SPIDERS: Low-Cost Wireless Glasses for Continuous In-Situ Bio-Signal Acquisition and Emotion Recognition

Jingping Nie*, Yigong Hu*, Yuanyuting Wang, Stephen Xia, and Xiaofan Jiang
Department of Electrical Engineering, Columbia University
{jn2551, yh3104, yw3241, sx2194}@columbia.edu, jiang@ee.columbia.edu

***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 (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, 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 up to 83.87% accuracy, as well as a pupillometry-based real emotion classifier with higher accuracy than other low-cost wearable platforms that use sensors requiring skin contact. SPIDERS costs less than $\$20$ to assemble and can continuously run for up to 9 hours before recharging. We demonstrate that SPIDERS is a truly wireless and portable platform that has the capability to impact a wide range of applications, where knowledge of the user's emotional state is critical.**

***Index Terms*—Wearables, emotion monitoring, pupillometry, facial expression detection, bio-signal acquisition**

I. Introduction

Emotion monitoring systems play an important role in improving the mental health conditions of the general public. They help psychotherapists diagnose and develop better treatments for patients, as well as inform the general population of their own mental states. Emotion monitoring platforms also have immense impacts for 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 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 [1]–[4].

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

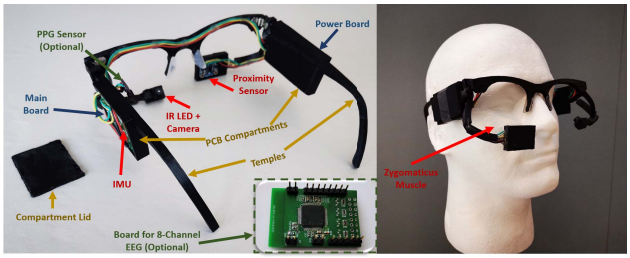


Fig. 1: Left: SPIDERS glasses platform (red: hardware sensors; blue: PCB boards; yellow: frame components; green-optional attachments). Right: SPIDERS glasses on a mannequin.

Pupillometry has been shown to be a strong indicator of real emotions [5]–[7]. However, even using the state-of-art algorithms and systems that utilize pupillometry, such as the *Tobii Pro Glasses 2*, cannot reliably predict real emotions due to inherent noise in the measurements caused by lighting changes, reflections, rapid eye movements, blinking, etc. As such 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.

Our contributions are summarized below:

- We present SPIDERS, a low-cost, energy-efficient, and portable wearable platform, shown in Figure 1 that

*These authors contributed equally to this work.

978-1-7281-6602-5/20/$31.00 ©2020 IEEE
DOI 10.1109/IoTDI49375.2020.00011

collects data from a variety of non-contact sensors to provide: 1) eye & eyebrow detection, 2) pupillometry, 3) zygomaticus muscle (smile muscle) movements, and 4) head movements. SPIDERS can be also extended to include functionalities provided by contact-based sensors, such as 5) heart rate and 6) 8-channel EEG. We demonstrate the robustness of our sensing modalities by comparing against commercial products.
- We present novel and robust vision-based techniques to perform pupillometry 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 field of view of the user.
- We develop a data processing pipeline that intelligently 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 ensure that our emotion recognition algorithms run in semi-real-time. We demonstrate that our wearable is able to last up to 9 hours.
- We propose a novel optical-flow based algorithm for estimating facial expressions (apparent emotions) by leveraging movements in the user's eyes as s/he transitions into a new expression, reaching an average classification rate of 83.87%, outperforming the state-of-the-art approach [4]. We also demonstrate SPIDERS's capability of estimating a user's affective state (real emotions), reaching a classification rate of 49.32%.
- We provide the community with: 1) anonymized dataset of biometrics measured from subjects during our controlled experiment setting 2) 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 emotion-based applications based on SPIDERS.

The rest of this paper is organized as follows. Section II presents related work and discusses limitations in current works and products that SPIDERS addresses. Section III details the sensing modalities included in SPIDERS and the hardware platform. Section IV describes the library of core functions, including the eye and eyebrow shape detection, pupillometry, and zygomatics movements, that SPIDERS enables. Section V prototypes and evaluates the facial expressions (apparent emotion) detection and real emotion classification. Section VI introduces the potential applications SPIDERS could enable. Section VII discusses limitations and future work, and Section VIII concludes the paper.

II. 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 then analyze different modalities of data that would be useful for emotion classification purposes, and explore algorithms which derive the types of data that cannot be directly measured by minimal sensors.

A. 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 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 [8]. *Muse* is a headband that provides five-channel electroencephalogram (EEG) data, heart-rate, and gyroscope sensors [9]. 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 example, the *Tobii Pro Glasses 2* leverages a wide-angle camera, microphone, gyroscope, and accelerometer to track the wearer's eye [10]. However, the 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 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 [11]–[13], but not apparent emotion. A second category of works leverage photoreflective IR sensors integrated within the eyeglasses frames to detect facial expressions [14] [15]. 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. [16] leverages sensor fusion, combining facial features extracted from a camera 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 skin and sensor, which is not stable if the wearer is on-the-move.

