown in Figure 9.



(a) Regular clear images captured by SPIDERS+. (b) Extreme cases with blurry images with squint eye.

Figure 9: Comparison between the ground-truth landmarks and the landmarks output by the CNN model (red: hand-labeled ground truth; green: output of the CNN model).

4.2. Pupillometry

Our approach for robust pupillometry using our gray-scale camera is outlined in Figure 10 and Algorithm 1. Determining the shape of the eye is a crucial first step in determining pupillometry as it helps in localizing the pupil by
 465 constraining the area to search for the pupil. First, we apply the eye and eyebrow shape detector, described in Section 4.1. The results of this step are shown in (a) and (b). The extracted IR eye-area **image** and **landmarks** are used as the input to the pupillometry module. Next, we crop out the eye area and discard the rest of the image with the knowledge of eyelid shape in (c) to
 470 get the *cropped_image*, since the pupil can only be located within the eye area.

Next, we generate a histogram of the light intensity of the pixels (d). We note that the pupil is generally a much darker color (lower intensity) than any part of the surrounding eye, and is thus located in the valley of the histogram of light intensities. Having this information, we set the threshold of light intensity to be
 475 the intensity corresponding to the first local minimum in the valley, as labeled in (d). Following the rule that any pixel in the image has lower light intensity than the threshold is set to be white (within the pupil area), otherwise is set to be black, we threshold the image to generate (e) get the *binary_image*. We remove light reflections outside and inside of the pupil area by applying standard
 480 morphological image operations and image filling algorithms. An example of the results after these operations is shown in (e). Next, we extract the raw contour (*raw_contour*) of the pupil (f). This is only a rough and noisy contour of the pupil since this contour includes part of the eyelids and effects due to light reflection.

To clean the contour in (f), we extract the curvature (g) along the raw
 485 contour starting from the lowest point on the contour and calculated pixel by pixel clockwise as demonstrated in (f). The negative peaks in (g) are the sharp turning points when we “walk” along the contour in (f). We use the negative peaks (*negative_peak_set*) as the critical points to chop the curvature into
 490 several segments. The sharp positive peak corresponds to the portion where the raw contour is distorted by reflecting light. Those smoother curvature segments either correspond to the contour of the pupil or occlusions caused by

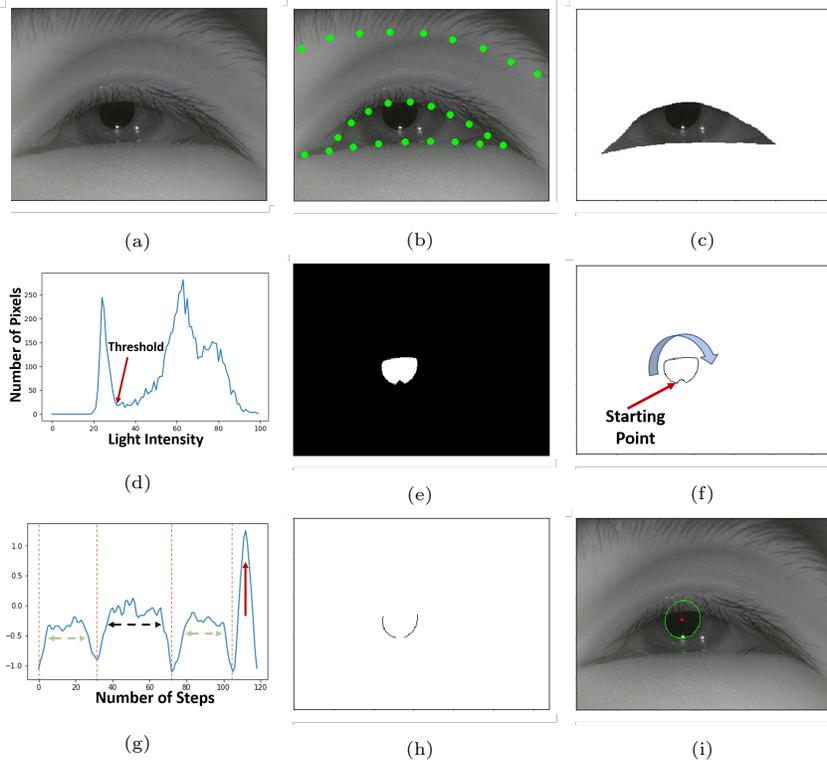


Figure 10: Pupil data processing pipeline. (a): The original image captured by SPIDERS+. (b): The 27 landmarks generated by the CNN model mentioned in Section 4.1. (c): The cropped eye area image is based on the landmarks for upper and lower eyelids. (d): The color histogram of (c), which is used to detect the threshold of the pupil light intensity. (e): Binary image based on the threshold (white: pupil). (f): The raw contour of the pupil. (g): The raw contour curvature and its segmentation according to the negative peaks. (h): The contour of the pupil. (i): The elliptic contour of the pupil.

eyelids. The latter factor can be identified by calculating the average distance
between the contour segment and the eyelids, remembering the coordinates of
the eyelids are provided by the eye and eyebrow shape detector in a previous
step. After we exclude the segments regarding the light reflections and eyelid
occlusions, the segments for the pupil contour (*detected_point_set*) can be de-
tected, as shown in (h). Finally, we fit an ellipse to the contour segments to
obtain the final denoised contour of the pupil shown in (i). In addition to the
parameters of the fitted ellipse that delineates the shape of the pupil, Algo-
rithm 1 also calculates the number of pixels of the pupil that is not covered by
the eyelid (**uncovered_pupil_pixels**) and estimates the area of the entire pupil
(**detected_pupil_pixels**). These values are used for fatigue detection, which is
described in Section 6.1.

Sometimes the contour of the extracted pupil is of low quality due to a

Algorithm 1: Pupillometry

Input:

image: IR eye image
landmarks: Output of the CNN model

Output:

detected_pupil_pixels: The number of pixels inside the portion of the pupil not covered by eyelid
uncovered_pupil_pixels: The number of pixels inside the exposed pupil
pupil_ellipse_parameter: Parameters of the whole pupil, including the center coordinates, length of the fitting ellipse’s axes and angle of rotation compared to the vertical axis

- 1 Crop the **image** based on the eye shape part of **landmarks** and get *cropped_image*.
 - 2 *binary_image* = binary of the *cropped_image* with adaptive threshold based on pixel value histogram of *cropped_image*.
 - 3 Morphology open operation for the *binary_image*.
 - 4 Morphology close operation for the *binary_image*.
 - 5 Remove the holes in *binary_image* by applying the image fill algorithm;
 - 6 **uncovered_pupil_pixels** = number of white pixels in *binary_image*;
 - 7 *raw_contour* = boundary of the white pixel block in *binary_image*;
 - 8 Calculate the *raw_contour* curvature and get the *negative_peak_set*, filter the *raw_contour* by *negative_peak_set* to get *detected_point_set*;
 - 9 Ellipse fitting *detected_point_set* to generating the **pupil_ellipse_parameter**;
 - 10 Generate the ellipse shaped *detected_boundary* by **pupil_ellipse_parameter**;
 - 11 **detected_pupil_pixels** = number of pixels bounded by *detected_boundary*;
 - 12 Output **uncovered_pupil_pixels**, **detected_pupil_pixels**, and **pupil_ellipse_parameter**;
-

number of reasons (e.g., a person blinks, closes his eyes, or squints). To filter out frames where the state of the eye prevents SPIDERS+ from estimating the shape of the eye accurately, we compare the shape of the true contour extracted in step (h) and the fitted ellipse extracted in step (i) using an error of fit function (EOF_2) [38]. If the quality is deemed too low using this test, we discard this frame, as most likely the user has closed his eyes, blinked, squinted, etc. In the end, we use a Hampel filter with a threshold of 10 times the Median Absolute Deviation (MAD) to remove the points where there is a rapid and unnatural change in pupil size. The removed values are then replaced with new values through linear interpolation. Though we provide our own implementation of pupillometry, we allow users and developers to save raw images on the client application to design and test their own algorithms. It is hard to detect the absolute size of the pupil, which is heavily dependent on the placement of the SPIDERS+.

