Improvements in Remote Cardio-Pulmonary Measurement Using a Five Band Digital Camera

Daniel McDuff, Student Member, IEEE, Sarah Gontarek, and Rosalind W. Picard, Fellow, IEEE

Abstract—Remote measurement of the blood volume pulse via photoplethysmography (PPG) using digital cameras and ambient light has great potential for healthcare and affective computing. However, traditional RGB cameras have limited frequency resolution. We present results of PPG measurements from a novel five band camera and show that alternate frequency bands, in particular an orange band, allowed physiological measurements much more highly correlated with an FDA approved contact PPG sensor. In a study with participants (n=10) at rest and under stress, correlations of over 0.92 (p<0.01) were obtained for heart rate, breathing rate and heart rate variability measurements. In addition, the remotely measured HRV spectrograms closely matched those from the contact approach. The best results were obtained using a combination of cyan, green and orange (CGO) bands; incorporating red and blue channel observations did not improve performance. In short, RGB is not optimal for this problem: CGO is better. Incorporating alternative color channel sensors should not increase the cost of such cameras dramatically.

Index Terms—heart rate variability (HRV), blood volume pulse (BVP), photoplethysmography (PPG), remote sensing.

I. INTRODUCTION

REMOTE detection of physiological parameters holds great potential for healthcare and affective computing. Applications that would benefit from non-contact measurement of heart rate (HR) and heart rate variability (HRV) include: infant monitoring [1], detection of cardiac diseases [2] and stress monitoring [3]. The current gold standard methods of measuring HR and HRV involve obtrusive devices attached to the body, in some cases requiring sticky gels and/or uncomfortable electrodes.

Heart rate variability spectrograms are a useful non-invasive measure of phenomena such as the cardiac regulatory system response [4], anxiety [5], sleep patterns [6] or cognitive stress [7]. Mental arithmetic tasks can increase low frequency components and low frequency/high frequency ratios in power spectral analysis of the heart rate variability [7]. We use a mental arithmetic task to cause changes in HRV and show that these can be measured accurately using a digital camera.

Photoplethysmography (PPG) is a low-cost and non-invasive technique for measuring the cardiovascular blood volume pulse (BVP) through variations in transmitted or reflected light [8]. Traditionally a dedicated light source and specialized sensor (e.g. IR light) are used to measure the PPG signal. However, recent work has demonstrated the measurement of the pulse signal using ambient light [9]. Furthermore, it is possible to accurately measure cardio-pulmonary parameters (heart rate, breathing rate (BR) and high and low frequency components of the HRV) using ambient light and a low-cost camera [10]. However, it was not shown whether detailed information about the HRV spectrogram (HRVS) and subtle changes over time could be measured using this approach. In addition, participants were seated close to (~0.5m) the camera. We present results that show we can recover, with high accuracy, both physiological parameters and HRV spectrograms from videos of the face taken using a digital camera placed 3m from the participant. This increased range opens up more potential applications in which remote PPG could be practically used. In addition, this new method shows performance improvements over the state of the art. As in our prior work the method works in ambient light, does not require a dedicated light source, and works well regardless of skin color.

Most digital single-lens reflex (DSLR) cameras capture three color channels (RGB) with 16-bits/channel. We use a novel DSLR sensor that has the capability to capture five color channels (16-bits/channel): red, green, blue, cyan and orange (RGBCO). Each pixel on the camera sensor measures one color. The sensor we use has pixels for detecting light in the orange and cyan frequency bands as well as pixels for detecting light in the red, green and blue bands. Therefore, we are able to measure more specific frequency information. Previous work has shown the green channel in a traditional RGB camera to capture the strongest BVP signal [9]. We performed experiments with different combinations of color channels and show that the set of channels including the orange band signals perform much better than the green signal alone and much better than the set of RGB signals.

Poh et al’s [10] method for recovering the BVP waveform from video uses independent component analysis (ICA). In its traditional form the number of source signals cannot exceed the number of observations. Therefore, by allowing for more observations (more color channels) we have greater flexibility in the number of source signals that may be present. Considering that there may be many sources of noise (e.g. lighting changes, rigid head motion, facial expressions, camera sensor noise) we test whether more observations will allow for more accurate recovery of the BVP.

