Full-Body Cardiovascular Sensing with Remote Photoplethysmography

Lu Niu, Jeremy Speth, Nathan Vance, Ben Sporrer, Adam Czajka, Patrick Flynn
University of Notre Dame
{lniu,jspeth,nvance1,bsporrer,aczajka,flynn}@nd.edu

Abstract

Remote photoplethysmography (rPPG) allows for non-contact monitoring of blood volume changes from a camera by detecting minor fluctuations in reflected light. Prior applications of rPPG focused on face videos. In this paper we explored the feasibility of rPPG from non-face body regions such as the arms, legs, and hands. We collected a new dataset titled Multi-Site Physiological Monitoring (MSPM), which will be released with this paper. The dataset consists of 90 frames per second video of exposed arms, legs, and face, along with 10 synchronized PPG recordings. We performed baseline heart rate estimation experiments from non-face regions with several state-of-the-art rPPG approaches, including chrominance-based (CHROM), plane-orthogonal-to-skin (POS) and RemotePulseNet (RPNet). To our knowledge, this is the first evaluation of the fidelity of rPPG signals simultaneously obtained from multiple regions of a human body. Our experiments showed that skin pixels from arms, legs, and hands are all potential sources of the blood volume pulse. The best-performing approach, POS, achieved a mean absolute error peaking at 7.11 beats per minute from non-facial body parts compared to 1.38 beats per minute from the face. Additionally, we performed experiments on pulse transit time (PTT) from both the contact PPG and rPPG signals. We found that remote PTT is possible with moderately high frame rate video when distal locations on the body are visible. These findings and the supporting dataset should facilitate new research on non-face rPPG and monitoring blood flow dynamics over the whole body with a camera.

1. Introduction

Heart rate is one of the most important vital signs. Photoplethysmography (PPG) [1] is an optical technique to detect light changes caused by blood flow to noninvasively estimate heart rate. Traditional PPG employs a sensor that transmits and measures reflected light of specific frequencies on vascularized skin of the fingertip, ears, or forehead. These sensors are affixed to the body, which can be incon-

venient and can present risks to neonates, the elderly, and patients with damaged skin [9, 34, 37, 56]. These concerns, plus increased interest in noncontact vital sign measurements, have yielded rapid progress in remote photoplethysmography (rPPG), which can estimate blood flow changes at a distance from a camera, including exotic applications such as measurement from UAV-mounted cameras [13].

The current major databases conducive to rPPG are mainly focused on facial skin pixels. MAHNOB-HCI [40] and PURE [44] contain face videos with limited or controlled head movements. MMSE-HR [59] collected face videos after inducing emotional responses. UBFC-rPPG [5] contains occasional movements and subjects under stress from mathematical games. DDPM [42, 47] was the first long-form dataset containing unconstrained facial movements. Synthetic rPPG datasets also only contain avatars' faces [27, 51]. As rPPG becomes ubiquitous for health monitoring, it is important to understand the limitations presented by partial or total occlusion of the face, as well as the corresponding opportunities when skin regions *not* on the face are visible.

This paper explores multi-site rPPG, where the blood volume pulse is simultaneously estimated from the face, arms, hands, and legs. From a physiological perspective, collecting rPPG signals from different skin regions could provide a richer map of blood transport through the skin's capillary networks, rather than a scalar measurement of periodic blood volume changes at a single location. Full-body rPPG also allows for noncontact measurements of phase offsets between different sites, also known as the pulse transit time (PTT) [1, 29, 39], which is correlated with high-level biomarkers such as pulse wave velocity [14, 24], blood pressure [4, 7, 15–17, 29], arterial stiffness [26, 58], and peripheral arterial disease [3]. Figure 1 displays the experimental process presented in this paper. Our main contributions are summarized as follows:

(a) The newly-created Multi-Site Physiological Monitoring (MSPM) dataset collected from eighty-seven subjects. This dataset is, to our knowledge, the first dataset that allows simultaneous rPPG estimation from multiple body sites including arms, legs, and hands in
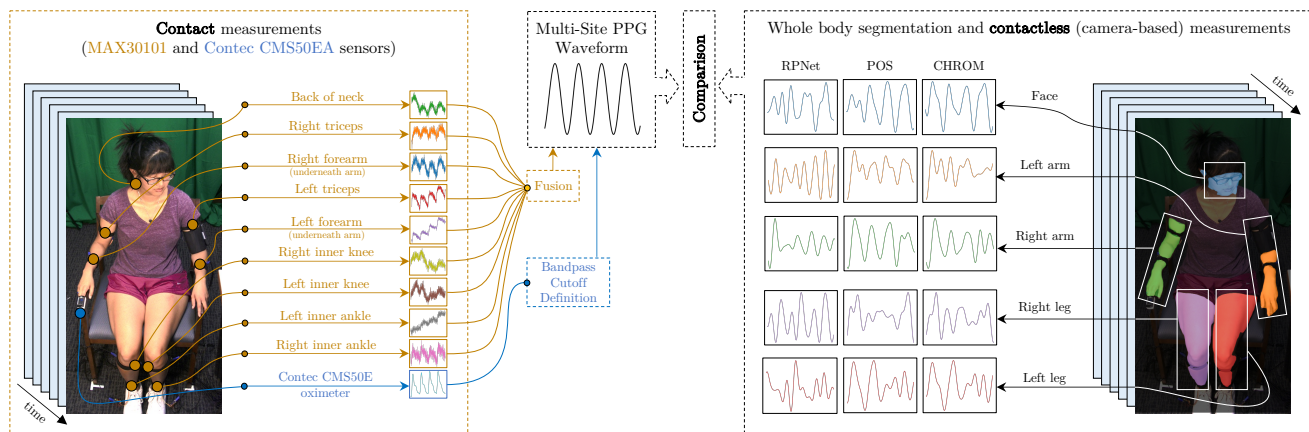
Figure 1. This paper explored the (a) Multi-Site Physiological Monitoring (MSPM) dataset with **contact** measurements from ten PPG sensors, including a fingertip oximeter, as shown on the left, with (b) **contactless** (camera-based) measurements of multiple skin areas carried out by three state-of-the-art rPPG methods (CHROM [9], POS [49], and RPNet [41]) as shown on the right. In this paper we also compared de-noised and combined PPG signals with the corresponding rPPG estimates.

addition to the face (see Sec. 3). A copy of the dataset can be requested at https://cvrl.nd.edu/projects/data.

(b) A comprehensive evaluation of baseline rPPG methods applied to video from multiple skin regions of the body. We used state-of-the-art rPPG methods including CHROM [9], POS [49], and RPNet [41] to retrieve signals from each body part, as well as sub-regions of each body part, to explore the overall performance and motivate future work (see Sec. 4, 6 and 5).

(c) A pulse transit time (PTT) analysis from both contact PPG and rPPG measurements at different sites on the body. We showed that pulse transit time can be estimated from a camera with lower frame rate than previous studies [15, 17] if the measurement sites on the body are further apart (see Sec. 7).

2. Related Work

2.1. Camera-based Measurements

rPPG methods estimate pulse rate based on skin pixel color changes from blood volume fluctuations. Several rPPG algorithms have been proposed based on different mechanisms, including Blind source separation (BSS) [52] and independent component analysis (ICA) [32, 33], which estimates pulse rate by applying different criteria to separate temporal RGB traces into uncorrelated or independent signal sources [49]. CHROM [9] assumes a standardized skin tone under white light and linearly combines the color signals for heart rate estimation. Spatial subspace rotation (2SR) [50] utilizes both spatial subspace and temporal rotation angle to calculate heart rate. POS [49] applies a plane orthogonal to the skin tone in the temporally normalized RGB space for pulse extraction.

More recently, deep learning models have been developed as the size of rPPG databases has increased. The first deep learning method [12] used a support vector regression model on both ICA and chrominance features. Convolutional neural networks (CNN) have successfully been applied to frame differences with several adaptations [8, 21, 22]. End-to-end waveform estimation from cropped videos passed through 3D-CNNs was first presented with Phys-Net [54] and later improved with internal dilated convolutions to improve temporal context [41]. Several unsupervised approaches have been introduced to reduce the need for simultaneous PPG ground truth. Most approaches leverage contrastive training strategies, where similar pairs of samples are pulled closer and different samples are repelled [10, 45, 48, 53]. The first non-contrastive approach leverages strong periodic priors to encourage the model to predict sparse signals in the frequency domain [43].