To evaluate our pupil detection algorithm, we compare our measurements with those from a commercial product, the Tobii Pro Nano (\$10000+ with the software), which is a bar-shaped, screen-based device that remotely tracks the user’s gaze and pupillometry with a sampling rate of 60 Hz from its installation point underneath a computer monitor. No additional equipment is needed to be worn by the user (aside from SPIDERS+), which ensured that the Tobii Pro Nano do not interfere with SPIDERS+. The sampling rate for SPIDERS+ is 10 Hz.

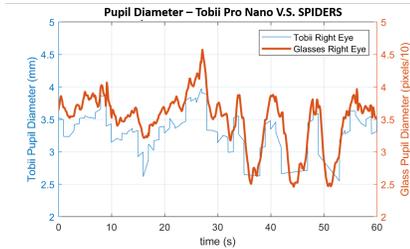


Figure 11: A 60-second segment of pupil diameters measured by both Tobii Pro Nano and SPIDERS+.

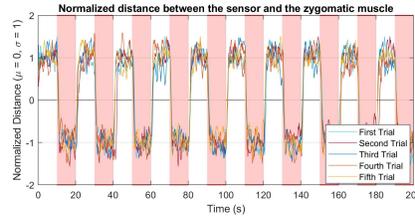


Figure 12: The relative distances between the proximity sensor and the zygomatic muscle while a subject alternates smile and calm facial expressions for 10 repetitions among 5 different trials (red shades: the ground truth when the subject is asked to smile).

We use both SPIDERS+ and Tobii Pro Nano (a table-mounted eye tracker) to take measurements on six subjects from different genders and ethnicities simultaneously, with each subject being measured for five minutes. The data collected from both Tobii and SPIDERS+ are timestamped. Unlike Tobii Pro Nano, which has multiple IR-band cameras to enable a 3-D model of the pupil and calculate the diameter of the pupil in millimeter, SPIDERS+ only contains one camera module. The pupil size output by SPIDERS+ could only be in the unit of the pixel. The correlation between the two time-series data with different units from SPIDERS+ and the Tobii Pro Nano is 0.946, demonstrating the accuracy of our pupillometry approach. Figure 11 exhibits an example of the pupil diameters provided by both platforms.

4.3. Zygomaticus Muscle Movements

As described in Section 3.2, the VL1680X proximity sensor is placed below the right lens of the glasses, facing the zygomaticus (cheek) muscle. This muscle flexes when a person smiles, resulting in a decrease in the distance between the proximity sensor and muscle [39]. To evaluate whether the proximity sensor could detect movements in the zygomaticus muscle, we ask six subjects to wear SPIDERS+ and design a stimulus to guide the subject to alternate between smiling for 10 seconds and remaining neutral for another 10 seconds for 10 repetitions. At the beginning of each experiment, the position of the proximity sensor is adjusted for each subject to face the zygomaticus muscle of each subject. To analyze the amount of calibration required for each user among different trials, each subject repeats the experiment procedure five times.

After the client application receives the sensor values, we first use a high pass filter with a cutoff frequency of $0.02Hz$ to get rid of the artifacts due to the relative distance change between the zygomaticus muscle and the sensor. A lowpass filter is utilized to remove the random noise, and the filtered data were normalized ($\mu = 0$ and $\sigma = 1$). Figure 12 shows the example readings from the proximity sensor for one subject among five different trials. The red shades indicate the periods when the subject was asked to smile. From visual inspection, there is a noticeable difference between a smiling and a neutral face

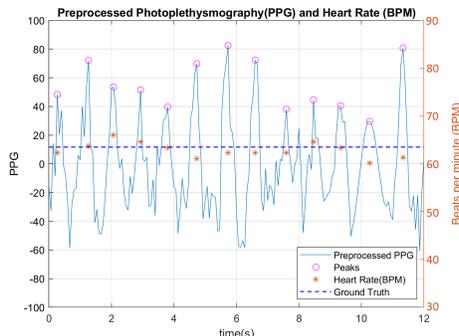


Figure 13: The filtered and smoothed PPG (blue solid line), estimated heart beat peaks (pink circles), calculated heart rate in BPM (orange stars), and the ground truth provided by a commercial heart rate monitor (blue dashed line).

with the signals being highly correlated between different trials. We conducted
 560 a two-sample t-test between the processed data of the smiling and calm trials
 from all subjects ($p < 0.01$). This shows that with simple adjustments on the
 position of the proximity sensor for each user before the first use, SPIDERS+ is
 able to robustly capture movements in the zygomaticus muscle to detect when
 a person is smiling, frowning, etc.

565 4.4. Heart Rate

SPIDERS+ measures the heart rate of a user using the MAX30102 PPG
 sensor. Since this is a contact-based sensor, we do not expect users to be
 comfortable using this sensor all day long; as such, heart rate is only an optional
 module. PPG is an inherently noisy signal. SPIDERS+ employs a few methods
 570 to clean estimate heart rate from these noisy measurements.

First, we remove any outliers that are more than three standard deviations
 away from the mean PPG reading and apply a high-pass filter with a cut-off
 frequency of 0.9 Hz to remove low frequency drift [40]. Once the PPG signal
 is filtered and smoothed, we estimate the heart rate by counting the number
 575 of peaks that occur in the raw PPG signal over a period of time. These peaks
 within the PPG signal correspond to heartbeats. We note that people generally
 have a heart rate of 120 beats per minute or below. As such, once SPIDERS+
 identifies a heartbeat (peak), it searches for the next peak beginning at least 0.5
 seconds later. This constraint helps reduce detecting false positive peaks that
 580 may yield heart rates that are too high.

To evaluate the performance of the heart rate module, we recruited 10 sub-
 jects and measured their heart rates for 60 minutes using the PPG sensor and
 heart rate module on SPIDERS+. We compared readings from SPIDERS+ with
 a Wellue FS20F Bluetooth Finger Oximeter [41]. We averaged 12-second win-
 585 dows of readings from the commercial device to use as ground truth. Over 95%
 of the BPM rates estimated from SPIDERS+ is within ± 3 BPM of the ground
 truth values collected by the commercial device. Figure 13 shows the data and

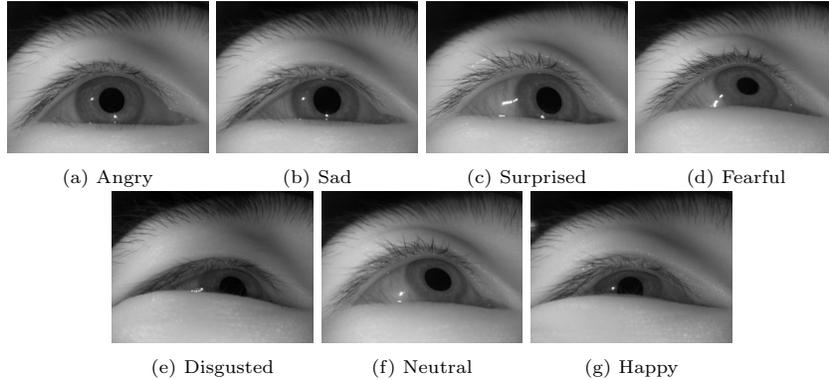


Figure 14: Eye images of seven common facial expressions. The constraints from the gray-scale image captured by a wearable eye tracker make it difficult to distinguish subtle differences among different emotions (eg. happy and disgusted).

