Improvements in Remote Cardiopulmonary Measurement Using a Five Band Digital Camera

Citation: McDuff, Daniel; Gontarek, Sarah and Picard, Rosalind W. "Improvements in Remote Cardiopulmonary Measurement Using a Five Band Digital Camera." IEEE Transactions on Biomedical Engineering 61, no. 10 (October 2014): 2593–2601. © 2014 Institute of Electrical and Electronics Engineers (IEEE)

As Published: http://dx.doi.org/10.1109/TBME.2014.2323695

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: http://hdl.handle.net/1721.1/109145

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

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 ($\sim$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

D. J. McDuff is with the Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. (e-mail: djmcduff@mit.edu).

S. Gontarek is. is with the Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: sgontare@mit.edu).

R. W. Picard is a professor at the Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. (phone: 617-253-0611; e-mail: picard@media.mit.edu).

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 digital 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].

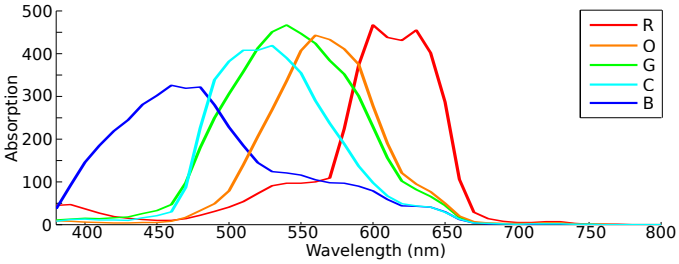

Fig. 1. Five band camera light sensitivity profile. In addition to red, green and blue light sensors this camera can measure orange and cyan bands.

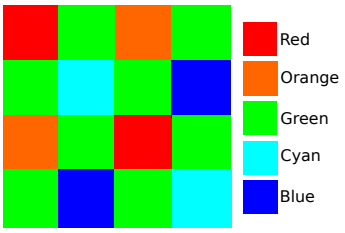

Fig. 2. Five band camera sensor layout. The image shows the arrangement of the colors in a 4x4 pattern that repeats across the sensor. Each pixel on the sensor measures one color as determined by its position.

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

$$m = \sum_{\lambda} e(\lambda)s(\lambda) \quad (1)$$

Where e(λ) is the energy of light at a given wavelength λ and s(λ) 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

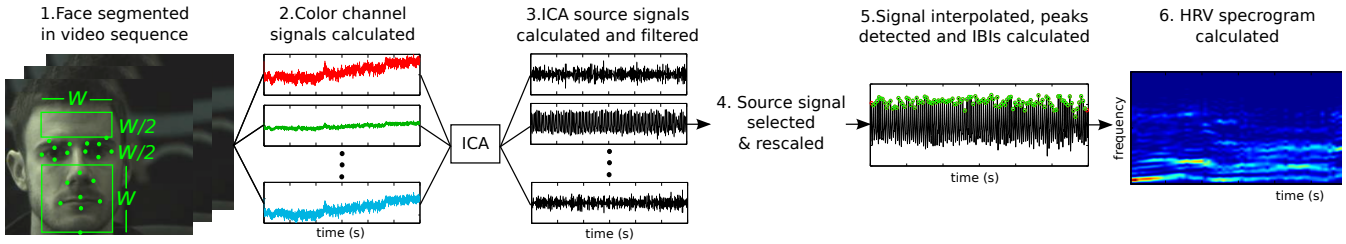

Fig. 3. Overview of the automated method used to recover the HRV spectrogram from videos of a human face. 1) Facial landmarks are detected using [20] and the face region of interest (ROI) segmented (excluding the region around the eyes), 2) spatial averages of each color channel in the ROI over time are calculated, 3) source signals, calculated via ICA, are filtered, 4) the channel with the estimated strongest BVP signal are selected and inverted if necessary, 5) BVP signal is interpolated to 256Hz, peaks detected, and IBIs calculated, 6) spectrogram are calculated with a moving window (window length 60s, step size 1s).

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.).