B. 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 and can lead to privacy concerns. On the other hand, wearable devices are becoming more and more popular, and provide in-situ sensing capabilities [17]–[20]. A camera mounted on the glasses can not only sense the pupil size and gaze angle, but also can see clearly the eyelids and the eyebrows. Previous work directly use raw images as an input to a machine learning classifier [16]. However, directly learning the features with raw images requires large amount of labeled data and can perform well within the dataset and perform poorly in real-world cases.

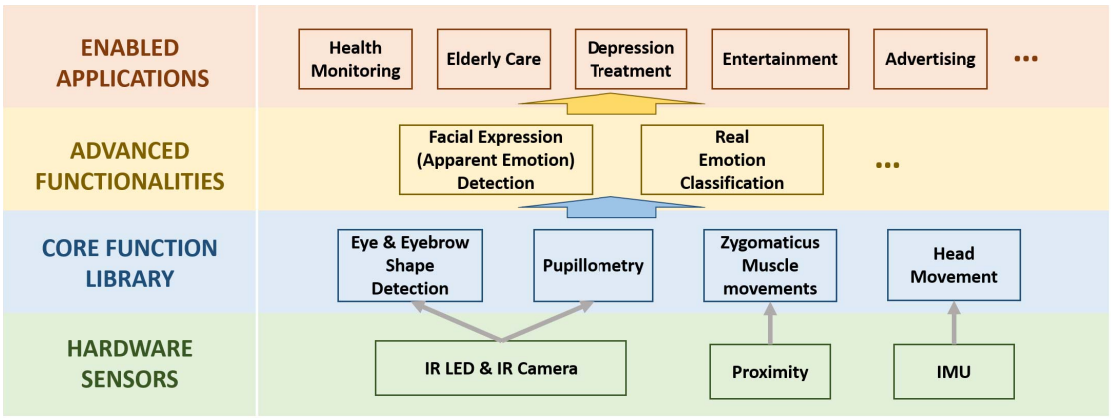


Fig. 2: Four-layer system architecture block diagram.

For eyelid and eyebrow detection, many previous works utilize RGB images, conducting adaptive thresholding, luminance valley detection, and edge detection to detect the shape of the eye lids and eyebrows [21]–[23]. However, these methods are computationally expensive and not robust. Also, in order to perform pupillometry, we have to use IR cameras which produce gray-scale images, making the boundary between the sclera (white outer layer of the eye) and skin harder to distinguish.

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 [24] [25], and some of these methods have been shown to run in real-time [26]. 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.

C. 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 [6], [27], [28].

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 [29] [30]. EEG head caps are also widely used in detecting emotion because it non-intrusively captures brain signals [31]–[33]. 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.

III. 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. Each layer in the architecture directly enables the applications or functionalities in the layer directly above. We take a top-down approach in designing SPIDERS based on the types of applications we aim to address. 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.

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* listed in Figure 2.

A. Sensing Modalities Overview

In this section, we discuss the types of bio-signals that we decided to incorporate into 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 II. 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. A person may smile while s/he is happy or sad.

Common signals used to determine someone's real emotions include pupillometry, skin conductance, EEG, and HRV. We omit facial expressions, as people'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 will 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 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 IV-B.

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 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 IV-A and Section V.

A summary of the different bio-signals that SPIDERS provides is shown in the *Core Function Library* layer of 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.

B. 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 open-source all of our glasses designs, code, and assembly instructions to the community to allow anyone to build upon or use SPIDERS for their own projects [1].

1) Processing and Wireless Transmission 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. The Espressif ESP32 supports both WiFi and Bluetooth, computational power at low power consumption, and direct memory access (DMA) support.

[1]https://github.com/Columbia-ICSL/SPIDERS

TABLE I: Price Breakdown.

Module	Unit Price (Retail) [U$]	Unit Price (Wholesale) [U$]
MCU (ESP-32)	4.80	4.80
Proximity Sensor (VL6180X)	4.50	2.18
IMU (MPU-6050)	7.28	3.52
CAMERA (OV2640)	8.50	8.50
Voltage Regulators (TPS63031DSKR)	1.87	0.79
Glasses (PLA Filament)	0.01	0.01
total	26.96	19.80

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 on-chip JPEG encoding, converting pixels into a machine-readable format without computation required from the MCU. We also install an $850nm$ 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.

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 the final SPIDERS glasses system.

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

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 main board and power board.

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 on have a certain degree of freedom to rotate, which enables customized sensor placement.

All boards and sensors are connected with wires running along the frame of the glasses. The total weight of SPIDERS,

TABLE II: Power consumption.

Module	Active Power	Sleep Power
ESP32 MCU	128.00 mW	2.64 mW
ESP32 WiFi Connection	29.60 mW	0
IMU (MPU6050)	16.30 mW	16.50 μW
Proximity Sensor (VL6180X)	82.10 mW	3.30 μW
IR LED	49.50 mW	0
Camera (OV2640)	168.00 mW	1.98 mW
ESP32 WiFi	286.40 mW	0
total	759.90 mW	4.6398 mW

with all boards, sensors, and battery installed, is below $100g$, which is around a quarter of the weight of the Tobii Pro Glasses 2.