590 measurements extracted for one 12-second trial. The blue solid line indicates the smoothed the PPG signal; the pink circles represent the detected peaks that correspond to heartbeats; the orange stars represent the calculated BPM at each prominent peak, and the blue dashed line shows the ground truth obtained from the commercial pulse oximeter. It can be seen that the heart rates estimated by SPIDERS+ match closely with those obtained from the commercial device.

5. Advanced Functionalities Prototyping

595 In this section, we prototype and evaluate our algorithms for facial expression (apparent emotion) detection and real emotion classification in the *Advanced Functionality Layer* to demonstrate the adequate performance of SPIDERS+ compared to other platforms, which are usually more expensive and bulky [13][42][43]. These make up the *Advanced Functionalities* layer of SPIDERS+ in Figure 2. All experiments presented in this section are approved by
600 the Human Research Protection Office and IRBs at our institution.

5.1. Facial Expression Detection

Previous works focus on facial expression (apparent emotion) detection using colored images of the entire face. Few works have explored facial expression
605 detection using gray-scale images and even fewer have investigated facial expression extraction using images only contain the eye area. Figure 14 shows images captured from our IR camera with a user making common facial expressions. The camera is angled very low and captures only the area around the eye and part of the eye brow. The partial view of the brow, subtle differences in the
610 shape of the eye between different expressions, and the inherent variations in the shape of the eye among different demographics all make it difficult for traditional algorithms, that match the shape of the eye to an expression, to perform well.

Table 3: Expressions and eye area features

Facial Expression	Eye State	Brow Movement
Angry/Fearful	Non-Squinting	Brow Head Moves Up
Neutral/Sad	Squinting/Normal	No Movement/Move Down
Surprised	Non-Squinting	Moves Up
Disgusted	Squinting	Brow Head Moves Down
Happy	Squinting	No Movement/Moves Up

Two features we observed that were more consistent among people of different races + gender and showed greater variation among different facial expressions are the movement of the eyebrow and the size change of the eye. Table 3 shows the relative state of the eyes and the movement of the eyebrows for the different facial expressions we studied. For instance, when a person becomes surprised, the head of the eyebrow (the portion of eyebrow closest to the nose ridge) raises and the eyes become wide open. Next, we present and evaluate our novel facial expression detection pipeline as shown in Figure 15, that leverages both the movement of the eyebrow and the shape of the eye, to classify facial expressions (apparent emotion) into one of five expressions listed in Table 3.

5.1.1. Expression Detection Approach

We propose that SPIDERS+ can perform facial expression (apparent emotion) detection with the following high-level approach:

1. Extract eye landmark features that provides the rough shape of the eyelids and eyebrows (Section 4.1).
2. Estimate the size of the eye (wide open vs. squinting).
3. Track the movement of the eyebrow (moving up vs. moving down).
4. Estimate facial expressions based on the size of the eye, movements of the eyebrow, and a binary decision tree.

Eye Size Estimation: The goal of this step of the algorithm is to determine if the eye is squinting or non-squinting. To accomplish this, we need to determine the location and the outline of the eye. The area encompassed by the outline of the eye is directly correlated to the state of the eye; if the area is large, the eyes are wide open, and if the eyes are narrow, then the person is squinting. The eye landmark generation algorithm we proposed in Section 4.1 generates discrete points that outline the shape of the eye and brow. Next, we interpolate between these points to draw the outline that encompasses the area of the eye. Once the area is computed, we manually set a threshold that achieves the greatest separation between squinting and non-squinting; eye areas above this threshold are considered non-squinting, and eyes below this threshold are considered squinting.

eyebrow Movement Tracking: Besides checking for squinting eyes, tracking the movement of eyebrows is helpful in providing information about when the subject frowns or raises the eyebrows. Figure 17 shows two consecutive frames taken when a user’s facial expression changing from angry to neutral, and the

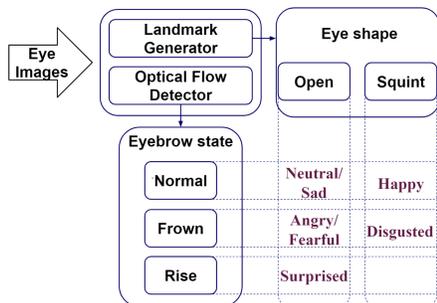


Figure 15: The pipeline of eye-image-based facial expression (apparent emotion) detection.

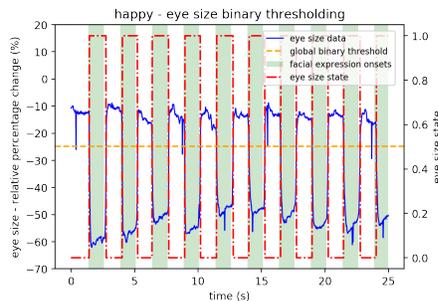


Figure 16: Example of binary thresholding conducted on baseline corrected eye size data (green shades: the ground truth when the subject is asked to smile to express happy).

650 corresponding motion vector map generated with a naive exhaustive search algorithm. From the motion map, We can easily observe that the eyebrow is moving towards the left. In order to achieve real-time performance, we adapt the PWC-Net proposed in [44] to calculate the optical flow between frames. The mean of the motion vectors close to the previously generated eyebrow position is taken as the eyebrow movement vector.

655 As the visualization in Figure 18 shows, the motion of the eyebrow between frames is manifest in the form of short, high-magnitude pulses. Here, the spikes and plunges of the horizontal and vertical motion components represent the derivative of general eyebrow states, i.e. the acceleration with which the eyebrow changes from one state to another. An equal-magnitude expansion outwards in both the x- and y-direction may represent a raise of the eyebrow (potentially induced by surprise), while a prominent contraction inwards in the x-direction, usually accompanied by a less prominent y-direction movement, may signify a frown.

660 We utilize these pulses of movements for analyzing the eyebrow state along the temporal dimension. Based on the observations made previously, three states of the eyebrow are considered in this model: neutral (state 0), raised (state 1), and frown (state -1). In order to construct the state model of the eyebrow, a few assumptions are made. First, since eyebrow movements involve the involuntary contraction and expansion of muscles, they happen over short periods of time, and thus each movement is always signaled by a short, high-power pulse. Second, we assume that each time the eyebrow state diverges from state 0, it will return to state 0 from either state 1 or state -1 before it diverges again. Third, as a result of the previous two assumptions, motion pulses are processed in pairs, with each eyebrow movement complete only after the appearance of both a positive diversion (a spike) and a negative diversion (a plunge). Hence, if two spikes or plunges appear consecutively on the time-series motion, the second signal succeeds the first in becoming the first of the following pair of movement, while the state of the eyebrow between the two spikes is still

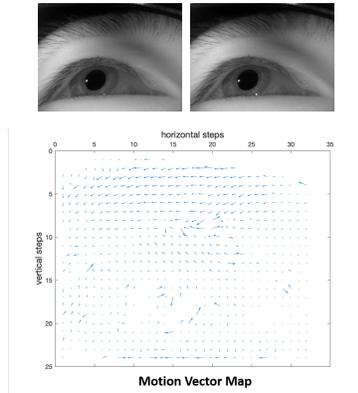


Figure 17: Motion tracking between two consecutive image frames (top). As shown in the motion vector map (bottom), the eyebrow region moves primarily leftwards from the image on the left to the image on the right.