The contributions of this paper are: 1) to show that a five band digital camera allows highly accurate measurement of physiological parameters from a distance of 3m and outperforms the traditional three color bands (RGB), 2) to show that alternative color bands (orange and cyan) between the red and
blue bands are particularly useful in recovering the BVP and 3) to present the first examples of remotely measured HRV spectrograms from individuals in both relaxed conditions and under cognitive stress that captures sympathetic modulation. The results presented suggest that subtle changes in the high and low frequency components of the HRV can be measured using this new approach.

II. RELATED WORK

Remote measurement of vital signs has been demonstrated using a number of methods. HR and BR measurements have been shown using laser doppler [11], microwave doppler [12], milliwave doppler [13] and thermal imaging [14], [15] methods. Verkruysse et al. [9] showed that PPG measurements could be made using ambient light in the visual spectrum. Poh et al. [16] presented a practical method allowing the BVP to be recovered using a low-cost webcam, which can then be used to calculate HR, BR and high and low frequency components of heart rate variability [10]. However, that work did not show that it was possible to gain an accurate measurement of HRV changes over time (e.g. showing accurate recovery of HRV spectrograms) and all measurements were taken with participants in a restful state (not under stress) at a distance of 0.5m.

Motion [16] can impact the accuracy of PPG measurements made using ambient light and a digital camera. However, there are many applications in which remote measurement of physiology could be very useful where there is little rigid head motion and slowly changing ambient lighting. Motion compensation is also possible in cases where motion artifacts may be problematic [17]. A method of magnifying the PPG signal in a video of the face allows the signal to be visualized [18], although the filtering frequencies need to be set manually. Balakrishnan et al. [19] presented a method for recovering the BVP waveform from motion of the human head allowing the heart rate to be detected even if no skin is visible. However, their system was outperformed by camera-based PPG measurements in some cases and is likely to be susceptible to noise due to rigid head motions. A pilot study in infant monitoring has demonstrated accurate measurement of HR using remote PPG measurements from a camera in realistic conditions [1], HRV measurements were not validated.

III. METHODS

A. Camera

The camera used to collect the video sequences for analysis was a standard single-lens reflex (DSLR) camera. The lens used was a standard Zuiko 50mm lens. The camera’s sensor has the capability of capturing five color bands including the typical three frequency band sensors (red, green and blue (RGB)) and also cyan and orange frequency band sensors (RGBCO). Figure 1 shows the sensitivities for the five band camera. In other respects it was a standard DSLR camera. Figure 2 shows the five band camera sensor layout. The image shows the arrangement of the colors in a 4x4 pattern that repeats across the sensor. Each pixel on the sensor measures one color as determined by its position. Further details about the sensor and demosaicking can be found in [21].

The raw image values captured by the camera can be calculated using the following formulation:

\[
m = \sum_{\lambda} e(\lambda)s(\lambda)
\]

Where \(e(\lambda)\) is the energy of light at a given wavelength \(\lambda\) and \(s(\lambda)\) is the camera sensitivity profile for a certain color channel. We compare the performance recovering physiological parameters using all 31 possible combinations of the color bands in Section IV. Custom image capture software was used to record raw images of each frame of video. All the videos were recorded with a frame rate of 30 frames per second (fps) and a resolution of 960 x 720. The recording were in color (80-bit image with five channels x 16 bits/channel).

B. Contact Sensors

For comparison of the camera measurements with a contact sensor, BVP, respiration and electrodermal activity (EDA) signals were measured using FDA-approved sensors (Flexcomp Infiniti by Thought Technologies, Inc.). BVP was calculated via the PPG signal from the index finger tip on the left hand. EDA was measured with finger sensors on the middle and ring fingers of both hands and respiration was measured using a chest strap. For the validation and analysis here we only consider the BVP and respiration measurements.

C. Experiments

This study was approved by the Institutional Review Board of the Massachusetts Institute of Technology (COUHES). All experiments were conducted indoors and with a varying amount of ambient light, provided by a changing combination.
of sunlight through a nearby window and indoor illumination. Participants were seated and the data were recorded on a laptop (Toshiba laptop running Windows 7).