C. Price Breakdown

Table I 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 as $\$19.80$, which is low-cost. Compared to existing commercial products like the Tobii Pro Glasses 2, SPIDERS is orders of magnitude less expensive.

D. Power Consumption

Table II lists the active and idle power consumption for all major components of SPIDERS. When sampling and transmitting data from all sensors, SPIDERS consumes around $759.9mW$. Powering the total system off of a $2000mAh$ 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 reduce power consumption further. Turning all sensors and the MCU processing unit to sleep consumes less than $5mW$ of energy.

E. 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 main board (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 near the normal body temperature of a person. The thermal images shown in Figure 3 suggests that the heat is mostly generated by the wireless MCU and the camera module. Since the temperature of SPIDERS is no higher than human body temperature, the user would not feel any discomfort from heat dissipation.

F. 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

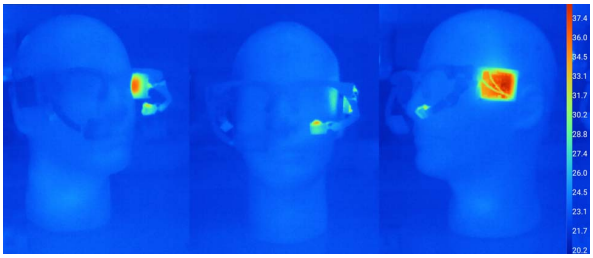


Fig. 3: Thermal image of SPIDERS in active mode.

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 V. 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. The client application computes and stores bio-signal data extracted from all of the sensors on the SPIDERS glasses platform. Users are able to view their processed bio-signal and sensor readings on the client application 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 $200ms$ or 2 seconds after the fact would likely not change the outcome or quality of care the person receives. Currently, the client application is implemented on an internet-connected desktop server. We leave a smartphone implementation for future work.

IV. 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. 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.

A. Eye & Eyebrow Shape Detection

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 6000 images using our

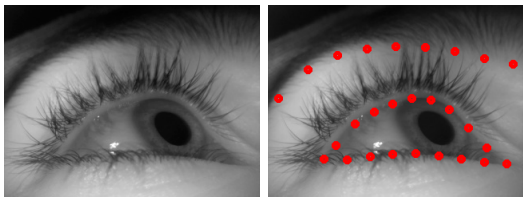
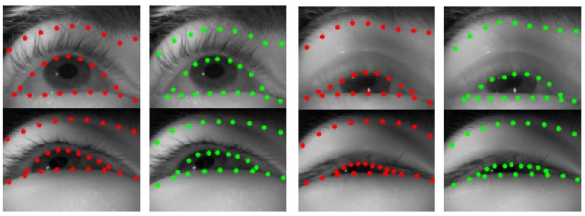


Fig. 4: Left: Example raw image of the eye. Right: Eye image with 27 landmark points generated by CNN (9 points on the upper eyelid, 9 on the lower eyelid, and 9 on the eyebrow).



(a) Regular clear images captured by SPIDERS.

(b) Extreme cases with blurry images with squint eye.

Fig. 5: 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).

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 [34]: 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 V. We manually label each image with 27 landmark points per image, as shown in Figure 4. 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. Our CNN consists of 8 convolutional layers and two fully connected layers that outputs the predicted locations of the 27 landmark points for each input image. Finally, we obtain the final contours of the eyelids and brows by fitting the generated landmark points onto a fifth-order polynomial curve.

2) Evaluation: 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 $5ms$ per input image. The average root mean square error (RMSE) of all estimated landmarks in the test set was 4.90 pixels. Figure 5 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 races, skin tone, and facial characteristics (e.g., different eye lash or hair lengths) and found similar performance across a wide range of demographics.

We show a set of eyelid and brow shape landmarks output by our model, compared to our hand-labeled ground truth, in Figure 5. 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 5.

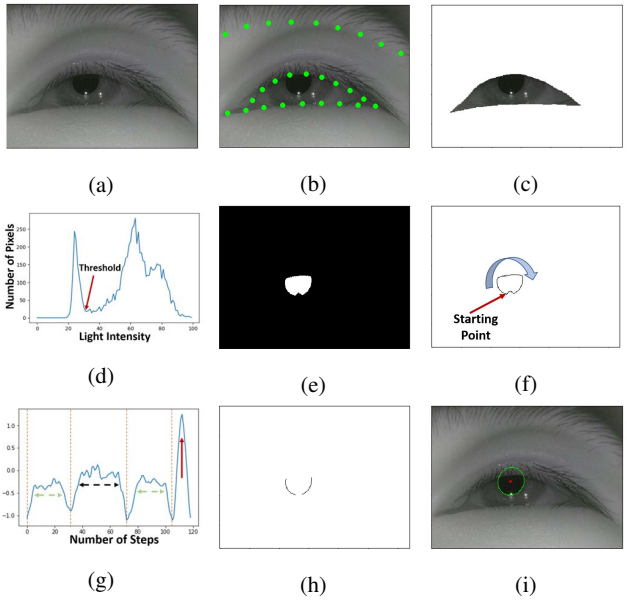


Fig. 6: Pupil data processing pipeline. (a): The original image captured by SPIDERS. (b): The 27 landmarks generated by the CNN model mentioned in Section IV-A. (c): The cropped eye area image based on the landmarks for upper and lower eyelids. (d): The color histogram of (c), which used to detect the 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.