2.2. Multiple Contact Measurements

Measuring contact PPG from multiple locations simultaneously has garnered research interest primarily for blood pressure estimation from features related to pulse transit time (PTT) [1, 4, 25, 29, 30, 35]. Unfortunately, such systems are cumbersome, since multiple sensors must be properly attached and synchronized. It is therefore desirable to gather the same physiological information from a single sensor without contact, thus providing additional motivation for this work. Another advantage is that the number of potential transit time differences is drastically increased when using a camera sensor over contact PPG.
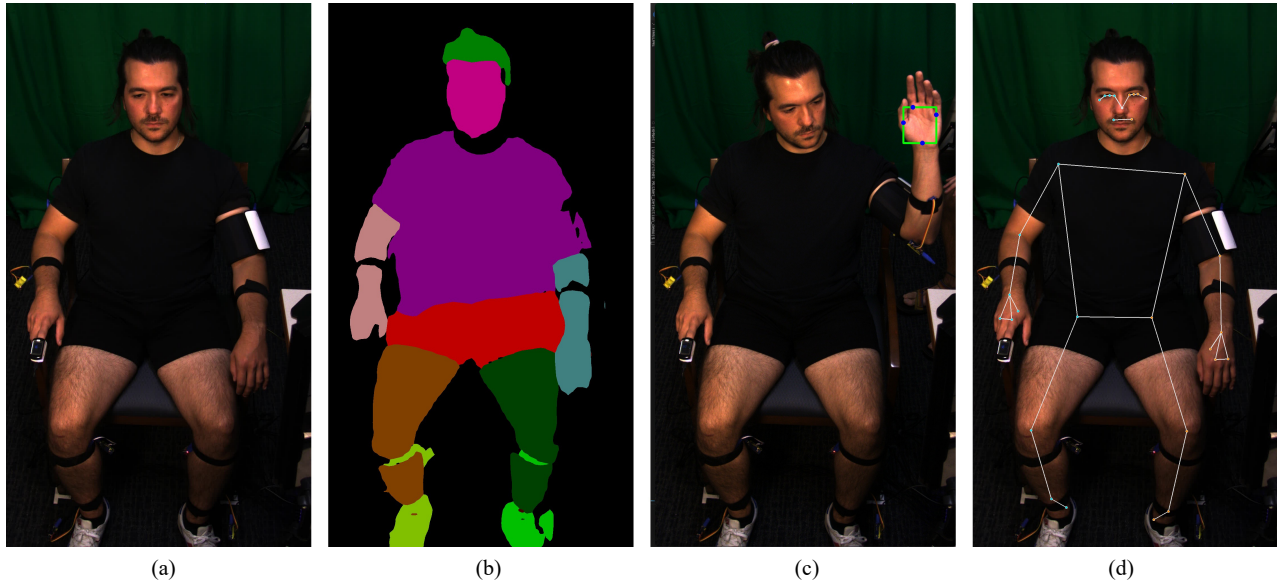
Figure 2. The subjects remained seated while collecting data (a). Using a SCHP [19] model trained on the LIP [20] dataset, the human body was segmented into 20 classes (b). rPPG Signals were extracted separately from 5 classes including 'Face', 'Left-arm', 'Right-arm', 'Left-leg', 'Right-leg'. We used MediaPipe Hand detection [57] to detect 4 points (shown in blue) of palm and use them to form a bounding box (shown in green) (c). MediaPipe Pose detection method [2] was applied and thirty-three keypoints of human body were generated (d).

2.3. Non-Facial rPPG

Most of the existing research utilizes the diffuse reflection from the face region to estimate the rPPG signal, with the exception of a few works analyzing the hand [6, 18, 36, 38, 46], arms [55], and thigh [11]. A gap in current rPPG research is the capture and estimation at multiple parts of the body simultaneously. Understanding rPPG signal quality across the entire body could allow for fewer constraints on the user and create new biomarkers as blood volume dynamics are measured in the peripheral vasculature.

3. Multi-Site Physiological Monitoring Dataset

We collected a large dataset from eighty-seven subjects seated in front of a video camera with several attached contact PPG sensors. Several characteristics were collected for each subject including sex (50 females, 35 males, and 2 non-binary subjects), age (24.6 ± 8.4 years), race, height (168.5 ± 10.3 cm), weight (69.6 ± 14.9 kg), smoking history, makeup worn, and recent caffeine consumption. Subjects were instructed to avoid wearing clothes that obstructed the arms and legs. A sample video frame is shown in Fig. 1, where the subject's face, arms, and legs are visible simultaneously. To the authors' knowledge, this is the first dataset that allows for simultaneous ground truthed rPPG at more than two sites on the body.

