Drowsiness Detection using Contactless Heart Rate Monitoring and Eye Tracking

Bradley Barnhart
bbarnhar@stanford.edu

Kartik Prabhu
kprabhu7@stanford.edu

Arun Seetharaman
aseethar@stanford.edu

***Abstract*—Detecting drowsiness has many important applications, particularly in driver safety. Blink duration and heart rate variability are two important metrics that can be used to determine how drowsy someone is, and can be determined using image processing techniques. In order to determine blink duration, training images of open and closed eyes are used to generate a fisherimage using Fisher linear discriminant analysis. This fisherimage can then be used to determine whether the eyes are open or closed at a given time, allowing for determination of blink duration. In order to determine heart rate variability, two methods were used to determine the blood volume pulse, independent component analysis and chrominance based. Eye detection yielded very good results, with 71% to 97% accuracy in classifiying open/closed eyes. Heart rate estimates were a little less accurate, with a mean error of roughly 16 BPM for the ICA method and 13 BPM for the chrominance based method. Overall, these results show that detecting drowsiness using simple RGB cameras is very promising.**

***Index Terms*—eye tracking, remote ppg, independent component analysis, fisher LDA, chrominance, openCV**

I. Background

Driving while fatigued is very similar to driving while intoxicated, as drowsy drivers exhibit much of the same behaviors as intoxicated drivers: slow reaction times, inability to concentrate, and bad awareness of their surroundings. A study by the AAA Foundation for Traffic Safety found that 21% of fatal crashes involved a drowsy driver [1]. The problem is particularly bad as it can be difficult for drivers to recognize signs of fatigue in themselves.

However, there are several physiological signs that can be used to determine if a driver is drowsy. One such sign is, unsurprisingly, trouble keeping eyes open. By tracking whether the eyes are open or closed, the blink duration can be found, which can then be used to determine how much trouble the driver is having to stay awake. Another characteristic is heart rate variability, which is the variation in time between heartbeats. Drowsy and non-drowsy people exhibit different heart rate variability metrics [2]. With this in mind, our goal is not to actually build a drowsiness detector but to explore the possibility of creating one using a simple RGB camera like the ones found in smart phones and laptops.

II. Related Works

There has been sufficient work done in measuring a person's heart rate and gaze using simple cameras like the ones available on most modern phones and laptops [3] [4]. While these studies did not explicitly explore drowsiness detection we believe that the methods described in these papers would prove useful for drowsiness detection. Heart rate detection using footage of the frontal face has been accomplished through a variety of methods using the change in average color over time [5], [6]. By detecting heart rate it stands to reason that we could then determine heart rate variability. Likewise there has been many different methods established for tracking eye movement using frontal face footage [4]. Although we are not interested in actual gaze detection we believe the fundamentals of segmenting the eyes found in these papers could be applied to our task of segmenting the eyes and then determining whether it is open or closed.

III. Preprocessing

All of our algorithms work in real time and work on a frame by frame basis for a video either from a webcam or a video file. First, the face is identified with the Viola-Jones object detection algorithm using pretrained cascade classifiers from the OpenCV toolbox [7]. This step outputs coordinates for the bounding box for the face, which is then used to crop the frame to only include the face.

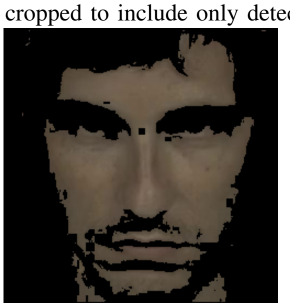
Further processing is needed on these cropped frames to extract regions of interest for both the eye detection and heart rate detection algorithms. In order to segment the eyes, another cascade classifier pretrained for eyes is used on the cropped frame to determine bounding boxes for the eyes. For the heart rate detection algorithm, two different region of interests were computed and evaluated, all facial skin and only cheeks. The facial skin was found by masking the image in the HSV colorspace. We ultimately discarded the HSV mask since we found that it excluded darker skin tones while potentially accepting some of the background making it ineffective, as seen in Figure 2. The cheeks were found by proportionally selecting two boxes from the bounding box of the face.

IV. Eye Detection

A. Training

1) Fisher LDA: To determine blink rate and blink duration, it was necessary to determine if eyes were open or closed. Because of the binary nature of this problem, and the predictable differences between open and closed eyes, a fisherimage was obtained from training data to classify the eyes.