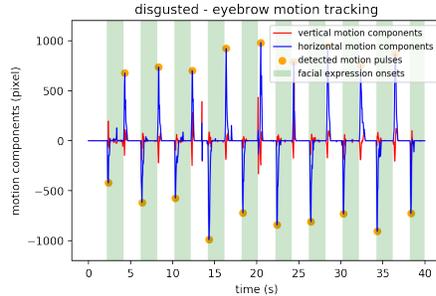


Figure 18: Tracking of eyebrow region motion in the form of overall motion vector components over time (green shades: the ground truth when the subject is asked to perform disgusted).

680 seen as the one signified by the first spike. A state diagram along the time axis is thus constructed, with 3 discrete y value levels of -1, 0, and 1, which tracks the state of the eyebrow at any time point during a trial.

5.1.2. Experiment Setup

685 Participants: 10 adults, including 5 males and 5 females, from different ethnicities (6 Asian, 3 White, and 1 African American) with self-reported normal hearing are recruited. Participants with impaired vision don't wear corrective lenses throughout the experiments since no reading is required.

690 Stimuli and Procedure: The subjects are asked to express seven different apparent facial expressions indicated in Table 3. For each apparent expression category, each subject is guided to repeat the process of explicitly performing a certain facial expression while images are taken with a 10 fps frame rate. Each subject is directed to repeat the process of explicitly performing a certain facial expression for 50 frames and 50 frames of neutral facial expression for ten times, resulting in 1000 images for one emotion.

695 We take the 10 frames in every 50 frames between the transitions as the ground truth for a certain facial expression or a neutral state. By directly calculating the eyebrow movement from an image with a neutral state to an image with a certain facial expression, we obtain the eyebrow movement of each facial expression compared to the neutral baseline. The eye sizes of the images are also calculated using the method previously described. The eyebrow movement plots and the histogram of the eye sizes are shown in Figure 19. The results in these two figures support what we describe in Table 3.

700

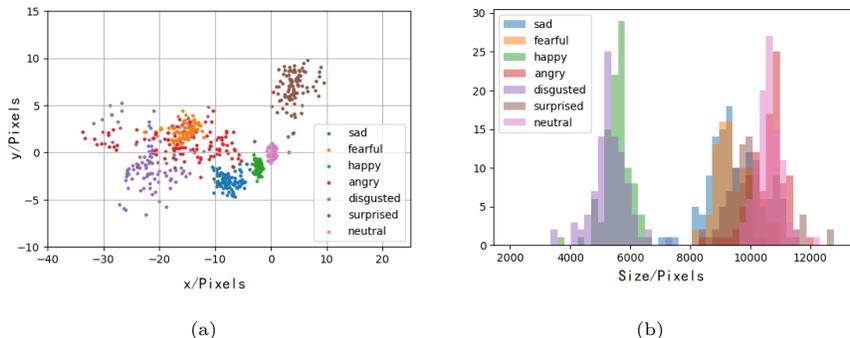


Figure 19: (a) Eye brow motion distribution and (b) eye size histogram for a same subject, we can see the feature of movements and size change described in Table 3.

5.1.3. Facial Expression Detection Results

By directly training a support vector machine (SVM) classifier on the eye size and eyebrow motion features mentioned before, we can get a classification accuracy ranging up to 94.47% for different subjects. However, the high accuracy results from over-fitting to the position of the glasses when taking certain trials of the experiment. Using a 3D pupil model or other computer vision method to compensate for glasses frame movements might solve this problem, but it is reserved for future work.

As for now, we validate the previously described more robust method, where time-series eye image data are passed into the two processing pipelines to obtain eye area data and eyebrow state data separately. With the 2×3 combinations of eye and eyebrow states, a facial expression classification is generated for each time point in the test set, with the logic shown in Figure 15. The result is a continuous prediction along the time dimension that specifies the subject's facial expression at every time point. The result thus generated for the test dataset is validated against the ground truth and an evaluation score is calculated for each trial, based on the percentage of time point-specific predictions made by the classification model. The results are shown in Figure 20.

The results show that our classification model, based solely on the detection of eye shape and the tracking of the eyebrow state with raw, gray-scale, partial images of the eye, achieves a satisfactory accuracy for facial expression classification, with the balanced average accuracy reaching up to 83.87% for the test dataset, exceeding the performance of Eyemotion, a state-of-the-art method for eye-area facial expression detection with IR cameras positioned in front of eyes (which will block the user's field of view) [6]. It is worth noting that instead of providing a probability distribution, our method provides a direct classification of each facial expression. It takes less than 100ms to process each frame, so the proposed method can achieve a frame rate of more than 10 fps.

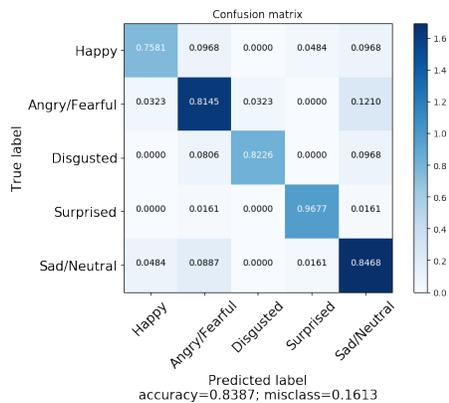


Figure 20: Confusion matrix for facial expression classification.



Figure 21: GUI of the application used to collect data for training and evaluating our real emotion classifier. Users are shown a stimulus through the GUI and are asked to select their arousal and valence level after each stimulus.

730 5.2. Real Emotion Classification

To show SPIDERS+’s effectiveness in discerning real emotional states, we implemented two 3-class Support Vector Machines (SVM), with pupil size measured from SPIDERS+ as input, to detect three classes of arousal and valence levels, which are two measures of real emotion.

735 To obtain data for training and evaluation, we recruited 12 adults with self-reported normal hearing, right-handed, and varying ages. Participants with impaired vision wore corrective lenses throughout the experiments, while everyone else had normal vision. Half of the subjects were males and the other half were females. During the experiment, all subjects were exposed to picture/sound/video test on a MATLAB graphical user interface (GUI) shown in Figure 21, while wearing SPIDERS+. Picture and sound tests were taken from the International Affective Picture System (IAPS) and International Affective Digital Sounds (IADS) datasets, while videos were selected from Youtube [45][46].

745 During each trial, the subject would first observe a grayscale picture with a black cross in the center of the screen for 6 seconds, followed by a target picture, sound, or video stimuli for 6 seconds on the MATLAB GUI. Afterwards, users would enter their arousal and valence levels using the Self-Assessment Manikin (SAM), a 9-point scale commonly used to measure arousal (excited vs. calm) and valence (pleasant vs. unpleasant) levels [47]. We consider one 12 second segment followed by a self-report of valence and arousal to be a single sample. The responses provided by the subjects would be used as labels to train and test the SVM model. Subject responses on the 9-point SAM scale were scaled to fit in a range from -1 to 1 , and we divided the range into three classes for both valence and arousal. The classes for valence are “Pleasant” (> 0.4), “Neutral” ($< 0.4, > -0.4$), and “Unpleasant” (< -0.4). The classes for arousal are “Excited” (> 0.4), “Neutral” ($< 0.4, > -0.4$), and “Calm” (< -0.4).