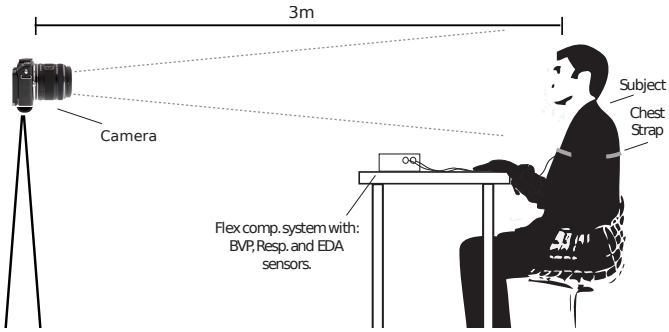

Fig. 4. Experimental set-up. Contact measurements of the blood volume pulse and electrodermal activity were collected using finger sensors and respiration was measured using a chest strap. A camera, placed 3m from the participant, was used to capture videos images at 30fps, 960x720 resolution.

The LEAR [20] facial landmark detector was used to find the x- and y- coordinates of facial landmarks on the participant's face in each frame of the video. As shown in Figure 3 step 1 we selected the full width between the outer eye corners (w) and a height twice the width (w above the eye corners to w below the eye corners) as a box encompassing the region of interest (ROI). We excluded a section around the eyes (the full width of the box and 25% of the height around the eye corners) to remove motion artifacts due to blinking and eye movements. This adjustment improved correlations with the contact sensor measurements. The mean number of pixels within the ROI was 125,000 pixels (st. dev. = 19,100 pixels). This represents less than 25% of the frame. The minimum facial ROI used across all videos was 95,000 pixels.

A spatial average of the color channel (red, green, blue, orange and cyan or a subset thereof) pixel values within the resulting ROI were calculated for each frame to form the raw signals $x_1(t)$, $x_2(t)$, ..., $x_N(t)$ respectively, where N is the number of color channels (between one and five). The raw traces were detrended using a technique based on a smoothness priors approach [22]. The smoothness parameter, λ, was set to 2000. This allowed only very low-frequency components of the signal to be removed, not damaging the high frequency information. The resulting signals were normalized by subtracting the mean and dividing by the standard deviation. We then apply ICA to recover source signals from the observations, maximizing the non-Gaussianity within the sources. Using a conventional ICA algorithm the number of recoverable sources cannot exceed the number of observations;

thus, we assumed N underlying source signals, represented by $s_1(t)$, $s_2(t)$, ..., $s_N(t)$. 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 arbitrarily (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 less 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 (μ_{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.