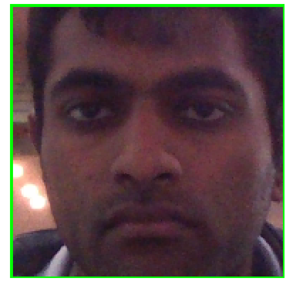
(a) Frame cropped to include only detected face.



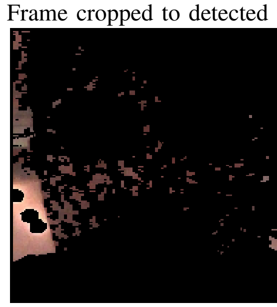
(b) Face masked with HSV skin mask.

(c) Face with bounding boxes for the cheeks.

Fig. 1: Results from the preprocessing step.



(a) Frame cropped to detected face.



(b) Face masked with HSV skin mask.

Fig. 2: Example of the skin mask working improperly.

2) Data Collection: The training data for the first user consisted of approximately 100 images of closed eyes and 100 images of open eyes. To collect the data, 20 images were taken at a time, at a sampling rate determined by the frame rate of the webcam. This was repeated 5 times for each class at different locations, illuminations, and distances from the camera.

3) Data Pre-Processing: Before performing linear discriminant analysis, the images were pre-processed to account for variations in illumination and alignment. First, global histogram equalization was applied to the images to enhance the contrast, which was especially important for images acquired in dim lighting. Next, two alignment protocols were executed. In both cases, the minimum MSE was used as the alignment criterion, and the algorithm was coarse-to-fine with step sizes of 2 and 1. The objective of the first alignment was to obtain a single image that could be used as a template for all future alignments. This process involved aligning each open eye to a single open eye, followed by aligning each closed eye to a single closed eye, and averaging across all images. This meaneye was then used for all future alignments, including training. Finally, once images were aligned, a 30x30 cropped region was extracted from the center of the image to ensure the sizes of the eyes were consistent.

4) Linear Discriminant Analysis: The top 100 eigenimages were computed from the training data using the Srirovich and Kirby algorithm. Fisher linear discriminant analysis was used to compute fisherimages from the eigenimages. Only the first fisherimage was used for analysis. Once the fisherimage was obtained, the training data was projected onto the fisherface and an appropriate threshold was chosen for future classification.

B. *Testing*

To test the eye classification performance, 200 consecutive frames were obtained for the user in a new setting. A chart listing positives, negatives, false positives, and false negatives was obtained (positive = closed, negative = open).

C. *Variations in User Appearance*

The previous methods were performed on one user. But the fisherimage obtained from the training data of one user is less likely to perform as well on other users, especially if the users vary significantly in skin tone. Below, user 1 refers to a User of a light skin tone, and User 2 refers to a user of a dark skin tone. The first test involved applying the fisherimage of user 1 on testing data from user 2. The second test involved combining training data for User 1 and User 2 to generate a new fisherimage, and then using this to classify eye images from both users.

D. *Calculating Blink Duration*

While the program is running, the blink duration of the previous blink is calculated in real time. This information is displayed in the command line, along with the blink status (binary, open or closed). One post-processing step was used to help eliminate false negatives, which was important when calculating blink duration. If one or two consecutive negatives (open) were discovered between any two positives (closed), the negatives were corrected to positives. This is valid because blinks typically do not happen in rapid succession, and if they do, it is appropriate to consider them one continuous blink since the eyes are not open for long enough to collect information.

V. Heart Rate Detection

As the heart pumps blood through the body, the volume of blood flowing through the blood vessels changes. This variation in blood volume is visible in color, and by monitoring the RGB signal, the blood volume pulse can be determined, from which the heart rate and other related metrics can be found.

Two different methods were used to estimate the blood volume pulse, independent component analysis (ICA) and chrominance-based. These two methods used a shared pipeline, as there were some similar steps. Using the region of interest found in the preprocessing step, the average red, green, and blue values are calculated and stored, generating a time series signal for the color channels.