The collection began with the subject holding their left hand upright facing the camera for 90 seconds. This allows

for evaluation of rPPG on the palm, a region with glabrous skin that is valuable for remote pulse estimation [6]. For the remainder of the session, subjects underwent a blood pressure measurement, guided breathing exercise, and relaxation period. On average, the sessions evaluated in this paper last for approximately 5.7 minutes.

3.1. Apparatus

Video of the full subject was recorded with a DFK 33UX290 RGB camera from The Imaging Source (TIS). Raw video was recorded at 90 FPS with a resolution of 1920×1080 pixels. To record blood oxygenation (SpO2) and pulse rate, we collected 60 samples per second from an FDA-certified Contec CMS50EA pulse oximeter attached to an index finger. To support research in pulse transit time, we attached nine MAX30101 contact PPG sensors that recorded red and near-infrared signals at 400 samples per second. Each leg and arm had two attached sensors via elastic straps, and the last sensor was attached to the back of the neck with medical tape. All sensors and the camera recorded raw data to a single collection machine with the associated timestamps for easy synchronization.

3.2. Data Preprocessing

To properly evaluate rPPG approaches, we created a global pulse rate estimate from the multiple contact sensors. Contact PPG signals recorded from the nine sites with the MAX30101 sensors contain varying levels of noise
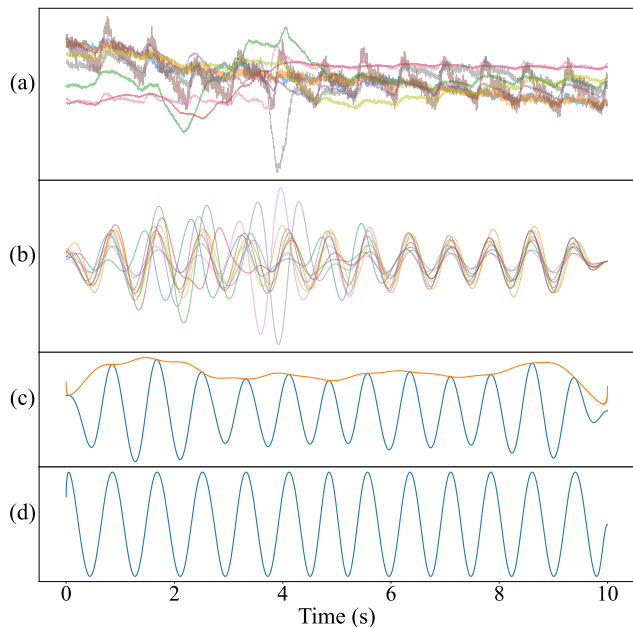
Figure 3. Processing of the contact PPG waveforms to produce a robust pulse rate. (a) The z-normalized waveforms from all sensors. (b) Bandpass filtered signals around the fingertip oximeter's pulse rate ± 30 bpm. (c) The signals were added together and the envelope is calculated. (d) The combined signal was divided by the envelope.

throughout the session depending on body movements. However, the pulse signal is likely present in at least one signal at any given point in time, since movements may be isolated to local regions.

Using this assumption, we combined the multiple pulse signals using a sliding window approach and bandpass filtering. We relied on the FDA-certified CMS50EA oximeter's pulse rate estimate to define the bounds of a narrow bandpass filter for the MAX30101 signals. Given the estimated pulse rate from the fingertip oximeter Y, we specified a padding around this value, $\Delta Y = 30$ bpm, and filtered the MAX30101 signals with a 2nd order Butterworth filter with lower and upper cutoffs of $Y - \Delta Y$ and $Y + \Delta Y$, respectively.

Specifically, for a sliding 10-second window with a stride of a single sample at 400 Hz, the waveforms underwent z-normalization, followed by filtering around the fingertip oximeter's pulse rate, then they were summed together into the combined waveform for that window. Finally, we divided the complete combined waveform by its envelope, which was calculated from the Hilbert transform. The process for signal combination is shown in Fig. 3. In the figure, between seconds 2 and 5, there appears to be noise from motion in multiple sensors, but the combined signal in subplot (d) remains clean.