We trained our SVMs on 800 samples and tested on 150 samples of pupil

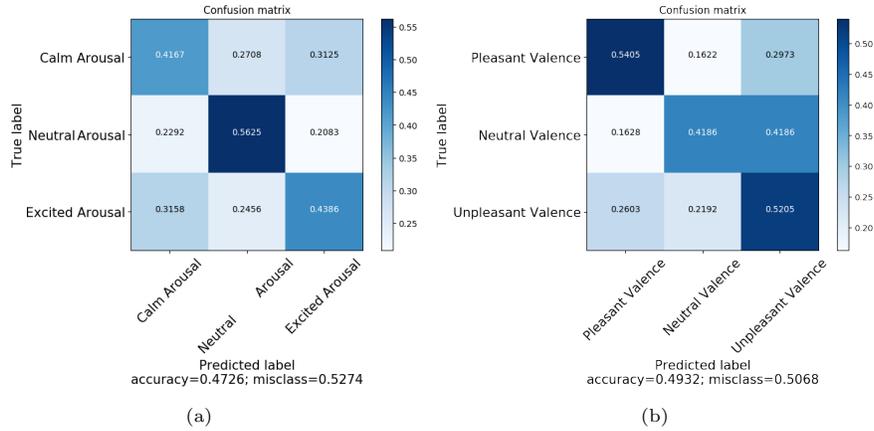


Figure 22: Confusion matrix for the (a) arousal and (b) valence-based classes.

size and valence/arousal pairs. The confusion matrices of our results for both
760 valence and arousal are shown in Figure 22. Classification for multi-class valence
and arousal levels based on physiological responses is difficult. Even commercial
products that use EEG sensing devices that cost over \$10,000 can only achieve
an accuracy of around 70% [48][49]. The average prediction accuracy across
different subjects for three-class valence and arousal in SPIDERS are 49.32%
765 and 47.25%, which is higher than the accuracy of other low-cost wearable de-
vices ([10] achieved 34.2% for detecting four-class states using physiological data
collected by contact-based sensors).

6. Enabled Applications

In this section, we discuss areas and applications where SPIDERS+ can make
770 potential impacts. In particular, we focus on an example application, fatigue
detection, and describe its detailed algorithms and implementations. We also
briefly describe other applications that can be enabled by SPIDERS+.

6.1. Fatigue Detection

People often exhibit specific facial expressions and cues when they are tired,
such as yawning or squinting their eyes. These actions generally change the
shape and position of the eyelids, eyebrows, and facial muscles. As such SPI-
DERS+ is capable of detecting these changes to infer fatigue in a user. The
processing pipeline for our fatigue detection application is shown in Figure 23.
In particular, SPIDERS+ generates two measurements using images from its IR
camera (**uncovered_pupil_pixels** and **detected_pupil_pixels**) that is used by
the smartphone and desktop applications to compute PERCLOS (Percentage
of eyelid closure over the pupil, over time), which is a widely used metric for

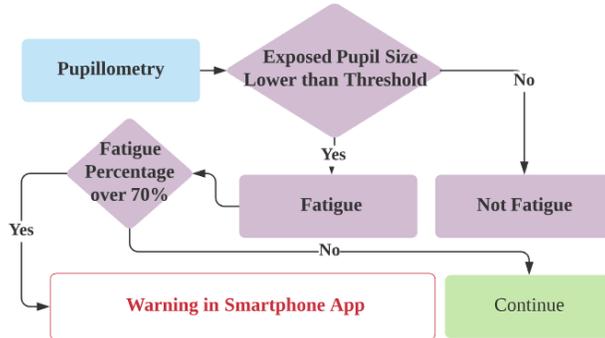


Figure 23: The processing pipeline for fatigue detection.

measuring fatigue in drivers [50]. The PERCLOS metric is computed as follows:

$$PERCLOS = \frac{\# \text{ of instant-fatigue frames}}{\# \text{ of total frames}} \times 100\%.$$

775 The smartphone and desktop applications estimate whether the user is fatigued or not using Algorithm 2. In general, people tend to squint or close their eyes when fatigued. As such, we determine *instant-fatigue frames* to be frames where the area of the pupil currently exposed (**uncovered_pupil_pixels**) is less than 20% of the total size of the pupil (**detected_pupil_pixels**). In
 780 other words, these are frames where the eyes are mostly closed. If the fraction of *instant-fatigue frames* is very high over a certain period of time, then we detect that the person is fatigued and send an alert through the smartphone/desktop applications. We evaluate the PERCLOS metric every 30 seconds (**cumulative_results**, $\mathbf{K} = 30\text{s} \times 10\text{fps} = 300$ frames in total an image sampling rate of 10 Hz). Through experimentation, we determine that setting
 785 the PERCLOS threshold for fatigue to 70% ensures that our application detects most cases of fatigue while maintaining a low false detection rate (e.g. scenarios where a user is blinking rapidly may yield many frames where the user has his/her eyes closed and cause the detector to falsely detect fatigue).
 790 In other words, if the PERCLOS metric is greater than 70%, then the smartphone/desktop application detects fatigue (**fatigue_output** = 1). The results of the fatigue detection are displayed on the user panel in both the desktop and smartphone applications described in Section 7.

6.2. Safe Driving and Concentration Monitoring

795 The number of vehicles has increased drastically in recent years. Additionally, there are approximately 100,000 police-reported crashes involving drowsy driving per year according to the National Highway Traffic Safety Administration [51]. SPIDERS+ can help reduce incidents of drowsy driving through

Algorithm 2: Fatigue Detection

Input: **uncovered_pupil_pixels:** The number of pixels inside the portion of pupil not covered by eyelid (output of Algorithm 1).
detected_pupil_pixels: The number of pixels inside the fitted ellipse (output of Algorithm 1).
cumulative_results: Vector of the instant fatigue detection results (*instant_result*) over the past **K** frames of eye images.
Output: **fatigue_output:** A Boolean value indicating fatigue (=1) or not fatigue (=0).

```
1 Initialize cumulative_results =  $\vec{0}$ , PERCLOS_threshold =  $70\% \times \mathbf{K}$  ;  
2 while True do  
3   pupil_fraction =  $\frac{\text{uncovered\_pupil\_pixels}}{\text{detected\_pupil\_pixels}}$  ;  
4   instant_result = pupil_fraction < 20% ;  
5   cumulative_results right shift by 1 ;  
6   First element of cumulative_results = instant_result ;  
7   PERCLOS_value = SUM(cumulative_results)  
8   fatigue_output = PERCLOS_value > PERCLOS_threshold ;  
9 end
```

fatigue detection, as described in Section 6.1. If a user is detected to be fatigued,
800 the smartphone application will send a five-second audio alert to wake up the
driver. SPIDERS+ is also comfortable to wear, wireless, and does not block the
driver’s field of view.

In addition, the fatigue detection system we outlined in Section 6.1 can
also be used in other application scenarios, where monitoring user attention is
805 important. For example, in various scientific studies, different levels of human
subject attentiveness can cause significant differences in observed results (e.g.,
in neuroscience experiments, visual attention will affect neural processing in
different ways [52]).

6.3. Mental Health Monitoring

810 According to the World Health Organization, more than 300 million peo-
ple suffer from depression and other emotional health disorders worldwide [53].
Proper treatment may require continuous monitoring of the patients, which cur-
rently happens only when patients come in for appointments [54][55]. Intelligent
wearable devices that can monitor aspects of emotion and mental health, like
815 SPIDERS+, allow doctors to continuously monitor their patients’ even outside
doctor visits. Additionally, many psychiatric and psychological diagnoses cur-
rently rely on subjective observations. SPIDERS+ provides measurements of
critical bio-signals that are correlated with emotion that can give healthcare
providers a more objective view of their patients’ mental and emotional health.
820 In many scenarios, patients are unable to make frequent trips to their health-
care provider (e.g., elderly patients with trouble moving, or patients living in
rural/isolated areas) [3]. It is especially difficult, during the current COVID-19
global pandemic, to arrange mask-off doctor visits and mental status exams.