Our experiments featured 10 participants of both genders (seven females), different ages (18-30) and multiple skin colors (Asian, Caucasian, Hispanic). Two participants were wearing glasses and one had facial hair. During the experiment participants were seated approximately 3m from the camera and asked to face the camera while their video was recorded. Figure 4 shows the setup used to record the data. Two-minute recordings of the participants were taken; the contact measurements and video sequences were time aligned by starting the recordings simultaneously.

**Measurements at rest:** In the first experiment participants were told to sit still, look toward the camera, and relax. The synchronized video and physiological recordings were taken for two minutes. For one of the sessions the contact finger PPG measurements were noisy due to motion artifacts; this session was not used for the comparison of the contact and remote methods. Although the camera method is also susceptible to motion artifacts this highlights some of the challenges associated with contact measurements, especially for tasks where people need to move their hands (e.g. typing).

**Measurements under cognitive stress:** In the second experiment participants were asked to perform a mental arithmetic test (MAT) silently. Starting with the number 4000 they were required to subtract 7 then subtract 7 again and so on, as quickly as possible. The synchronized video and physiological recording were taken for two minutes. The participants started the task immediately after the recordings were started. In order to increase the cognitive stress induced we told the participants that they were competing against the other people to reach the lowest number after two minutes. The participants consistently reported this task to be more stressful.

**D. Recovery of Physiology from Camera**

We propose a new fully-automated method for recovering the HRV spectrogram from the recorded videos by making alterations to the method presented in [16]. The alterations improve the resulting measurements and are described below. Figure 3 provides an overview of the method. The videos were exported in an uncompressed format. The physiological and video recordings were analyzed offline using custom software written in MATLAB (The Mathworks, Inc.).
thus, we assumed N underlying source signals, represented by \(s_1(t), s_2(t), \ldots, s_N(t)\). We use the JADE implementation of ICA [23]. Each of the source signals was band-pass filtered using a Hamming window filter with low- and high-frequency cut-offs at 45 beats-per-minute (bpm) (0.75Hz) and 180 bpm (3Hz) respectively. These cut-off frequencies reflect conservative lower and upper limits in heart rate.

ICA has two properties that make automated analysis challenging. Firstly, the source signals are returned in a random order and therefore it is not always the same source which has the strongest BVP waveform. Secondly, the source signals can be scaled arbitarily (and subsequently flipped if scaled by a number < 0). A flipped BVP signal is problematic when it comes to peak detection as the calculated inter-beat intervals (IBI) are typically much more accurate. The following steps are designed to find the optimal source signal and invert it if the BVP component within the source has been flipped.

The appropriate source signal was selected by calculating the normalized fast Fourier transform (FFT) of each source and choosing the source signal with the greatest frequency peak within the range 45 - 180 bpm (the same limits as the bandpass filter 3dB points above). The FFTs were normalized to give a total power across all frequencies equal to one. This is similar to the method used by Poh et al. [10]. We verified this approach by manually choosing the optimal source signal for the five band signal case and the FFT method agreed on all occasions. Figure 5a shows examples of source signals and the power spectra of each source. Clearly, the power spectrum with the highest peak corresponds to the signal with the strongest BVP; all other source signals have flatter spectra.

As mentioned above, when using ICA the BVP waveform may be inverted. In order to automatically predict whether the selected source signal has an inverted BVP waveform we calculate the mean absolute peak (\(\mu_{\text{peakamp}}\)) and trough (\(\mu_{\text{troughamp}}\)) amplitudes (the source signals are returned with zero mean). For an inverted BVP signal the mean trough amplitude is likely to be greater than the mean peak amplitude due to the shape of the BVP waveform. Therefore, if \(\mu_{\text{peakamp}} < \mu_{\text{troughamp}}\) the selected source would be inverted (multiplied by -1). Figure 5 shows examples of a non-inverted and an inverted BVP signal. Poh et al. [10] did not propose any technique for detecting inverted BVP signals. In Section IV we show that this addition improves the accuracy of the resulting physiological parameters.