A. Independent Component Analysis Method

ICA is a method that tries to find an optimal linear transform to turn n input signals into n statistically independent output signals. In practice, many papers showed that performing ICA on the three normalized color channels of a person's skin yielded one output signal that corresponded to blood volume pulse [5] [3]. Equation 1 shows how the red channel was normalized with the green and blue channels following the same procedure. Equation 2 demonstrates the relationship between input $x(t)$ and output $s(t)$ and the unmixing matrix W that ICA solves for. Figure 3 also illustrates what ICA accomplishes. While ther reference paper used the JADE algorithm, we used FastICA from the scikit toolbox since we could find no functional difference [5].

$$r_{norm} = (r - \mu_r)/\sigma_r \tag{1}$$

$$s(t) = Wx(t) \tag{2}$$

First, we use FastICA on the normalized red, green, and blue signals to determine an unmixing matrix that will be saved and used for all future videos and webcam footage. This process can be thought of as using ICA to train our unmixing matrix. While we initially ran FastICA multiple times to determine a new unmixing matrix every few iterations we decided it was better to come up with a single unmixing matrix to keep results consistent between different sources. Another reason was by keeping one matrix we did not have to dynamically decide which of the three signals corresponded to blood volume pulse each time.

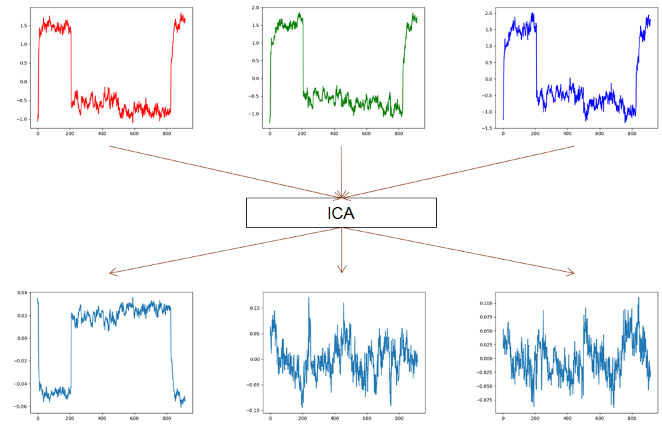


Fig. 3: Color Channels Transformed into Indpendent Signals

However this approach makes the choice of one of the three output signals as blood volume pulse even more important. Unfortunately none of the signals seemed analogous to blood volume pulse by visual inspection so we chose the signal by taking the FFT of each of the three signals and looking at the dominant frequency. We only looked at frequencies between 0.67 Hz and 3 Hz which correspond to a range of reasonable heartbeats between 40 bpm and 180 bpm respectively. Using the ECG data of the subject in the video we picked signal whose dominant frequency had sufficient magnitude and whose frequency was closest to the heart rate of 71.25 bpm calculated from the ECG. The output of the matrix that corresponded to this signal is the one used to estimate blood volume pulse and heart rate.

Once the linear transform was settled upon we used it in our script to estimate blood volume pulse from the color channel time series signals. Ten seconds worth of color channel data is analyzed at a time. The heart rate is estimated by taking the FFT of the blood volume pulse and looking at the dominant frequency between 0.67 Hz and 3 Hz and then multiplying the frequency by 60 to turn Hz to bpm. We did originally try using bandpass filtering in the process to emphasize the frequencies of interest but found that there was no appreciable increase in performance.

B. Chrominance Based Method

Haan et al. proposed a chrominance based method to calculate heart rate, as an alternative to blind-source separation methods, like independent component analysis [6]. They reason that this method is more robust to motion. Two types of reflections are captured by a camera as light reflects off of skin: specular and diffuse. The specular reflection is related only to the light source, and not to any physiolocal characteristics, while the diffuse reflection is light that has been scattered from the blood vessels, and thus contains the information needed to determine the blood volume pulse.

The authors propose that by assuming a standardized skin tone, which they found to be similar across all skin types, temporally normalized RGB can be projected into the plane

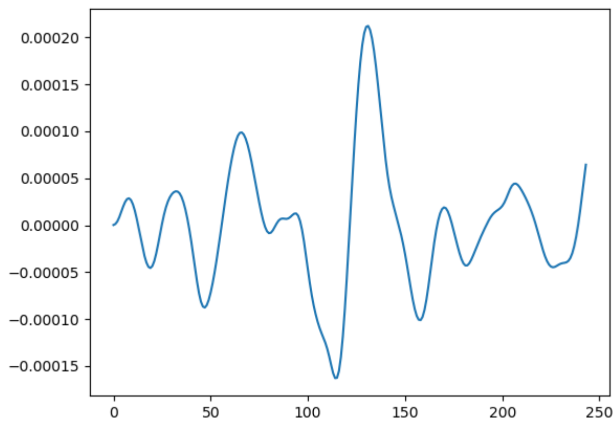


Fig. 4: Signal showing the blood volume pulse in the time domain. The peak to peak time appears to be roughly 60 frames, for a video recorded at 60 fps, corresponding to a heart rate of 60 BPM.