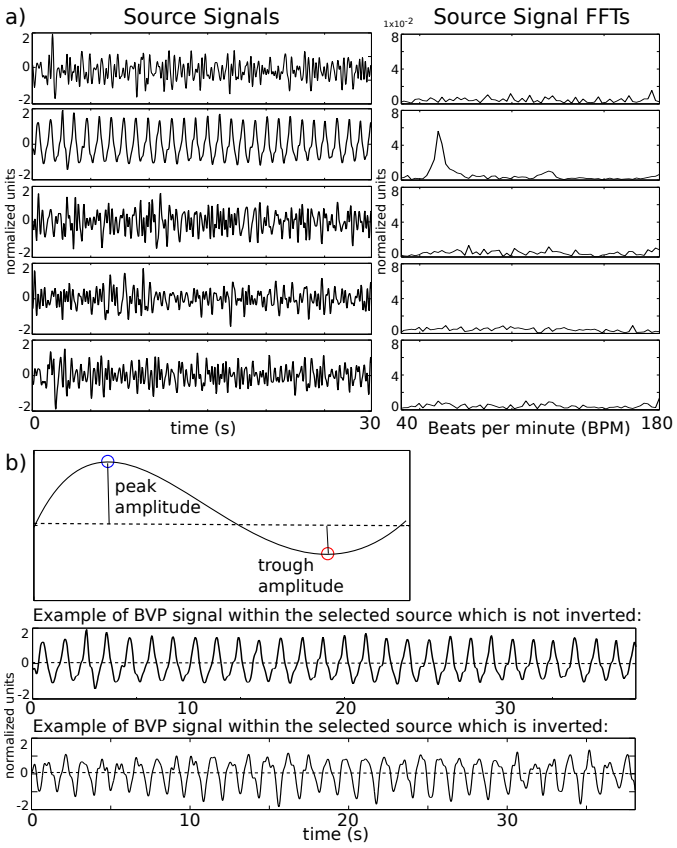

Fig. 5. a) ICA returns the source signals in a random order. Source signal selection is solved by choosing the signal with peak of greatest power in the normalized FFT spectrum (between 40 and 180 bpm). In the above example this would mean selecting the second source signal. b) ICA returns source signals with arbitrary scaling. Problems due to inverted signals are solved by finding the mean absolute peak and trough heights and inverting the signal if $\mu_{\text{peakamp}} < \mu_{\text{troughamp}}$.

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 frequency 0.4Hz. In 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

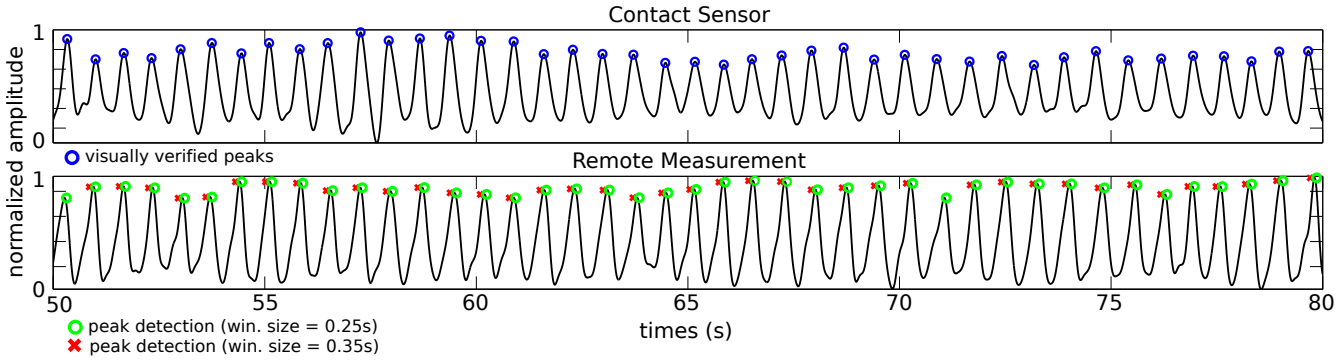

Fig. 6. Comparison of 30 second PPG waveforms from the contact sensor and the remote method. Top) Contact sensor waveform with visually verified peaks (blue). Bottom) Remote measurement waveform with peaks located using a time window of 0.25s (green) and 0.35s (red) (peak locations offset for comparison). Notice the beats at 50.3s and 71s are missed by the peak detection algorithm with window size 0.35s but not with window size 0.25s.

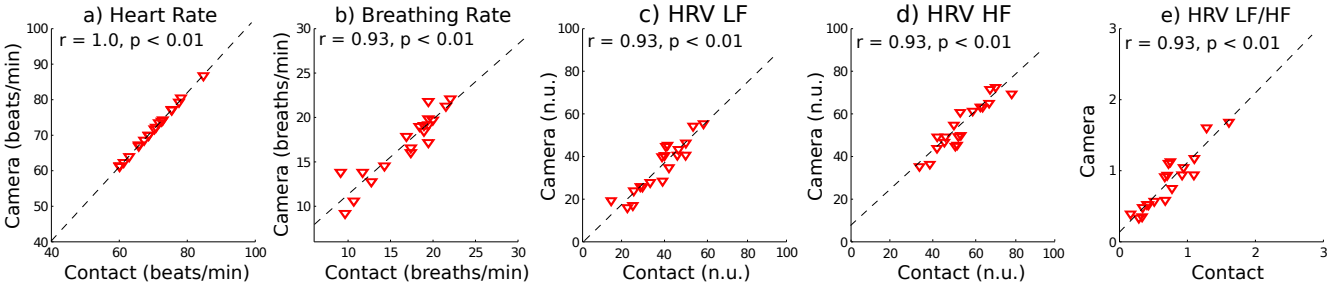

Fig. 7. Scatter plots comparing the measurements of (a) heart rate (HR), (b) breathing rate (BR), (c) HRV LF component, (d) HRV HF component, (e) HRV LF/HF ratio for the contact sensor (finger BVP for HR and HRV and chest strap for BR) and the non-contact method described in Section III using the GCO band observations. High agreement was observed across all measures. In all cases $p < 0.01$. n.u. = normalized units. Comparisons made across 19 sessions.