4. Multi-Site rPPG Approach

Our pipeline to perform remote pulse estimation on our MSPM dataset consists of preprocessing, signal extraction and pulse rate estimation. We applied three pulse estimation algorithms: a chrominace-based method using color difference signals (CHROM) [9], plane-orthogonal-to-skin tone (POS) [49], and a 3D-CNN deep learning method (RP-Net) [41]. These three methods perform well for both stationary and moving subjects, which are characteristic of our MSPM dataset. The RPNet model was trained on face videos from the large and challenging deception detection and physiological monitoring dataset (DDPM) [42,47]. The model was not fine-tuned on any data from MSPM, so we tested how well the model could transfer to new subjects, lighting, standoff, and movement. Since the model was only trained on cropped faces, we also investigated its ability to transfer to spatially dissimilar regions such as the arms and legs. One of the benefits of CHROM and POS is that they do not have any spatial priors and only expect RGB traces from skin pixels.

4.1. Preprocessing

We selected skin pixels from the face, arms, and legs as the five ROIs. We applied Self-Correction Human Parsing (SCHP) [19] for body segmentation to the eighty-seven subjects of MSPM. Frame (b) of Fig. 2 shows masks of the twenty parts from SCHP for a single frame. Masked skin pixels of each ROI were averaged and supplied to CHROM and POS for signal extractions. For each masked ROI, we also calculated the minimum and maximum values of (x, y) locations of each mask to form bounding boxes for RPNet. Similarly, we created bounding boxes for palms using four key points of the palm generated by MediaPipe Hand detection [57]. Frame (c) of Fig. 2 shows the bounding box for the palm.

We applied MediaPipe Pose detection [2] to detect key points of the human skeleton, which were used to align the body for all subjects in our local error estimation experiments (described later in Sec. 6). There are thirty-three keypoints which define locations of many joints, including detailed locations for hands. Frame (d) of Fig. 2 shows an example of the skeleton detection results.

4.2. Signal Extraction

For each body part, we calculated spatial averages of skin pixels to reduce camera quantization error and generate 1D signals for each channel of RGB. CHROM [9] and POS [49] combined the signals of the three channels linearly to remove noise from movement or lighting to generate more robust pulse signals.

We used the generated bounding box coordinates to crop ROIs from each frame for each body part and down-

Table 1. Pulse rate estimation results. MAE: Mean Absolute Error; r: Pearson correlation coefficient.

| Methods | Both relaxed and hand raise | | | | | | | | | | | | Relaxed | | Hand raise | | | |
| | Face | | Right leg | | Left leg | | Right arm | | Left arm | | Left arm | | Left arm | | Palm | |
	MAE (bpm)	r	MAE (bpm)	r	MAE (bpm)	r	MAE (bpm)	r	MAE (bpm)	r	MAE (bpm)	r	MAE (bpm)	r	MAE (bpm)	r
CHROM [9]	2.38	0.85	10.92	0.42	11.07	0.41	9.13	0.50	9.81	0.41	11.57	0.35	4.26	0.71	5.01	0.67
POS [49]	**1.38**	**0.93**	**6.96**	**0.54**	**7.11**	**0.54**	**3.60**	**0.78**	**6.04**	**0.64**	**6.88**	**0.61**	**3.40**	**0.75**	**3.88**	**0.76**
RPNet [41]	2.27	0.87	29.50	0.14	30.42	0.11	23.94	0.15	23.15	0.16	27.01	0.11	11.06	0.38	6.70	0.52

sized these ROIs to 64x64 pixels using bicubic interpolation. The RPNet model we used was trained on the DDPM dataset [42, 47] where the frame rate is 90 frames per second (fps), which is the same as our MSPM dataset. The trained RPNet model was fed video clips of 135 frames (1.5 seconds) as described in the original paper [41].

4.3. Filtering and Pulse Rate Calculation

Remote pulse estimation over non-face regions is challenging, because of their lower signal-to-noise ratio than the face. To improve signal quality for all approaches, we applied a 4th order Butterworth bandpass filter with cutoff frequencies of 40 bpm and 180 bpm. Bandpass filtering was not used in the original POS and RPNet implementations, but we found the POS estimates in particular to contain high frequency noise before filtering. To compute the dominant pulsatile component in the rPPG waveforms, we applied the short-time Fourier transform (STFT) over a sliding 10-second window (900 frames in our videos), and selected the spectral peak as the pulse rate.