B. Pupillometry

Our approach for robust pupillometry using our gray-scale camera is outlined in Figure 6. Determining the shape of the eye is a crucial first step in determining pupillometry as it helps in localizing the pupil by constraining the area to search for the pupil. First, we apply the eye and eyebrow shape detector, described in Section IV-A. The result of this step is shown in (a) and (b). Next, we crop out the eye area and discard the rest of the image with the knowledge of eyelid shape in (c), since the pupil can only be located within the eye area.

Next, we generate a histogram of light intensity of the pixels (d). We note that 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 the intensity correspond 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). Next, we extract the raw contour of the pupil (f). This is only a rough and noisy contour

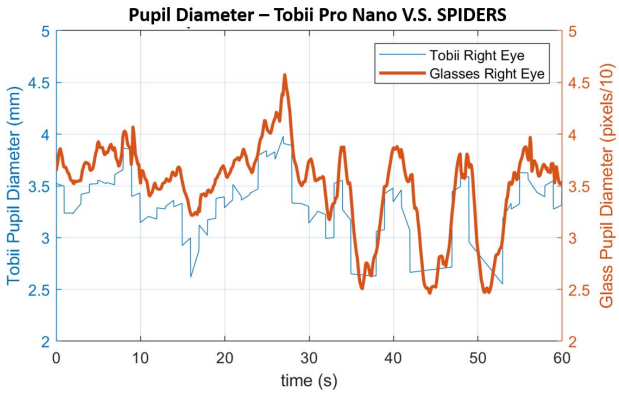


Fig. 7: A 60-second segment of pupil diameters measured by both Tobii Pro Nano and SPIDERS.

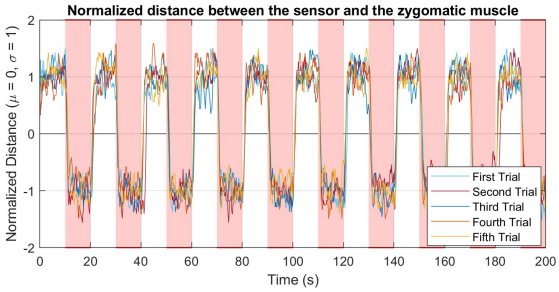


Fig. 8: The relative distances between the proximity sensor and the zygomaticus muscle while a subject alternates smile and calm facial expressions for 10 repetitions among 5 different trials (red shades: the ground truth when the subject is asked to smile).

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 contour starting from a the lowest point on the contour and calculated pixel by pixel in 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 as the critical points to chop the curvature into several segments. The sharp positive peak correspond to the portion where the raw contour is distorted by reflecting light. Those smoother curvature segments either correspond to the contour of pupil or occlusions caused by eyelids. The later 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 to the light reflections and eyelid occlusions, the segments for the pupil contour can be detected, 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).

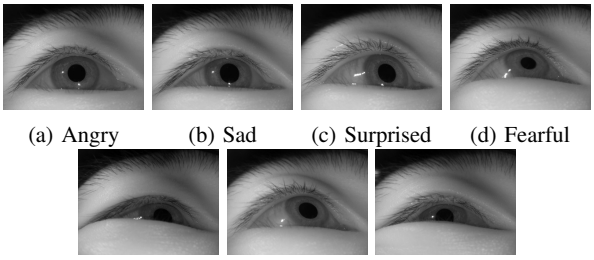
Sometimes the contour of the extracted pupil is of low quality due to a number of reasons (e.g., a person blinks, closes his eyes, or squints). To filter out frames where the state of the eye prevents SPIDERS from estimating the shape of the eye accurately, we compare the shape of the true contour extracted in step (h) and the fitted ellipse extracted in step (i) using a error of fit function (EOF_2) [35]. 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 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 $60Hz$ 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 $10Hz$.

We use both SPIDERS and Tobii Pro Nano to take measurements on six subjects from different genders and enthnicities 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 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 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 7 exhibits an example of the pupil diameters provided by both platforms.

C. Zygomaticus Movements

As described in Section III-B, 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 [36]. To evaluate whether the proximity sensor could detect movements in the zygomaticus muscle, we ask six subjects 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 for five times.



(a) Angry (b) Sad (c) Surprised (d) Fearful

(e) Disgusted (f) Neutral (g) Happy

Fig. 9: 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).

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 8 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 with the signals being highly correlated between different trials. We conducted 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 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.

V. Advanced Functionalities Prototyping

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 adequate performance of SPIDERS compared to other platforms, which are usually more expensive and bulky [10] [37] [38]. These make up the *Advanced Functionalities* layer of SPIDERS in Figure 2. All experiments presented in this section are approved by the Human Research Protection Office and IRBs at our institution.