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 $60/\overline{IBI}$, where $\overline{IBI}$ 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_{resppeak} in the PSD of the recorded respiratory waveform as $60/f_{\text{resppeak}}$. 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 measurements 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 ($\overline{r}$). The GCO combination had the greatest mean correlation

TABLE I
COMPARISON OF THE CORRELATIONS BETWEEN THE CONTACT FINGER SENSOR MEASUREMENTS AND CAMERA MEASUREMENTS FOR ALL COMBINATIONS OF THE CAMERA COLOR CHANNEL SIGNALS. FOR ALL CORRELATIONS $p < 0.01$. ON THE RIGHT ARE THE CHANNEL COMBINATIONS ORDERED FROM LOWEST MEAN CORRELATION TO HIGHEST MEAN CORRELATION. THE GCO CHANNEL COMBINATION PERFORMED BEST. COMPARISONS MADE ACROSS 19 SESSIONS.

	HR	BR	LF	HF	LF/HF
R	0.99	0.95	0.60	0.60	0.57
G	0.99	0.91	0.63	0.63	0.63
B	0.99	0.93	0.68	0.68	0.70
C	0.85	0.44	0.64	0.64	0.64
O	0.83	-0.02	0.43	0.43	0.34
RG	0.97	0.66	0.72	0.72	0.74
RB	0.95	0.89	0.47	0.47	0.47
RC	0.99	0.67	0.69	0.69	0.73
RO	1.00	0.93	0.88	0.88	0.89
GB	0.89	0.75	0.44	0.44	0.44
GC	0.99	0.83	0.82	0.82	0.82
GO	1.00	0.98	0.88	0.88	0.88
BC	0.99	0.68	0.61	0.61	0.65
BO	1.00	0.92	0.87	0.87	0.87
CO	0.99	0.67	0.40	0.40	0.48
RGB	0.85	0.67	0.45	0.45	0.46
RGC	0.99	0.75	0.67	0.67	0.71
RGO	1.00	0.92	0.83	0.83	0.86
RBC	0.99	0.69	0.71	0.71	0.68
RBO	1.00	0.92	0.83	0.83	0.83
RCO	1.00	0.90	0.91	0.91	0.89
GBC	0.99	0.77	0.80	0.80	0.78
GBO	1.00	0.93	0.84	0.84	0.83
GCO	1.00	0.93	0.93	0.93	0.93
BCO	0.99	0.84	0.69	0.69	0.77
RGBC	0.99	0.89	0.72	0.72	0.68
RGBO	1.00	0.81	0.79	0.79	0.81
RGCO	1.00	0.90	0.87	0.87	0.86
RBCO	1.00	0.90	0.81	0.81	0.77
GBCO	1.00	0.72	0.83	0.83	0.80
RGBCO	1.00	0.74	0.81	0.81	0.79

Lowest $\overline{r}$
O
RGB
CO
GB
C
RB
BC
R
RC
RBC
G
RGC
RG
BCO
B
RGBC
GBC
RGBCO
GBCO
RGBO
GC
RBCO
RBO
GBO
RGO
RGCO
BO
RO
RCO
GO
GCO
Highest $\overline{r}$

TABLE II
COMPARISON OF THE CORRELATIONS BETWEEN THE CONTACT FINGER SENSOR MEASUREMENTS AND CAMERA MEASUREMENTS FOR THE GCO CHANNEL COMBINATION - WITH DIFFERENT PEAK DETECTION WINDOW SIZES. WIN. SIZE: 0.25S SHOWS THE HIGHEST CORRELATION.

Win. Size	HR	BR	LF	HF	LF/HF
0.15s	0.95	0.92	0.68	0.68	0.61
0.20s	1.00	0.88	0.94	0.94	0.92
0.25s	1.00	0.93	0.93	0.93	0.93
0.30s	1.00	0.83	0.88	0.88	0.86
0.35s	0.96	0.63	0.83	0.83	0.83

across all measures. Interestingly, the O band alone had the lowest mean correlation across all measures, but is present in all of the 10 top-performing combinations.