The estimated BVP signal was interpolated with a cubic spline function at a sampling frequency of 256Hz. Peak detection was performed using a custom algorithm with a moving time window of length 0.25s. Within the moving window, if the signal maximum was greater than that in the previous window the next window would be considered. If the maximum within the window was less than that in the previous window then the previous maximum was selected as a peak and the process would repeat. We tested different window sizes and found that a window of 0.25s gave the results that were most closely correlated with the visually verified contact sensor measurements, results are shown in Table II. Figure 6 shows a 30 second comparison of the PPG waveforms from the contact sensor and the remote method for one individual.

The visually verified peaks are shown on the contact waveform and the automatically detected peaks are shown on the remote waveform. Notice how when using a window size of 0.35s the beats at 50.3s and 71s are missed but with a window size of 0.25s they are not.

To avoid artifacts (such as motion or ectopic beats) which can impact the HRV analysis, the resulting IBIs were filtered using the non causal of variable threshold (NC-VT) algorithm [24] with a tolerance of 30% (Poh et al. [10] found 30% tolerance to be effective). Finally, inter-beat intervals were filtered using a low pass filter with cut-off at 0.4Hz. This analysis we were interested in measuring the high (0.15-0.4Hz) and low frequency (0.04-0.15Hz) components of the HRV power spectra and therefore we filtered with a cut-off at 0.4Hz. We construct the HRV spectrograms by calculating the power spectral density from the IBIs for sequential windows. For each window the power spectral density (PSD) of the inter-beat intervals was calculated using the Lomb periodogram. In this analysis we use a moving window of one minute and the sessions were two minutes in length, the step size was one second. We chose these parameters as we wanted to have a large enough window to measure high and low frequency components of the IBIs (between 0.04-0.4Hz) accurately within each window but also to capture the temporal dynamics in

![Figure 5](image-url)
the spectrograms - there was a trade-off in the choice of the window size between frequency and temporal resolution.

E. Recovery of Physiology from Contact Sensors

To calculate the HRV spectrogram using the PPG measurements from the contact finger sensor we use the same parameters as above. PPG peak detection was performed with a moving time window of length 0.25 seconds. For comparison with the remote measurements the accuracy of the peak detection was then confirmed using visual inspection, any missed peaks or false detections were corrected. Spectrograms were computed by calculating the power spectral density from the IBIs for a moving window of length one minute, step size one second, over each two-minute session.

F. Quantification of Physiological Parameters

For a quantitative comparison between the contact sensor and the camera method, and also comparing to results in our previous work, we calculate the HR, BR and HRV low frequency and high frequency components. HR was calculated as $60/TBI$, where $TBI$ is the mean of the inter-beat intervals. BR can be estimated from the high frequency component of the HRV [25]. We determine the BR from the frequency peak ($f_{HFpeak}$), center frequency of the highest peak between 0.15 and 0.4Hz, of the HRV power spectrum. For the contact measurements the BR was calculated from the frequency of the dominant peak $f_{respeak}$ in the PSD of the recorded respiratory waveform as $60/f_{respeak}$. The low frequency (LF) and high frequency (HF) powers of the HRV were calculated as the area under the PSD curve corresponding to 0.04-0.15 and 0.15-0.4Hz respectively. These were quantified in normalized units in order to minimize the impact of a difference in total power. The LF component is modulated by baroreflex activity and contains both sympathetic and parasympathetic activity [26]. The HF component reflects parasympathetic influence on the heart and is connected to respiratory sinus arrhythmia (RSA). An estimate of sympathetic modulation (the sympatho/vagal balance) can be made by considering the ratio LF/HF.

IV. RESULTS

Using the analysis described in Section III we calculated physiological parameters from the BVP waveforms extracted from the camera signals and the contact sensor.