SPIDERS+ is an intelligent, low-power, wearable device that is not constrained to artificial lab settings or clinical environments and can provide almost continuous emotional health monitoring of patients without the need to make frequent trips to the doctor. SPIDERS+ is able to collect and process bio-signal measurements that can be used to more accurately discern emotion and mental state in daily life. Patients would be able to live out their lives normally while wearing SPIDERS+ because the SPIDERS+ glasses platform is truly wireless and portable. In addition to medical use, SPIDERS+ is suitable for general users to conduct everyday mental health monitoring. Users can easily view the distribution of the five classes of their facial expressions once they start wearing SPIDERS+ on the service panel either running on the desktop or smartphone application. The details of the desktop and smartphone applications will be discussed in Section 7.

6.4. Entertainment and Gaming

There are many ways in which SPIDERS+ could enrich different categories of entertainment. For instance, a film-maker could use the spectators' emotional feedback to decide how to make the movie more sympathetic and impressive [56]. When facial expression recognition is integrated into an intelligent game system, the gaming experience can become more interactive, vivid, and attractive. A video game could dynamically increase the difficulty of the game if it detects that the player is getting bored or complacent [6]. A musical therapy system can utilize the real-time emotional state of the user and adjust the music therapy to the patient's needs [57]. Leveraging the optional contact-based sensors in SPIDERS+ can also enable applications like providing real-time feedback on medication and relaxation through measuring brain activity and HRV [58]. In general, media can dynamically change (and enrich) content based on the current user state of apparent and real emotions.

6.5. Marketing and Advertisements

Companies can use emotion predictions provided by SPIDERS+ to enhance their marketing campaigns and better reach their target audience. For example, companies can use the emotional response of large groups of people to determine which advertisements generate more interest in their products to help them create better advertisements. Companies can also use emotional responses to their advertisements to determine their target market and send targeted advertisements only to the people that are most likely to buy their products [59]. Moreover, with customers' emotional responses, the government can enhance the impact of public service announcements, such as advertisements for safe driving and water saving [5]. Customers would greatly benefit from this because they would only receive advertisements about products and services they are more willing to explore, rather than receiving advertisements about products that are not of their interests.

865 6.6. Elderly Care

In the United States, there are more than 40 million people who are 65 and older, accounting for around 15% of the total population. Around 20% of the elderly population have experienced some type of mental health issue. These mental health issues may be from a variety of reasons, including changes
870 in the environment, dementia-causing illnesses, chronic illnesses, and physical illnesses [60]. It is difficult to continuously monitor the mental health conditions of elderly citizens even after moving them into an assisted living community due to personnel and resource limitations. SPIDERS+ can continuously monitor the mental health of seniors, reducing the burden of caregivers and family members.
875 SPIDERS+ can also monitor aspects of physical health by measuring HRV through its PPG sensor module. The information provided by SPIDERS+ can provide early warnings for mental and physical health issues in senior citizens.

7. Desktop and Smartphone Applications

We developed a basic set of user-facing desktop and smartphone applica-
880 tions to enable the user to better visualize information from SPIDERS+. The smartphone/desktop application can be modified by researchers and developers to build custom applications on top of SPIDERS+, such as the fatigue detection application introduced in Section 6.1. The smartphone/desktop application utilizes outputs from the bottom three layers of SPIDERS+'s architecture (Fig-
885 ure 2) to enable a wide range of applications.

Both the smartphone and desktop applications have a user panel that displays the key features extracted from the *Core Function Library* and *Advanced Functionalities*. Specifically, the user panel displays the eye and eyebrow shape detection results (the 27 landmark points) and pupillometry results (fitted ellipse) as an overlay on top of the IR image of the user's eye. The user panel
890 also displays 1) pupil size, 2) eye size, and 3) eyebrow movements that were measured by the *Core Function Library*. Additionally, current estimates of the user's facial expression and a cumulative average of the user's facial expression over time from the *Advanced Functionalities* layer can also be viewed. Figure 24
895 illustrates the desktop application displaying the user panel. This example application was also adapted to estimate and display fatigue (Section 6.1).

The user panel is streaming to a web-page for the desktop application, while it is streaming to an embedded video player in the smartphone application. As shown in Figure 25, the homepage of the smartphone application allows the
900 user to input the URL of the SPIDERS+'s client application to obtain the video stream. The user can also choose whether to turn on the alarm or not prior to use. The user can enable the fatigue alarm function by simply clicking on the radio button next to *alarm* (outlined by the red box in the top picture in Figure 25) before clicking on the *start* button to visualize the user panel.
905 The application page contains the user panel and notification messages with the alarm sound, where the latter one will only appear if the users selected the *alarm* function on the homepage.



Figure 24: The interface of the user panel on the web-page-based desktop application.



Figure 25: Screenshots of SPIDERS+'s phone application. Top: The homepage of the smart-phone application. Bottom: The user panel with the same layout as the desktop application. The alarm notification pops up along with a warning sound, if the user turns on the alarm on the homepage.

8. Limitations and Future Work

The current SPIDERS+ system has several limitations. First, the primary
 910 contribution of SPIDERS+ is addressing the challenges of power consumption,
 cost, and portability, and creating a low-power wearable platform that enables
 researchers and engineers to develop real-time applications that benefit or re-
 quire continuous in-situ bio-signals monitoring. As such, in this paper, we pro-
 915 totype facial expression detection and emotion classification models and show
 that these models can be run on SPIDERS+ with adequate performance com-
 pared to the works done by other groups in the research community [6, 10].
 In the next phase, we will explore better feature engineering, modality fusion,
 and algorithmic development strategies that could boost the performance of the
 classification models drafted in this paper.

To validate the performance of SPIDERS+, we conduct the experiments in
 920 controlled situations similar to the experimental setups in [61, 62]. We will per-
 form further research to develop algorithms that are more robust to different
 lighting conditions (indoor and outdoor), different users, and body movements
 based on the five functions in SPIDERS+'s *Core Function Library*. We will also
 925 perform in-the-wild experiments with multiple days of continuous wearing. We
 will continue adding more functionality and exploring more applications that can
 be built upon SPIDERS+. Specific improvements include more efficient data
 processing algorithms and a communication platform that connects doctors and
 patients. We are also working on improving the robustness and security of SPI-

930 DERS+ for future medical deployments and studies with real patients. These improvements to SPIDERS+ will further reduce societal barriers to obtaining proper and accessible mental medical services.

9. Conclusion

In this paper, we present SPIDERS+, a low-cost wearable glasses platform that provides four sensing modalities (eye and brow shape detection, pupillometry, zygomaticus movements, head movements) from three non-contact sensors (IR camera, proximity sensor, IMU), with the option of including sensors that require skin contact (heart rate, skin conductance, EEG). We show that SPIDERS+ is able to robustly classify five different classes of facial expressions (apparent emotions) and three levels of valence and arousal (real emotion). We also develop a basic desktop/smartphone application for visualizing SPIDERS+'s data to enable various applications. SPIDERS+'s low cost, portability, and long battery life, combined with its capability to estimate both facets of human emotion (apparent and real) make it the perfect platform for exploring applications that can utilize human emotions.