Figure 7 shows the correlations between the non-contact and contact measurements for the a) HR, b) Breathing Rate (BR), c) HRV LF, d) HRV HF and e) HRV LF/HF measurements. The camera values were computed using the GCO channels. These results show high agreement between the remote method and the finger PPG measures and are improved over those presented by Poh et al. [10] despite the camera being much further from the subject here (3m compared to 0.5m).

HRV Spectrograms: We calculated the HRV spectrograms from the two minute sessions using a one minute sliding window with one second increments. This is the first example we are aware of that shows the HRV spectrograms calculated from non-contact video sequences. Figure 8 shows a comparison of spectrograms recovered from the three band (RGB) and three band (GCO) recordings next to those from the contact finger measurements. On the left are examples from sessions in which the participants were at rest and on the right are examples in which the participants were under cognitive stress.

TABLE III
COMPARISON OF THE CORRELATIONS BETWEEN THE CONTACT FINGER SENSOR MEASUREMENTS AND CAMERA MEASUREMENTS FOR THE GCO CHANNEL COMBINATION - WITH AND WITHOUT INVERSION CORRECTION OF THE SELECTED SOURCE SIGNAL AFTER ICA. FOR ALL CORRELATIONS $p < 0.01$.

	HR	BR	LF	HF	LF/HF
without inversion correction	1.00	0.90	0.86	0.86	0.88
with inversion correction	1.00	0.93	0.93	0.93	0.93

Impact of Inversion Correction Step: Table III shows the impact of inverting the source signal output from ICA. Without this step some of the source signals have an inverted BVP and in these cases calculation of the IBIs is problematic. Our new method of detecting and correcting for inversion gives a closer estimation of the HRV spectra and subsequent parameters when compared to the contact measures.

Performance in Rest and Stress Conditions: Figure 7 shows the correlations between the non-contact and contact 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 almost all of the 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