Comparison of Color Channels: In order to compare the results to those that would be recorded with most conventional digital cameras and to analyze the impact of each color band on the performance we repeated the analysis described in Section III using: 1) each signal band alone, 2) all pairs of bands, 3) all combinations of three bands, 4) all combinations of four bands and 5) all five band (RGBCO) channel signals. Table I shows the correlation between the camera measures and the contact finger measures for all the cases on the left side. For the comparisons peaks detected in the contact measures were confirmed visually. On the right side we show the color channel combinations ordered with respect to ascending mean HR, BR, LF, HF and LF/HF correlation ($r$). The GCO combination had the greatest mean correlation.
are aware of that shows the HRV spectrograms calculated from the two minute sessions using a one minute sliding window. The camera values were computed using the GCO channels. The best performing combination of channels was GCO, which was only outperformed by other combinations for the BR correlation. The orange band featured in the top ten combinations of channels: 1.0, 0.91, 0.97, 0.97, 0.95 respectively. The performance of the algorithm for the rest and stress conditions measurements for all the data. Here we compare the performance of the algorithm for the rest and stress conditions separately. In the rest condition the HR, BR, HRV LF, HF and LF/HF correlations with the contact sensor were: 1.0, 0.90, 0.87, 0.87, 0.86 respectively. In the stress condition the HR, BR, HRV LF, HF and LF/HF correlations with the contact sensor were: 1.0, 0.91, 0.97, 0.97, 0.95 respectively. The results show good performance in both conditions. Although the correlations were slightly lower in the rest condition perhaps due to the slower breathing rates which may be harder to estimate within the time window.

**V. DISCUSSION**

Our results in Table I show high correlation between the contact measures and the camera measures, despite the camera being placed 3m from the participant. The best performing combination of channels was GCO, which was only outperformed by other combinations for the BR correlation. The orange band featured in the top ten combinations of channels: This suggests it is capturing significant information. However, other observations are also needed to help boost the signal to noise ratio. The orange band is close to the green band and this supports previous work that showed strong measurement of the BVP in the green frequency range [9]. In order to improve remote PPG measurements using digital cameras these results suggest one should include color channel sensors closer to the orange, green and cyan frequencies.

Heart rate measurements were highly correlated across all channels. This is because the dominant frequency measurement is not highly susceptible to poor peak detection on the BVP. However, the correlations of other measures - which all rely on accurate BVP peak detection - varied greatly. The benefit of the additional color bands stands out as they allow for the recovery of a much cleaner BVP
that an orange color channel helps boost the performance of human face. Using a novel five band camera sensor we show HRV LF and HF components) from camera images of the response in the stress condition).

Eight of ten participants had a higher HRV LF/HF ratio during significantly higher (p < 0.005) than in the rest condition (0.51). Eight of ten participants had a higher HRV LF/HF ratio during the stress condition compared to the rest condition. Higher EDA response was also observed in eight of ten participants in the stress condition relative to the rest condition (seven participants had higher HRV LF/HF ratio and higher EDA response in the stress condition).

**VI. CONCLUSIONS AND FUTURE WORK**

We have presented physiological measurements (HR, BR, HRV LF and HF components) from camera images of the human face. Using a novel five band camera sensor we show that an orange color channel helps boost the performance of physiological measurement using a digital camera. A combination of cyan, green and orange color channels was the best performing. The RGB channel combination was one of the poorest performing. The GCO channel combination outperformed the RGB channels for all individuals showing that the best performance occurred across a range of skin tones and under varying ambient lighting conditions. We compared the camera measurements with those from traditional contact measurements. The agreement between the contact and camera measurements was very high. The measurements were made with a digital camera placed 3m from the face of the participant, a greater distance than in results presented previously.

We present the first examples of HRV spectrograms calculated from videos of the human face. Qualitative comparisons between the spectrograms measured using the camera and the contact sensors show close agreement.

There are certain limitations that should be noted when considering these results. In these experiments the participants were free to move; however, they were seated and did not turn away from the camera. A real-time system would need to address issues such as artifacts due to incorrect/missing face tracking results, rigid head motions and dramatic ambient light changes. We have tested the system against data from 10 individuals and demonstrated very strong performance. However, for certain applications - such as infant monitoring - testing would need to be performed on a representative population. Future work will consider measurements as people perform computer tasks and investigation of whether cognitive stress can be predicted from remotely measured changes in cardio-pulmonary activity. Comparisons between the measurements made by the camera and electro-cardiogram (ECG) measurements of HR and HRV components would also be a useful extension. In this work all the analysis was performed off-line. We leave a real-time implementation of the approach to future work.
ACKNOWLEDGMENT
This work was funded by the MIT Media Lab Member Consortium. Daniel McDuff was supported by an NEC fellowship.

REFERENCES