5. Global rPPG Experiments

5.1. Global Evaluation

We refer to rPPG with all pixels in a region of interest as "global" rPPG. We evaluated global rPPG performance over the face, arms, and legs with the color transformation approaches (CHROM [9] and POS [49]), and a 3D-CNN approach (RPNet [41]). We compared the rPPG quality by evaluating the pulse rate performance. We applied the same method for pulse rate estimation on both the ground truth signals and the rPPG signals for fair evaluation [28]. We used popular error metrics to compare the pulse rate estimates, including mean absolute error (MAE) and Pearson's correlation coefficient (r).

5.2. Global Results

The results of global rPPG experiments with different body parts are shown in Table 1. We show results for the entire sequence as well as for separate components in which the left arm is either relaxed or raised with left palm frontally presented. We observed the best MAE and r for

the face region. CHROM and RPNet give similar performance on the face region, with RPNet giving slightly better performance. RPNet performed poorly on the non-face regions, however, likely indicating that the model has learned spatial priors to look for facial features. Additionally, the deep learning model may be overfit to the skin thickness, melanin concentration, and microcirculation present in the glabrous skin of the face. The improved performance for RPNet on the palm (also with glabrous skin) helps justify this explanation.

The POS algorithm gives the best performance for all body regions in terms of both MAE and r. This is especially impressive given that it is a simple linear method. It also shows that color changes from blood volume are similar over different skin thicknesses and underlying microvasculature. For POS and CHROM, the order of performance from best to worst is face, arms, then legs. The palm also gives good performance for all approaches, due to physical similarities to the skin on the face.

Figure 4 shows predicted waveforms from a 10-second window for the same subject at different locations. The nearest contact PPG waveforms are bandpass filtered and plotted against the predictions to show that many of the predictions contain the underlying dominant pulse even in the presence of higher frequency noise. The difference in waveform morphology across body location shows how informative full-body rPPG can be. For this particular segment, the legs contain a strong second harmonic, which may arise from either the closure of the aortic valve during forward wave propagation, or wave reflections occurring at structural discontinuities along the femoral arteries [23]. Future studies on this dataset will explore different waveform morphologies at a finer scale and their relation to arterial stiffness and blood pressure.

Figure 5 shows pulse rates calculated from POS predictions over the entire session for 3 subjects. Interestingly, the errors generally occur as spikes of short duration rather than sustained periods of large offset. Even for the left arm, which is the worst-performing region for subject 0, the errors appear to be due to short periods where the second harmonic contains more power than the first harmonic. It is possible that simple heuristics during postprocessing could remove these transient errors. In general, the global POS
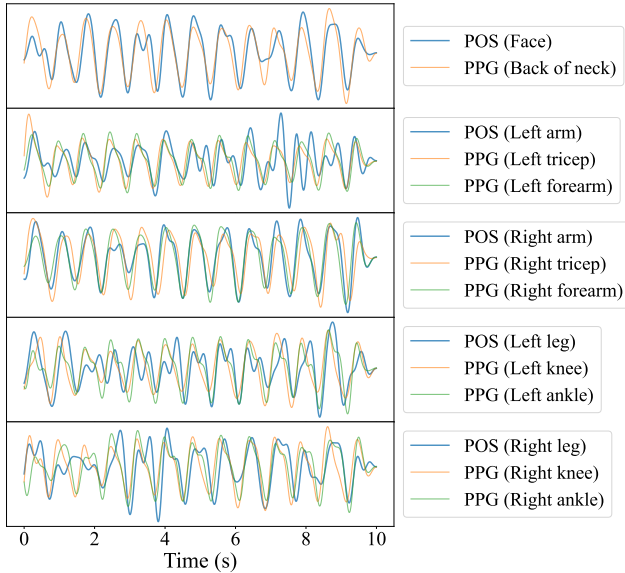
Figure 4. Pulse waveform predictions from POS overlayed with the nearest contact-PPG waveforms for a 10 second window.
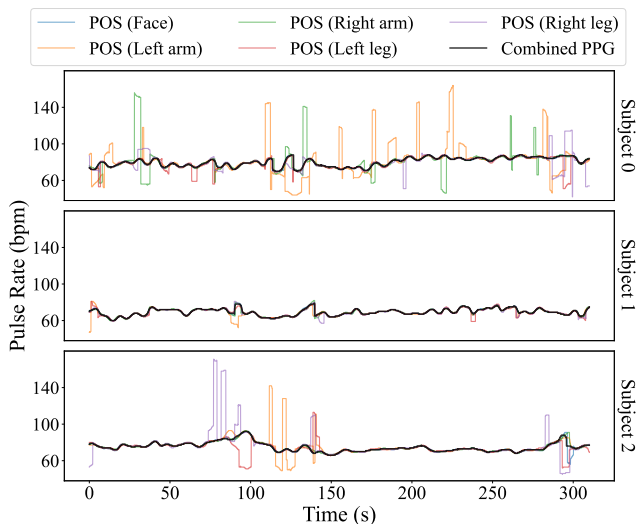