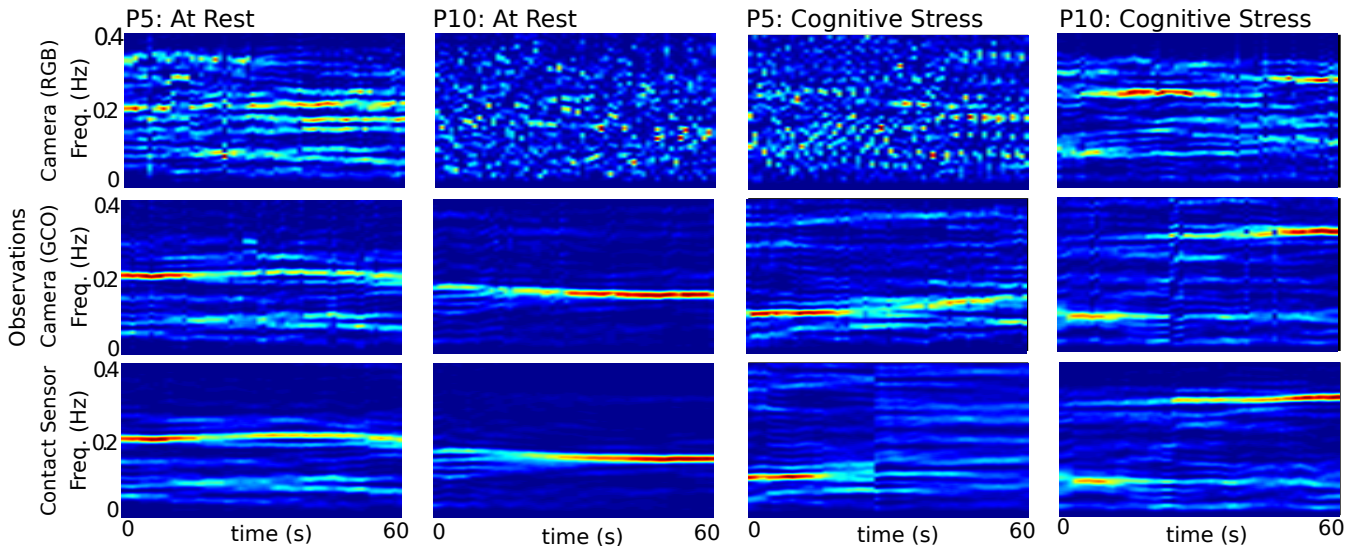

Fig. 8. Heart rate variability spectrograms calculated using: Top) RGB camera signals, middle) GCO camera signals and bottom) contact finger sensor. Qualitatively the measurements from the GCO channels more closely match those from the contact sensor (reinforcing the quantitative comparisons in Table I). On the left are sessions in which participants 5 and 10 were at rest and on the right are sessions in which participants 5 and 10 were under cognitive stress. There are stronger low frequency components in the latter as we would expect less parasympathetic nervous system activity. The measurements for participant 5, under cognitive stress, made using the camera suggest that the camera (using GCO bands) may have been more accurate than the finger sensors.

signal and therefore more accurate source signal selection, scaling and peak detection.

Interestingly, the RGB combination of color channels, used in today's standard digital cameras, was one of the worst performing combinations of channels. The measurements using the GCO combination of color channels show high agreement between the remote method and the finger PPG measures and are greater than those presented by Poh et al. [10] despite the camera being much further from the subject (3m compared to 0.5m) and our inclusion of variable lighting, different colors of skin and subjects wearing glasses. The spectrogram calculated from the five band observations is closer to that of the finger measurements and in example P5 of Fig. 8 actually seems to be more accurate than the finger measurements - perhaps due to motion artifacts as a result of the fingers moving. However, we cannot be certain that this is the case - we will investigate the impact of motion more in future work.

We can see from the spectrograms that there is greater low frequency power in the HRV spectra for those individuals under cognitive stress, this is what we would expect due to less parasympathetic activity. Across all participants the mean HRV LF/HF ratio in the stress condition (0.81) was 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

[1] L. A. Aarts, V. Jeanne, J. P. Cleary, C. Lieber, J. S. Nelson, S. Bambang Oetomo, and W. Verkruysse, "Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unita pilot study," *Early human development*, vol. 89, no. 12, pp. 943–948, 2013.

[2] R. E. Kleiger, J. P. Miller, J. T. Bigger Jr, and A. J. Moss, "Decreased heart rate variability and its association with increased mortality after acute myocardial infarction," *The American journal of cardiology*, vol. 59, no. 4, pp. 256–262, 1987.

[3] J. Delaney and D. Brodie, "Effects of short-term psychological stress on the time and frequency domains of heart-rate variability," *Perceptual and motor skills*, vol. 91, no. 2, pp. 515–524, 2000.

[4] M. V. Kamath, D. N. Ghista, E. L. Fallen, D. Fitchett, D. Miller, and R. McKelvie, "Heart rate variability power spectrogram as a potential noninvasive signature of cardiac regulatory system response, mechanisms, and disorders," *Heart and vessels*, vol. 3, no. 1, pp. 33–41, 1987.

[5] P. Jönsson, "Respiratory sinus arrhythmia as a function of state anxiety in healthy individuals," *International journal of psychophysiology*, vol. 63, no. 1, pp. 48–54, 2007.

[6] C. C. Yang, C.-W. Lai, H. Y. Lai, and T. B. Kuo, "Relationship between electroencephalogram slow-wave magnitude and heart rate variability during sleep in humans," *Neuroscience letters*, vol. 329, no. 2, pp. 213–216, 2002.

[7] A. Moriguchi, A. Otsuka, K. Kohara, H. Mikami, K. Katahira, T. Tsunetoshi, K. Higashimori, M. Ohishi, Y. Yo, and T. Ogihara, "Spectral change in heart rate variability in response to mental arithmetic before and after the beta-adrenoceptor blocker, carteolol," *Clinical Autonomic Research*, vol. 2, no. 4, pp. 267–270, 1992.

[8] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, vol. 28, no. 3, p. R1, 2007.

[9] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics express*, vol. 16, no. 26, pp. 21 434–21 445, 2008.

[10] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 1, pp. 7–11, 2011.