A. 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 detection using gray-scale images and even fewer have investigated facial expression extraction using images only contain the eye area. Figure 9 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 shape of the eye between different expressions, and the inherent variations in the shape of the eye among different

TABLE III: Expressions and eye area features

Apparent 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

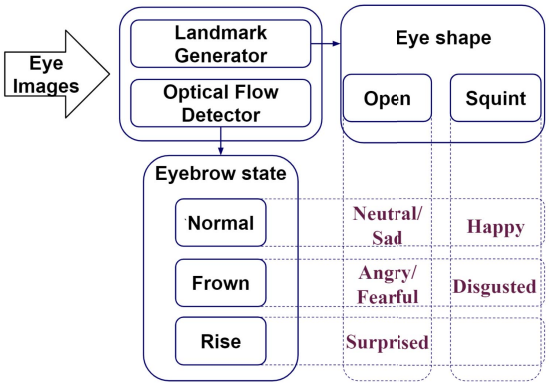


Fig. 10: The pipeline of eye-image-based facial expression (apparent emotion) detection.

demographics all make it difficult for traditional algorithms, that match the shape of the eye to an expression, to perform well.

Two feature 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 eye brow and the size change if the eye. Table III 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 (portion of eyebrow closest to the nose ridge) raises and the eyes become wide open. In the following sections, we present and evaluate our novel facial expression detection pipeline as shown in Figure 10, 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 III.

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 IV-A).
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

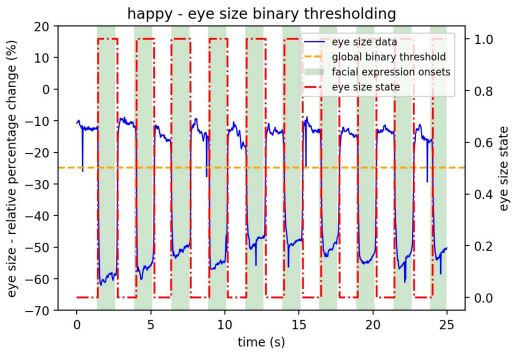


Fig. 11: Example of binary thresholding conducted on baseline corrected eye size data (green shades: the ground truth when the subject is asked to smile to express happy).

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 IV-A 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 12 shows two consecutive frames taken when a user's facial expression changing from angry to neutral, and the corresponding motion vector map generated with 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 [39] 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.

As the visualization in Figure 13 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.

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

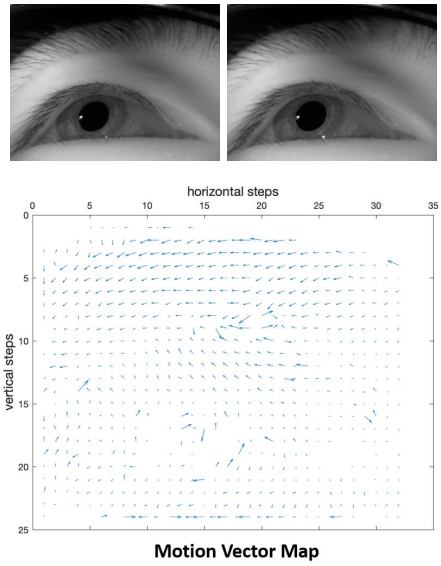


Fig. 12: Motion tracking between two consecutive image frames (top). As shown in the motion vector map (bottom), the eyebrow region moves primarily leftwards from the image on the left to the image on the right.

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 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.

2) Experiment Setup: Participants: 10 adults, including 5 male and 5 female, 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.

Stimuli and Procedure: The subjects are asked to express seven different apparent facial expressions indicated in Table III. For each apparent expression category, each subject is guided to repeat the process of explicitly performing the certain facial expression while images are taken with 10 fps frame rate. Each subject is directed to repeat the process of explicitly performing the certain facial expression for 50

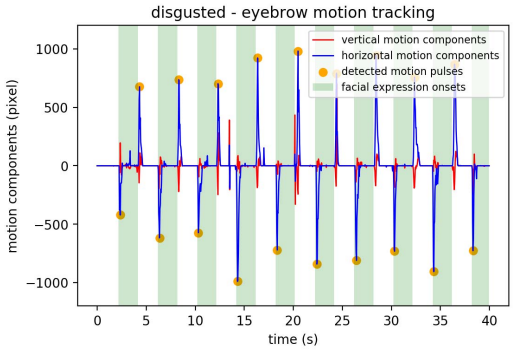


Fig. 13: Tracking of eyebrow region motion in the form of overall motion vector components over time (green shades: the ground truth when the subject is asked to perform disgusted).

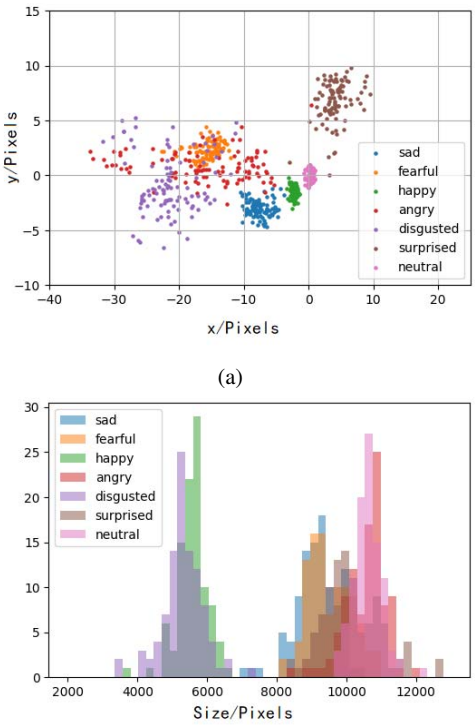
frames and 50 frames of neutral facial expression for ten times, resulting in 1000 images for one emotion.