Additionally, we present and discuss many high-level applications that SPIDERS+ will impact and plan to explore some of these problems in future work. In addition to the applications discussed, there are many other concrete problems such as detecting lies, detecting drowsy driving, or determining engagement levels in a class that SPIDERS+ has the chance to impact. As such, we have open-sourced all of our code and designs to the public and will continue to promote SPIDERS+ to foster a community of developers and researchers who will collectively improve and develop new applications for SPIDERS+. We envision SPIDERS+ as the go-to open-source platform for investigating applications in mobile emotion monitoring.

References

- [1] J. Nie, Y. Hu, Y. Wang, S. Xia, X. Jiang, SPIDERS: Low-cost wireless glasses for continuous in-situ bio-signal acquisition and emotion recognition, in: 2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI), IEEE, 2020, pp. 27–39.
- [2] Y. Hu, J. Nie, Y. Wang, S. Xia, X. Jiang, Demo abstract: Wireless glasses for non-contact facial expression monitoring, in: 2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), IEEE, 2020, pp. 367–368.
- [3] R. LiKamWa, Y. Liu, N. D. Lane, L. Zhong, Moodscope: Building a mood sensor from smartphone usage patterns, in: Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys '13, Association for Computing Machinery, New York, NY, USA, 2013, p. 389–402. doi:10.1145/2462456.2464449.
URL <https://doi.org/10.1145/2462456.2464449>

- [4] K. Rowan, D. D. McAlpine, L. A. Blewett, Access and cost barriers to mental health care, by insurance status, 1999–2010, *Health affairs* 32 (10) (2013) 1723–1730.
- 975 [5] N. Hamelin, O. El Moujahid, P. Thaichon, Emotion and advertising effectiveness: A novel facial expression analysis approach, *Journal of Retailing and Consumer Services* 36 (2017) 103–111.
- [6] S. Hickson, N. Dufour, A. Sud, V. Kwatra, I. A. Essa, Eyemotion: Classifying facial expressions in VR using eye-tracking cameras, *CoRR* abs/1707.07204. [arXiv:1707.07204](https://arxiv.org/abs/1707.07204).
980 URL <http://arxiv.org/abs/1707.07204>
- [7] R. R. Henderson, M. M. Bradley, P. J. Lang, Emotional imagery and pupil diameter, *Psychophysiology* 55 (5). doi:10.1111/psyp.13050.
URL <https://doi.org/10.1111/psyp.13050>
- [8] T. Partala, V. Surakka, Pupil size variation as an indication of affective processing, *Int. J. Hum.-Comput. Stud.* 59 (2003) 185–198.
985
- [9] M. M. Bradley, L. Miccoli, M. A. Escrig, P. J. Lang, The pupil as a measure of emotional arousal and autonomic activation, *Psychophysiology* 45 (4) (2008) 602–607.
- [10] J. Kwon, L. Kim, Emotion recognition using a glasses-type wearable device via multi-channel facial responses, *CoRR* abs/1905.05360. [arXiv:1905.05360](https://arxiv.org/abs/1905.05360).
990
- [11] Our technology.
URL <https://www.fitbit.com/technology>
- [12] How it works.
995 URL <https://choosemuse.com/how-it-works/>
- [13] T. Technology, Tobii pro glasses 2.
URL <https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/>
- [14] A. Mayberry, Y. Tun, P. Hu, D. Smith-Freedman, D. Ganesan, B. M. Marlin, C. Salthouse, Cider: Enabling robustness-power tradeoffs on a computational eyeglass, in: *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, MobiCom '15*, ACM, New York, NY, USA, 2015, pp. 400–412. doi:10.1145/2789168.2790096.
1000 URL <http://doi.acm.org/10.1145/2789168.2790096>
- 1005 [15] T. Li, X. Zhou, Battery-free eye tracker on glasses, in: *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking, MobiCom '18*, ACM, New York, NY, USA, 2018, pp. 67–82. doi:10.1145/3241539.3241578.
URL <http://doi.acm.org/10.1145/3241539.3241578>

- 1010 [16] B. Li, H. Fu, D. Wen, W. LO, Etracker: A mobile gaze-tracking system with near-eye display based on a combined gaze-tracking algorithm, *Sensors* 18 (5). doi:10.3390/s18051626.
- [17] K. Masai, Y. Sugiura, K. Suzuki, S. Shimamura, K. Kunze, M. Ogata, M. Inami, M. Sugimoto, Affectivewear: Towards recognizing affect in real life, in: *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, ACM, New York, NY, USA, 2015, pp. 357–360. doi:10.1145/2800835.2800898.
- 1015
- [18] K. Masai, K. Kunze, M. sugimoto, M. Billinghamurst, Empathy glasses, in: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ACM, New York, NY, USA, 2016, pp. 1257–1263. doi:10.1145/2851581.2892370.
- 1020
- [19] S. Xia, D. de Godoy Peixoto, B. Islam, M. T. Islam, S. Nirjon, P. R. Kinget, X. Jiang, Improving pedestrian safety in cities using intelligent wearable systems, *IEEE Internet of Things Journal* 6 (5) (2019) 7497–7514.
- 1025
- [20] J. Jia, C. Xu, S. Pan, S. Xia, P. Wei, H. Y. Noh, P. Zhang, X. Jiang, Conductive thread-based textile sensor for continuous perspiration level monitoring, *Sensors* 18 (11).
- [21] D. de Godoy, B. Islam, S. Xia, M. T. Islam, R. Chandrasekaran, Y.-C. Chen, S. Nirjon, P. R. Kinget, X. Jiang, Paws: A wearable acoustic system for pedestrian safety, in: *2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI)*, IEEE, 2018, pp. 237–248.
- 1030
- [22] D. Hong, B. Zhang, Q. Li, S. Nirjon, R. Dickerson, G. Shen, X. Jiang, J. A. Stankovic, Demo abstract: Septimu—continuous in-situ human wellness monitoring and feedback using sensors embedded in earphones, in: *2012 ACM/IEEE 11th International Conference on Information Processing in Sensor Networks (IPSN)*, IEEE, 2012, pp. 159–160.
- 1035
- [23] V. Vezhnevets, A. Degtiareva, Robust and accurate eye contour extraction, in: *Proc. Graphicon*, 2003, pp. 81–84.
- 1040
- [24] Y. Ji, S. Wang, Y. Lu, J. Wei, Y. Zhao, Eye and mouth state detection algorithm based on contour feature extraction, *Journal of Electronic Imaging* 27 (5) (2018) 051205.
- [25] A. Soetedjo, Eye detection based-on color and shape features, *Editorial Preface* 3 (5).
- 1045
- [26] H. Wu, J. Feng, X. Tian, E. Sun, Y. Liu, B. Dong, F. Xu, S. Zhong, EMO: real-time emotion recognition from single-eye images for resource-constrained eyewear devices, in: *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, 2020, pp. 448–461.