[11] S. S. Ulyanov and V. V. Tuchin, "Pulse-wave monitoring by means of focused laser beams scattered by skin surface and membranes," in *OE/LASE'93: Optics, Electro-Optics, & Laser Applications in Science& Engineering*. International Society for Optics and Photonics, 1993, pp. 160–167.

[12] E. Greneker, "Radar sensing of heartbeat and respiration at a distance with applications of the technology," in *Radar 97 (Conf. Publ. No. 449)*. IET, 1997, pp. 150–154.

[13] S. Bakhtiari, T. W. Elmer, N. M. Cox, N. Gopalsami, A. C. Raptis, S. Liao, I. Mikhelson, and A. V. Sahakian, "Compact millimeter-wave sensor for remote monitoring of vital signs," *Instrumentation and Measurement, IEEE Transactions on*, vol. 61, no. 3, pp. 830–841, 2012.

[14] M. Garbey, N. Sun, A. Merla, and I. Pavlidis, "Contact-free measurement of cardiac pulse based on the analysis of thermal imagery," *Biomedical Engineering, IEEE Transactions on*, vol. 54, no. 8, pp. 1418–1426, 2007.

[15] J. Fei and I. Pavlidis, "Thermistor at a distance: unobtrusive measurement of breathing," *Biomedical Engineering, IEEE Transactions on*, vol. 57, no. 4, pp. 988–998, 2010.

[16] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics Express*, vol. 18, no. 10, pp. 10 762–10 774, 2010.

[17] Y. Sun, S. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers, and Y. Zhu, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise," *Journal of Biomedical Optics*, vol. 16, no. 7, pp. 077 010–077 010, 2011.

[18] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 4, p. 65, 2012.

[19] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3430–3437.

[20] B. Martinez, M. F. Valstar, X. Binefa, and M. Pantic, "Local evidence aggregation for regression-based facial point detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 5, pp. 1149–1163, 2013.

[21] Y. Monno, M. Tanaka, and M. Okutomi, "Multispectral demosaicking using guided filter," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2012, pp. 82 990O–82 990O.

[22] M. P. Tarvainen, P. O. Ranta-aho, and P. A. Karjalainen, "An advanced detrending method with application to hrv analysis," *Biomedical Engineering, IEEE Transactions on*, vol. 49, no. 2, pp. 172–175, 2002.

[23] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," in *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 6. IET, 1993, pp. 362–370.

[24] J. Vila, F. Palacios, J. Presedo, M. Fernández-Delgado, P. Felix, and S. Barro, "Time-frequency analysis of heart-rate variability," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 16, no. 5, pp. 119–126, 1997.

[25] T. E. Brown, L. A. Beightol, J. Koh, and D. L. Eckberg, "Important influence of respiration on human rr interval power spectra is largely ignored," *Journal of Applied Physiology*, vol. 75, no. 5, pp. 2310–2317, 1993.

[26] S. Akselrod, D. Gordon, F. A. Ubel, D. C. Shannon, A. Berger, and R. J. Cohen, "Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control," *science*, vol. 213, no. 4504, pp. 220–222, 1981.

Daniel McDuff received the bachelor's degree, with first-class honors and master's degree in engineering from Cambridge University. He is a PhD candidate in the Affective Computing Group at the MIT Media Lab. He is interested in computer vision and machine learning to enable the automated recognition of affect. He is also interested in technology for remote measurement of physiology. He is a student member of the IEEE.

Sarah Gontarek is an electrical engineering and computer science student at MIT, and a political science student at Wellesley College. She is interested in signal processing, neuroscience, and machine learning. She enjoys work in the Affective Computing Group in the Media Lab, because she is especially interested in biomedical applications for common digital communications technology.

Rosalind W. Picard received the ScD degree in electrical engineering and computer science from MIT. She is a professor of Media Arts and Sciences at the MIT Media Lab, founder and director of the Affective Computing Group at the MIT Media Lab. She is also a co-founder of Affectiva, Inc. Her current research interests focus on the development of technology to help people comfortably and respectfully measure and communicate affective information, as well as on the development of models of affect that improve decision-making and learning. She is a fellow of the IEEE and member of the IEEE Computer Society.