We take the 10 frames in each 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 14. The results in these two figures support what we describe in Table III.

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 overfitting to the position of the glasses when taking certain trials of the experiment. Using 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 10. 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 15.

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



(a)

(b)

Fig. 14: (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 III.

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) [4]. It is worth noting that instead of providing a probability distribution, our method provides direct classification into 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.

B. 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.

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 on a MATLAB graphical user interface, while wearing SPIDERS. Picture and sound tests were taken from the International Affective Picture System

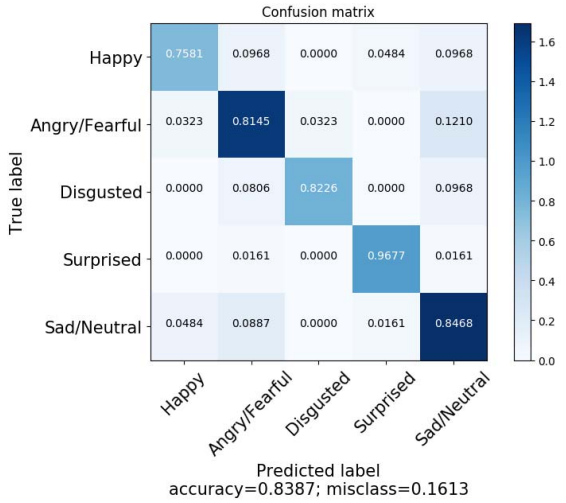


Fig. 15: Confusion matrix for facial expression classification.

(IAPS) and International Affective Digital Sounds (IADS) datasets, while videos were selected from Youtube [40] [41].

During each trial, the subject would first observe a gray scale 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. 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 [42]. 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 size and valence/arousal pairs. The confusion matrices of our results for both valence and arousal are shown in Figure 16. 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% [43] [44]. The average prediction accuracy across different subjects for three-class valence and arousal in SPIDERS are 49.32% and 47.25%, which is higher than the accuracy of other low-cost wearable devices ([16] achieved 34.2% for detecting four-class states using physiological data collected by contact-based sensors).

VI. Enabled Applications

In this section, we discuss areas and problems that SPIDERS has the potential to impact.

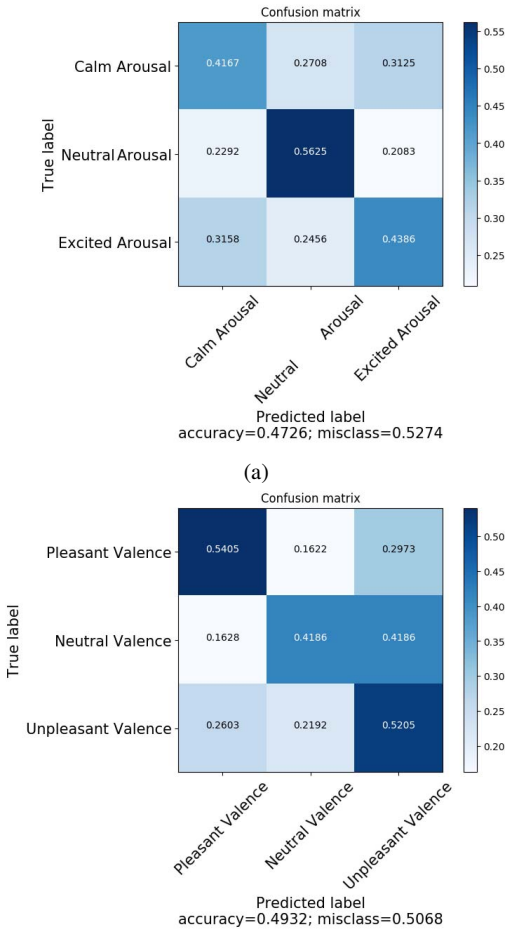


Fig. 16: Confusion matrix for the (a) arousal and (b) valence-based classes.

A. Emotional Health Monitoring

According to the World Health Organization, more than 300 million people suffer from depression and other emotional health disorders worldwide [45]. Proper treatment may require consistently monitoring the patient, which currently happens only when patients come in for appointments [46] [47]. In many scenarios, patients are unable to make frequent trips to their healthcare provider (e.g. elderly patients with trouble moving, or patients in rural/isolated areas) [1]. SPIDERS would be able to provide almost continuous emotional health monitoring of patients without the need to make frequent trips to the doctor. Additionally, patients would be able to live out their lives normally while wearing SPIDERS because the SPIDERS glasses platform is truly wireless and portable.

B. 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' emotion feedback to decide

how to make the movie more sympathetic and impressive [48]. A video game could dynamically increase the difficulty of the game if it detects that the player is getting bored or complacent [4]. A musical therapy system can utilized real-time emotional state of the user and adjust the music therapy to the patient's needs [49]. In general, media can dynamically change (and enrich) content based on the current user state.