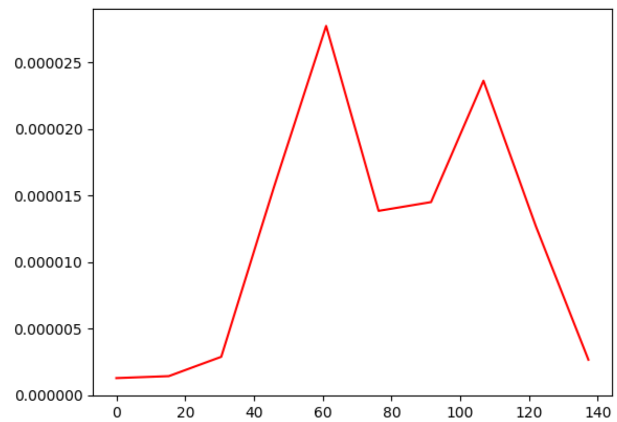


Fig. 5: The signal in the frequency domain, showing a peak at roughly 60 beats per minute.

orthogonal to the specular reflection component, generating two chrominance signals, X and Y. These signals are band-pass filtered, using a Butterworth bandpass filter of order 3 for frequencies 0.67-3 Hz, and then used to compute the blood volume pulse, S. Equations 3-6 summarize the computations.

$$X = 3R - 2G \tag{3}$$

$$Y = 1.5R + G - 1.5B \tag{4}$$

$$\alpha = \frac{\sigma(X_{filtered})}{\sigma(Y_{filtered})} \tag{5}$$

$$S = X_{filtered} - \alpha Y_{filtered} \tag{6}$$

The FFT of the blood volume pulse is computed, and the heart rate is then determined by the dominant frequency in the range 0.67 - 3 Hz.

Figure 4 shows what the signals looks like in the time domain, while figure 5 shows the signal in the frequency domain.

VI. Evaluation

We evaluated our eye detection and heart rate detection algorithms on a dataset of videos, as well as on healthy participants with a wide range of skin types. The dataset used is MAHNOB-HCI, which contains videos of participants watching video clips and their physiological data, including ECG signals, from which a ground truth for the heart rate can be determined [8].

A. Eye Detection

1) Eye Classification- Training: Figure 6 shows the fisher-image, and figure 7 shows the projections of the training data onto the fisherimage for User 1.

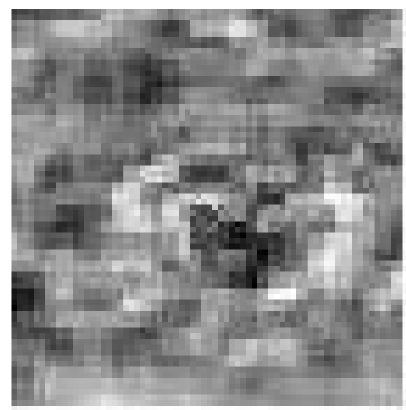


Fig. 6: Fisherimage

2) Eye Classification- Testing: Table I shows User 1 testing results for blink classification of 200 images. A total of 12 blinks occurred, lasting 1-8 frames.

3) Variations in User Appearance: Applying the fisher-image of User 1 on testing data from User 2 gave very poor results for both open and closed eyes. It was quickly determined that a fisherimage trained exclusively for User 1 would not be useful for User 2. Therefore, we combined training data for User 1 and 2 and trained a new fisherimage. The histogram of the projections of the training data on the fisherimage is shown in Figure 9.

Table II shows User 1 testing results for blink classification of 200 images. A total of 11 blinks occurred, lasting 1-8

Truth	Correctly Classified	Incorrectly Classified
Closed	40	0
Open	158	2

TABLE I: Testing results for blink classification of User 1 with 200 images.