Figure 5. Pulse rate predictions from POS for 3 subjects over the course of a full session compared to the ground truth.

signals give meaningful predictions for most applications, even in non-face regions.

6. Local rPPG Experiments

6.1. Local Evaluation

In addition to analyzing performance for the entire face, arm, and legs, we performed a "local" analysis on subregions of pixels in the native video resolution. This gives a more detailed evaluation on rPPG quality over the visible skin on the body. We selected POS [49] as the rPPG

method, since it was the best performing for each of the global regions.

We used the ROIs described in section 4.1 to define bounds for a non-overlapping sliding window of 20x20 pixel subregions. The POS algorithm was applied on 10-second non-overlapping time windows of the spatially averaged subregions for the whole video. The predictions were then linearly interpolated up to the native image resolution (i.e. 20 times as large along the x- and y-axes). Next, we computed error metrics at each pixel location for the 10-second windows, which results in 9 error frames for the hand raising portion, and around 25 error frames for the sitting portion per subject.

For error metrics we used MAE between the predicted and ground truth pulse rate and signal-to-noise ratio (SNR). The SNR was calculated similarly to [9, 31], where the sum of spectral power around the signal bands in the first and second harmonics (±6 bpm) was divided by the total power outside the signal bands.

To remove spurious local errors outside of the skin pixel regions, we applied the average SCHP mask (see frame (b) in Fig. 2) for the time window to the error frames. Since subjects sat in slightly different positions throughout the interview, we aligned each body part across subjects for an accurate physiological error map. To do this we used the pose keypoints from MediaPipe (see frame (d) in Fig. 2). We first calculated the average pose across all subjects for the hand raising and relaxed portions as the target poses. Then for each error frame for every subject, we applied a homographic transformation from the time window's pose to the hand raising or relaxed portion's target pose.

6.2. Local Results

Figure 6 shows a heatmap of local pulse estimation performance for the POS algorithm. The hand raising and sitting portions are visualized separately, since the body positions were much different. In general, we found the face to hold the strongest rPPG signal. In the hand raising heatmap, we can see that the palm gave relatively low errors compared with the arms and legs. This supports the hypothesis that signal quality is higher on glabrous than non-glabrous skin [6], and shows promise for substituting the face region with the hand if necessary.

It is useful to analyze each body part separately to assess the signal quality. Firstly, the face appears to give low MAE and high SNR in all regions except the eyes and mouth. This is in line with past research that mainly utilizes the forehead and cheek regions. The arms give relatively low signal quality, but there is a slight improvement visible above the forearm over other regions in the SNR plots. The legs give perhaps the weakest rPPG signal as evidenced by the global results in Table 1. Within the local analysis, we can see that the thighs give higher SNR than the shins. This could

Mean Absolute Error (MAE)   Signal-to-Noise Ratio (SNR)

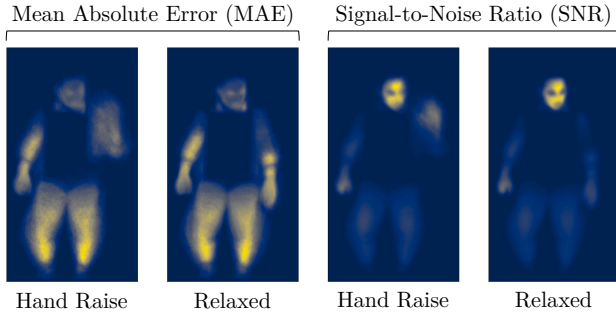Hand Raise   Relaxed   Hand Raise   Relaxed

Figure 6. A heatmap of the local POS signal quality across the visible skin. POS signals are predicted over 20x20 pixel subregions, masked with SCHP [19], and the human pose is aligned with a homographic transformation.

be due to the thigh's closer proximity to the illuminators, whereas the shins are visibly darker in the video.