C. 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 [50]. Moreover, with customers' emotional responses, the government can enhance the impact of public service announcements, such as advertisements for safe driving and water saving [3]. 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 they do not care for.

VII. Limitations and Future Work

The current SPIDERS system has several limitations. First, the primary 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 require continuous in-situ bio-signals monitoring. As such, in this paper, we prototype facial expression detection and emotion classification models and show that these models can be run on SPIDERS with adequate performance compared to the works done by other groups in the research community [4] [16]. 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 valid the performance of SPIDERS, we conduct the experiments in controlled situations similar to the experiment setups mentioned in [51] [52]. We will perform further research to develop algorithms that are more robust to different lighting conditions (indoor and outdoor), different users, and body movements based on the four functions in SPIDERS *Core Function Library*. We will also perform in-the-wild experiments with multiple days of continuous wearing.

VIII. 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). 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.

Acknowledgment

This research was partially supported by the National Science Foundation under Grant Numbers CNS-1704899, CNS-1815274, and CNS-1943396. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Columbia University, NSF, or the U.S. Government or any of its agencies.

References

[1] R. LiKamWa, Y. Liu, N. D. Lane, and 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*, ser. MobiSys '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 389–402. [Online]. Available: https://doi.org/10.1145/2462456.2464449

[2] K. Rowan, D. D. McAlpine, and L. A. Blewett, "Access and cost barriers to mental health care, by insurance status, 1999–2010," *Health affairs*, vol. 32, no. 10, pp. 1723–1730, 2013.

[3] N. Hamelin, O. El Moujahid, and P. Thaichon, "Emotion and advertising effectiveness: A novel facial expression analysis approach," *Journal of Retailing and Consumer Services*, vol. 36, pp. 103–111, 2017.

[4] S. Hickson, N. Dufour, A. Sud, V. Kwatra, and I. A. Essa, "Eyemotion: Classifying facial expressions in VR using eye-tracking cameras," *CoRR*, vol. abs/1707.07204, 2017. [Online]. Available: http://arxiv.org/abs/1707.07204

[5] R. R. Henderson, M. M. Bradley, and P. J. Lang, "Emotional imagery and pupil diameter," *Psychophysiology*, vol. 55, no. 5, Dec 2017. [Online]. Available: https://doi.org/10.1111/psyp.13050

[6] T. Partala and V. Surakka, "Pupil size variation as an indication of affective processing," *Int. J. Hum.-Comput. Stud.*, vol. 59, pp. 185–198, 2003.

[7] M. M. Bradley, L. Miccoli, M. A. Escrig, and P. J. Lang, "The pupil as a measure of emotional arousal and autonomic activation," *Psychophysiology*, vol. 45, no. 4, pp. 602–607, 2008.

[8] Our technology. [Online]. Available: https://www.fitbit.com/technology

[9] How it works. [Online]. Available: https://choosemuse.com/how-it-works/

[10] T. Technology. Tobii pro glasses 2. [Online]. Available: https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/

[11] A. Mayberry, Y. Tun, P. Hu, D. Smith-Freedman, D. Ganesan, B. M. Marlin, and C. Salthouse, "Cider: Enabling robustness-power tradeoffs on a computational eyeglass," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '15. New York, NY, USA: ACM, 2015, pp. 400–412. [Online]. Available: http://doi.acm.org/10.1145/2789168.2790096

[12] T. Li and X. Zhou, "Battery-free eye tracker on glasses," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '18. New York, NY, USA: ACM, 2018, pp. 67–82. [Online]. Available: http://doi.acm.org/10.1145/3241539.3241578

[13] B. Li, H. Fu, D. Wen, and W. LO, "Etracker: A mobile gaze-tracking system with near-eye display based on a combined gaze-tracking algorithm," *Sensors*, vol. 18, no. 5, 2018.

[14] K. Masai, Y. Sugiura, K. Suzuki, S. Shimamura, K. Kunze, M. Ogata, M. Inami, and 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*. New York, NY, USA: ACM, 2015, pp. 357–360.

[15] K. Masai, K. Kunze, M. sugimoto, and M. Billinghurst, "Empathy glasses," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2016, pp. 1257–1263.

[16] J. Kwon and L. Kim, "Emotion recognition using a glasses-type wearable device via multi-channel facial responses," *CoRR*, vol. abs/1905.05360, 2019.

[17] S. Xia, D. de Godoy Peixoto, B. Islam, M. T. Islam, S. Nirjon, P. R. Kinget, and X. Jiang, "Improving pedestrian safety in cities using intelligent wearable systems," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7497–7514, Oct 2019.

[18] J. Jia, C. Xu, S. Pan, S. Xia, P. Wei, H. Y. Noh, P. Zhang, and X. Jiang, "Conductive thread-based textile sensor for continuous perspiration level monitoring," *Sensors*, vol. 18, no. 11, 2018.

[19] D. de Godoy, B. Islam, S. Xia, M. T. Islam, R. Chandrasekaran, Y.-C. Chen, S. Nirjon, P. R. Kinget, and 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.

[20] D. Hong, B. Zhang, Q. Li, S. Nirjon, R. Dickerson, G. Shen, X. Jiang, and 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.