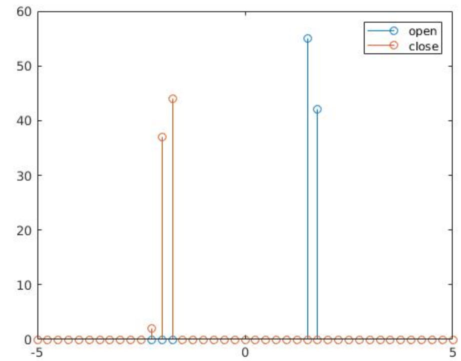
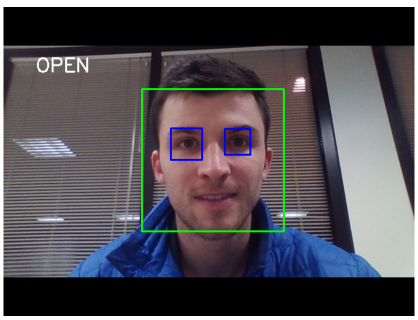
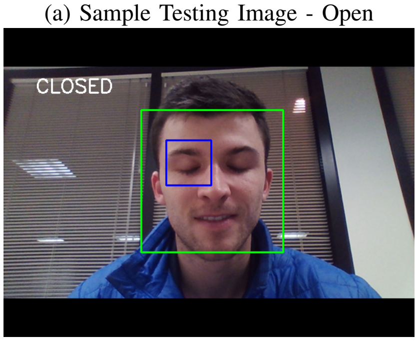


Fig. 7: Training data projections on fisherface obtained from User 1 training data



(a) Sample Testing Image - Open



(b) Sample Testing Image - Closed

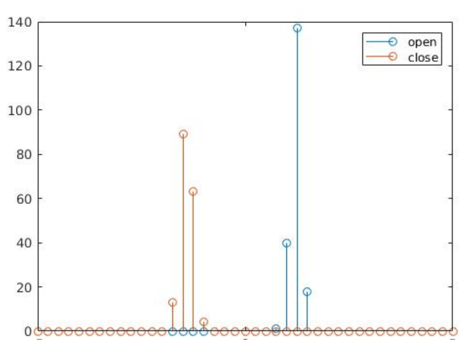


Fig. 9: Training data projections on fisherface obtained from data from User 1 and User 2

Truth	Correctly Classified	Incorrectly Classified
Closed	35	6
Open	159	0

TABLE II: Testing results for blink classification of User 1 with 200 images, without correcting for false negatives.

Truth	Correctly Classified	Incorrectly Classified
Closed	41	0
Open	159	0

TABLE III: Testing results for blink classification of User 1 with 200 images, after correcting for false negatives.

frames.

Table III shows User 1 testing results for blink classification of 200 images, after correcting for false negatives using the data post-processing step mentioned in section IV-D.

Table IV shows User 2 testing results for blink classification of 200 images. A total of 9 blinks occurred, lasting 1-10 frames. For the 58 incorrectly labeled eyes, the OpenCV pretrained classifier did not detect either eye in 48 cases, and did not detect more than one eye in the other 10 cases.

B. Heart Rate Detection

In order to evaluate our heart rate detection methods we compared the reported heart rate over time to the ground truth heart rate. For videos in the dataset we used the heart rate calculated from the ECG data as the ground truth and for webcam footage we used a PPG sensor on a smartphone to determine the ground truth. We calculate the heart rate to be the median of the recorded heart rates for 30 seconds of data. Table V summarizes our findings.

The mean error for the ICA method is around 16 BPM, while for the chrominance method is around 13 BPM. These numbers are thrown off by a relatively inaccurate trial, the one with a ground truth of 120 BPM. Excluding this bad trial, the mean error drops to 10 BPM for ICA and 8 BPM for ICA.

Truth	Correctly Classified	Incorrectly Classified
Closed	3	58
Open	139	0

TABLE IV: Testing results for blink classification of User 2 with 200 images.

Truth	ICA		Chrominance	
	Median	Standard Deviation	Median	Standard Deviation
Participants				
82	71	30	78	22
50	54	5.08	58	2.86
48	72	29	69	21
120	65	22	76	26
Videos				
71	67	25	73	30
71	64	26	80	42
73	65	3.21	77	19

TABLE V: Median and standard deviation of heart rate in beats per minute for the ICA and chrominance methods.

VII. Discussion

A. Eye Detection

The initial training for User 1 showed that the fisherimage separated the training data very effectively and would provide at least a good starting point for distinguishing between open and closed eyes.

User 1 testing was 99% accurate. The two false positives for closed eyes can be explained by the bounding box around the eye. It was observed that as the eyelid opened, the pretrained classifier gave a bounding box that was much smaller around one of the eyes, which may have caused the false negative. Since this occurred while the eye was opening, it caused the blink duration to be slightly overestimated.

User 1 testing was still 97% accurate when using the fisherimage obtained from users 1 and 2. This is significant because User 1 had light skin and User 2 had dark skin. The 6 false negatives were removed by the post-processing step when calculating blink duration (see methods), so there were no true errors. This indicates that it is possible to perform the classification task for multiple users using the same fisherimage. This would be important if multiple drivers shared the same vehicle.