- 1050 [27] K. Sun, Y. Zhao, B. Jiang, T. Cheng, B. Xiao, D. Liu, Y. Mu, X. Wang,
W. Liu, J. Wang, High-resolution representations for labeling pixels and
regions, arXiv preprint arXiv:1904.04514.
- [28] Y. Sun, X. Wang, X. Tang, Deep convolutional network cascade for facial
point detection, in: Proceedings of the IEEE conference on computer vision
1055 and pattern recognition, 2013, pp. 3476–3483.
- [29] M. Kopaczka, J. Schock, D. Merhof, Super-realtime facial landmark detec-
tion and shape fitting by deep regression of shape model parameters, arXiv
preprint arXiv:1902.03459.
- [30] M. J. R. Oliva, A. Anikin, Pupil dilation reflects the time course of emotion
1060 recognition in human vocalizations, in: Scientific Reports, 2018.
- [31] V. Raudonis, G. Dervinis, A. Vilkauskas, A. Paulauskaite-Taraseviciene,
G. Kersulyte-Raudone, Evaluation of human emotion from eye motions,
Evaluation 4 (8).
- [32] J. C. A. Costa, F. Guimbretière, M. F. Jung, T. Choudhury, Boostmeup:
1065 Improving cognitive performance in the moment by unobtrusively regulat-
ing emotions with a smartwatch, IMWUT 3 (2019) 40:1–40:23.
- [33] B. M. Appelhans, L. J. Luecken, Heart rate variability as an index of reg-
ulated emotional responding, Review of general psychology 10 (3) (2006)
229–240.
- 1070 [34] M. Li, B.-L. Lu, Emotion classification based on gamma-band eeg, in: 2009
Annual International Conference of the IEEE Engineering in medicine and
biology society, 2009, pp. 1223–1226.
- [35] M. Li, H. Xu, X. Liu, S. Lu, Emotion recognition from multichannel eeg
signals using k-nearest neighbor classification, Technology and Health Care
1075 26 (S1) (2018) 509–519.
URL <https://www.ncbi.nlm.nih.gov/pubmed/29758974>
- [36] H. Dabas, C. Sethi, C. Dua, M. Dalawat, D. Sethia, Emotion classification
using eeg signals, in: Proceedings of the 2018 2Nd International Conference
on Computer Science and Artificial Intelligence, CSAI '18, ACM, New York,
1080 NY, USA, 2018, pp. 380–384. doi:10.1145/3297156.3297177.
URL <http://doi.acm.org/10.1145/3297156.3297177>
- [37] P. Ekman, An argument for basic emotions, Cognition & emotion 6 (3-4)
(1992) 169–200.
- [38] L. Świrski, A. Bulling, N. Dodgson, Robust real-time pupil tracking in
1085 highly off-axis images, in: Proceedings of the Symposium on Eye Tracking
Research and Applications, ETRA '12, ACM, New York, NY, USA, 2012,
pp. 173–176. doi:10.1145/2168556.2168585.
URL <http://doi.acm.org/10.1145/2168556.2168585>

- [39] B. Farnsworth, Zygomaticus major – relating muscle movement to emotion (2019).
1090 URL <https://imotions.com/blog/zygomaticus-major/>
- [40] Z. Tong, X. Chen, Z. He, K. Tong, Z. Fang, X. Wang, Emotion recognition based on photoplethysmogram and electroencephalogram, in: 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMP-
1095 SAC), Vol. 02, 2018, pp. 402–407. doi:10.1109/COMPSAC.2018.10266.
- [41] Fs20f bluetooth finger oximeter.
URL <https://getwellue.com/pages/fs20f-bluetooth-finger-oximeter>
- [42] The imotions platform.
URL <https://imotions.com/platform/>
- 1100 [43] M. Kassner, W. Patera, A. Bulling, Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction, in: Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication, 2014, pp. 1151–1160.
- [44] D. Sun, X. Yang, M.-Y. Liu, J. Kautz, Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume, in: The IEEE Conference on
1105 Computer Vision and Pattern Recognition (CVPR), 2018.
- [45] P. J. Lang, M. M. Bradley, B. N. Cuthbert, International affective picture system (iaps): Technical manual and affective ratings, NIMH Center for the Study of Emotion and Attention 1 (1997) 39–58.
- 1110 [46] M. M. Bradley, P. J. Lang, The international affective digitized sounds (; iads-2): Affective ratings of sounds and instruction manual, University of Florida, Gainesville, FL, Tech. Rep. B-3.
- [47] M. M. Bradley, P. J. Lang, Measuring emotion: the self-assessment manikin and the semantic differential, *Journal of behavior therapy and experimental psychiatry* 25 (1) (1994) 49–59.
1115
- [48] M. Soleymani, M. Pantic, T. Pun, Multimodal emotion recognition in response to videos, *IEEE transactions on affective computing* 3 (2) (2011) 211–223.
- [49] Y. Lu, W.-L. Zheng, B. Li, B.-L. Lu, Combining eye movements and eeg to enhance emotion recognition, in: Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015.
1120
- [50] W. W. Wierwille, S. Wreggit, C. Kirn, L. Ellsworth, R. Fairbanks, Research on vehicle-based driver status/performance monitoring; development, validation, and refinement of algorithms for detection of driver drowsiness. final report, Tech. rep. (1994).
1125

- [51] T. N. S. Council, Drivers are falling asleep behind the wheel (2020).
URL <https://www.nsc.org/road-safety/safety-topics/fatigued-driving>
- 1130 [52] S. K. Ungerleider, L. G. Mechanisms of visual attention in the human cortex, *Annual review of neuroscience* 23 (1) (2000) 315–341.
- [53] W. H. Organization, Depression (March 2018).
URL <https://www.who.int/news-room/fact-sheets/detail/depression>
- 1135 [54] M. Vidal, J. Turner, A. Bulling, H. Gellersen, Wearable eye tracking for mental health monitoring, *Computer Communications* 35 (11) (2012) 1306–1311.
- [55] J. A. Naslund, K. A. Aschbrenner, S. J. Bartels, Wearable devices and smartphones for activity tracking among people with serious mental illness, *Mental health and physical activity* 10 (2016) 10–17.
- 1140 [56] K. Kimura, S. Haramizu, K. Sanada, A. Oshida, Emotional state of being moved elicited by films: A comparison with several positive emotions, *Frontiers in Psychology* 10 (2019) 1935. doi:10.3389/fpsyg.2019.01935.
URL <https://www.frontiersin.org/article/10.3389/fpsyg.2019.01935>
- 1145 [57] O. Sourina, Y. Liu, M. K. Nguyen, Real-time eeg-based emotion recognition for music therapy, *Journal on Multimodal User Interfaces* 5 (1) (2012) 27–35. doi:10.1007/s12193-011-0080-6.
URL <https://doi.org/10.1007/s12193-011-0080-6>
- 1150 [58] J. Lagopoulos, J. Xu, I. Rasmussen, A. Vik, G. S. Malhi, C. F. Eliassen, I. E. Arntsen, J. G. Sæther, S. Hollup, A. Holen, et al., Increased theta and alpha eeg activity during nondirective meditation, *The Journal of Alternative and Complementary Medicine* 15 (11) (2009) 1187–1192.
- 1155 [59] N. Hristova, G. M. P. O’Hare, Ad-me: wireless advertising adapted to the user location, device and emotions, in: *37th Annual Hawaii International Conference on System Sciences*, 2004. Proceedings of the, 2004, pp. 10 pp.–. doi:10.1109/HICSS.2004.1265673.
- [60] S. Stevenson, 10 symptoms of mental illness in the elderly (2018).
URL <https://www.aplaceformom.com/blog/2013-10-7-mental-illness-in-the-elderly/>
- 1160 [61] W.-L. Zheng, B.-N. Dong, B.-L. Lu, Multimodal emotion recognition using eeg and eye tracking data, in: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE*, 2014, pp. 5040–5043.

- ¹¹⁶⁵ [62] E. Gatti, E. Calzolari, E. Maggioni, M. Obrist, Emotional ratings and skin conductance response to visual, auditory and haptic stimuli, *Scientific data* 5 (2018) 180120.