[21] V. Vezhnevets and A. Degtiareva, "Robust and accurate eye contour extraction," in *Proc. Graphicon*, 2003, pp. 81–84.

[22] Y. Ji, S. Wang, Y. Lu, J. Wei, and Y. Zhao, "Eye and mouth state detection algorithm based on contour feature extraction," *Journal of Electronic Imaging*, vol. 27, no. 5, p. 051205, 2018.

[23] A. Soetedjo, "Eye detection based-on color and shape features," *Editorial Preface*, vol. 3, no. 5, 2012.

[24] K. Sun, Y. Zhao, B. Jiang, T. Cheng, B. Xiao, D. Liu, Y. Mu, X. Wang, W. Liu, and J. Wang, "High-resolution representations for labeling pixels and regions," *arXiv preprint arXiv:1904.04514*, 2019.

[25] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3476–3483.

[26] M. Kopaczka, J. Schock, and D. Merhof, "Super-realtime facial landmark detection and shape fitting by deep regression of shape model parameters," *arXiv preprint arXiv:1902.03459*, 2019.

[27] M. J. R. Oliva and A. Anikin, "Pupil dilation reflects the time course of emotion recognition in human vocalizations," in *Scientific Reports*, 2018.

[28] V. Raudonis, G. Dervinis, A. Vilkauskas, A. Paulauskaite-Taraseviciene, and G. Kersulyte-Raudone, "Evaluation of human emotion from eye motions," *Evaluation*, vol. 4, no. 8, 2013.

[29] J. C. A. Costa, F. Guimbretière, M. F. Jung, and T. Choudhury, "Boostmeup: Improving cognitive performance in the moment by unobtrusively regulating emotions with a smartwatch," *IMWUT*, vol. 3, pp. 40:1–40:23, 2019.

[30] B. M. Appelhans and L. J. Luecken, "Heart rate variability as an index of regulated emotional responding," *Review of general psychology*, vol. 10, no. 3, pp. 229–240, 2006.

[31] M. Li and 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.

[32] M. Li, H. Xu, X. Liu, and S. Lu, "Emotion recognition from multichannel eeg signals using k-nearest neighbor classification," *Technology and Health Care*, vol. 26, no. S1, pp. 509–519, 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/29758974

[33] H. Dabas, C. Sethi, C. Dua, M. Dalawat, and D. Sethia, "Emotion classification using eeg signals," in *Proceedings of the 2018 2Nd International Conference on Computer Science and Artificial Intelligence*, ser. CSAI '18. New York, NY, USA: ACM, 2018, pp. 380–384. [Online]. Available: http://doi.acm.org/10.1145/3297156.3297177

[34] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[35] L. Świrski, A. Bulling, and N. Dodgson, "Robust real-time pupil tracking in highly off-axis images," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, ser. ETRA '12. New York, NY, USA: ACM, 2012, pp. 173–176. [Online]. Available: http://doi.acm.org/10.1145/2168556.2168585

[36] B. Farnsworth. (2019) Zygomaticus major – relating muscle movement to emotion. [Online]. Available: https://imotions.com/blog/zygomaticus-major/

[37] The imotions platfrom. [Online]. Available: https://imotions.com/platform/

[38] M. Kassner, W. Patera, and 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.

[39] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[40] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (iaps): Technical manual and affective ratings," *NIMH Center for the Study of Emotion and Attention*, vol. 1, pp. 39–58, 1997.

[41] M. M. Bradley and 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*, 2007.

[42] ——, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[43] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE transactions on affective computing*, vol. 3, no. 2, pp. 211–223, 2011.

[44] Y. Lu, W.-L. Zheng, B. Li, and B.-L. Lu, "Combining eye movements and eeg to enhance emotion recognition," in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[45] W. H. Organization, "Depression," March 2018. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[46] M. Vidal, J. Turner, A. Bulling, and H. Gellersen, "Wearable eye tracking for mental health monitoring," *Computer Communications*, vol. 35, no. 11, pp. 1306–1311, 2012.

[47] J. A. Naslund, K. A. Aschbrenner, and S. J. Bartels, "Wearable devices and smartphones for activity tracking among people with serious mental illness," *Mental health and physical activity*, vol. 10, pp. 10–17, 2016.

[48] K. Kimura, S. Haramizu, K. Sanada, and A. Oshida, "Emotional state of being moved elicited by films: A comparison with several positive emotions," *Frontiers in Psychology*, vol. 10, p. 1935, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2019.01935

[49] O. Sourina, Y. Liu, and M. K. Nguyen, "Real-time eeg-based emotion recognition for music therapy," *Journal on Multimodal User Interfaces*, vol. 5, no. 1, pp. 27–35, 2012. [Online]. Available: https://doi.org/10.1007/s12193-011-0080-6

[50] N. Hristova and 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*, Jan 2004, pp. 10 pp.–.

[51] W.-L. Zheng, B.-N. Dong, and 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.

[52] E. Gatti, E. Calzolari, E. Maggioni, and M. Obrist, "Emotional ratings and skin conductance response to visual, auditory and haptic stimuli," *Scientific data*, vol. 5, p. 180120, 2018.