User 2 testing was only 71% accurate, but this can be attributed to problems with the pretrained classifier. In 48/58 cases no eyes were detected, and only one eye was detected in the remaining 10. In addition, the neck was detected as an extra face in almost every image, the closed eyes that were detected had unreliable size and position, and some of the bounding boxes overlapped. Since the program runs in real time, it is necessary to detect eyes for each frame, so if an eye is not detected correctly, the algorithm does not have a chance to correctly classify the image. This was much more a problem for User 2 than User 1. The pretrained classifier is extremely poor at detecting closed eyes in users with dark skin under certain lighting conditions, which is a limitation in the tools used for this process, rather than the algorithm itself. Labeling an image with no eyes as closed would improve the accuracy of the testing results, but would not be valid in a real drowsiness detector, so that method was not considered.

Overall, the accuracy of the detecting blinks and estimating blink duration is very good, especially for lighter skinned users. Blink duration is able to be calculated from the algorithm; typical blinks last 200-600 milliseconds. For darker skinned users, limitations in the accuracy of the bounding boxes obtained from OpenCV prevented closed eyes from being detected reliably.

B. Heart Rate Detection

The results were relatively good, with only one bad trial. Although not perfect, our methods were able to get quite close to the truth. For most trials, the measured heart rate varied quite a bit, as seen by the standard deviation. In order to mitigate the effects of these outliers, we used the median as the measure of central tendency. The two methods, ICA and chrominance, try to reduce the effects of noise, however, they are not perfect, and thus result in some times where the true blood volume pulse is not detected. In addition, the videos from the dataset are compressed, which has been found to significantly degrade the signal-to-noise ratio for the blood volume pulse [9]. We found it very difficult to figure out how to improve the accuracy of our algorithm, as it isn't very easy to isolate sources of noise. Instead, we found ourselves iterating on tweaking various values and testing performance in order to gain better results.

VIII. Conclusion

Overall, this project proved more difficult than we anticipated. The accuracy of eye detection is excellent under certain conditions and users, but in some conditions it is limited by the pretrained OpenCV classifiers, especially for darker skinned users. Exploring alternate methods to obtain bounding boxes around the eyes would be a future improvement. With additional training and refinement it could serve as a great detector of ocular indicators of drowsiness, especially blink duration.

Although we lack knowledge of the underlying biophysical characteristics, we were able to implement a relatively good heart rate monitor. The performance was less accurate than reported in other papers, despite following their procedures, likely due to the fact that our evaluation was not performed in a very controlled environment. For our application of monitoring drivers, the environment is not very controlled, and therefore we tried to obtain results that would be in line with real world results. The estimated heart rate varied wildly every once in a while, which means that heart rate variability metrics are likely to not be very accurate. However, the fact that we were able to even get a ballpark estimate of the heart rate from a simple RGB camera is a great feat, in and of itself. This project is a first step towards a proper drowsiness detector, and the fact we were able to run everything in real time is very encouraging and suggests that with refinement of our algorithms we could design a more reliable drowsiness detector.

References

[1] B. Tefft, "Prevalence of motor vehicle crashes involving drowsy drivers, united states, 2009 - 2013," 2014.

[2] J. Vicente, P. Laguna, A. Bartra, and R. Bailón, "Drowsiness detection using heart rate variability," *Medical & biological engineering & computing*, vol. 54, no. 6, pp. 927–937, 2016.

[3] C. Wang, T. Pun, and G. Chanel, "A comparative survey of methods for remote heart rate detection from frontal face videos," *Frontiers in Bioengineering and Biotechnology*, vol. 6, p. 33, 05 2018.

[4] A. Al-Rahayfeh and M. Faezipour, "Eye tracking and head movement detection: A state-of-art survey," in *IEEE Journal of Translational Engineering in Health and Medicine*, 2013.

[5] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation." *Opt. Express*, vol. 18, no. 10, pp. 10 762–10 774, May 2010. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-18-10-10762

[6] G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rppg," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, 2013.

[7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *null*. IEEE, 2001, p. 511.

[8] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2012.
[9] D. J. McDuff, E. B. Blackford, and J. R. Estepp, "The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*. IEEE, 2017, pp. 63–70.

APPENDIX

Bradley Barnhard worked on eye detection.

Kartik Prabhu worked on heart rate detection using chrominance based method and heart rate preprocessing.

Arun Seetharaman worked on heart rate detection using independent component analysis and heart rate preprocessing.