In both the hand raising and relaxed portions, the hand gives the second best SNR. In nearly all cases the subject's right hand was resting with the palm facing downwards, indicating that the back of the fingers and hand on non-glabrous skin is still feasible for rPPG. During the hand raising, we see that the signal quality is high on the glabrous skin of the palm. With more sophisticated methods for aligning each subject's palms, we believe the error maps would give even higher signal quality.

7. Pulse Transit Time Experiments

7.1. Pulse Transit Time for Contact PPG Sensors

Contact PPG measurements at different sites on the body contain small phase offsets due to different proximities to the heart. These offsets are sometimes referred to as differential pulse transit times (dPTT) [4] and are negatively correlated with blood pressure. Throughout the paper we treat PTT and dPTT interchangeably. When fusing the PPG measurements at multiple sites into a single ground truth, we summed the waveforms without applying a phase shift. To justify this simplification, we calculated the differential pulse transit time between all pairs of sensors. Fig. 7a shows a heatmap of the phase offsets between all pairs of PPG signals in milliseconds.

The phase differences were calculated with a sliding cross-correlation. We used a window size of 5 seconds (2,000 points) with a stride of 10 milliseconds (4 points), and a maximum lag 300 milliseconds, which is much higher than a typical transit time [4]. The sliding cross-correlation approach for PTT analysis is simple, but could be improved in future work by using foot-finding methods that measure time differences at the diastolic foot [29]. The index with the maximum correlation was selected as the phase shift

for each window. All pairs of transit times formed a skew-symmetric matrix.

In general, the phase offset between different sites is quite small relative to the pulse rate frequency. For example, the largest average phase offset of 51 milliseconds occurs between the left tricep and the right ankle. Given even a high pulse rate of 180 bpm, the phase angle for such a shift is only 27.66 degrees. Since most phase offsets are lower, there is very little risk of interference when summing the waveforms. The pulse transit times provide interesting physiological measurements of blood flow throughout the body. As mentioned previously, the largest difference occurs between the triceps and ankle. The triceps receive blood faster via the brachial artery and a closer proximity to the aorta than the lower legs. We had originally theorized that the neck sensor would sense the pulse wave before the other sensors, but our analysis refutes this.

7.2. Pulse Transit Time from rPPG

Beyond the benefit of contactless monitoring, one of the most powerful properties of a camera is the ability to sense at multiple spatial locations. To leverage this, we performed remote pulse transit time (rPTT) with a sliding cross-correlation from the POS rPPG waveforms at multiple sites on the body. The POS waveforms are noisier than the PPG waveforms in the legs and arms, but the infrequent pulse rate errors (see Fig. 5) indicate that the signal quality is high enough for rPTT.

Figure 7b shows a heatmap of the rPTT between all pairs of the face, legs, and arms. The maximum observed rPTT is 46.75 ms between the face and left leg, which is within a reasonable range compared to the largest contact-PTT measurements between triceps and legs (51.23 ms and 48.82 ms). An interesting observation is the bilateral asymmetry visible in the rPTT left and right sides of the body. The pulse waves from the right side of the body were consistently observed sooner than the left side of the body.

To compare the rPTT measurements with the PPG-derived PTT values we selected the nearest contact sensors to the rPTT measurement sites, and discarded the rest (*e.g.* ankles and triceps). Figure 7c shows boxplots of the PTT values for all 5-second time windows in the relaxation portion of the dataset. The groups were sorted by the median PTT at the pair of measurement sites for easier viewing. The means of the rPTT measurements via POS are not perfectly aligned with the PTTs — but upon further inspection, we can see that rPTT and PTT correlate well across the pairs of body sites. A potential reason for the bias in the rPTT values is that the signal quality is not uniform over the measured body sites. For example, the POS signal on the arms is likely shifted in phase towards the pulse wave at the hands. Similarly, the POS wave is likely shifted more towards the phase of the thighs than the lower legs. Fur-
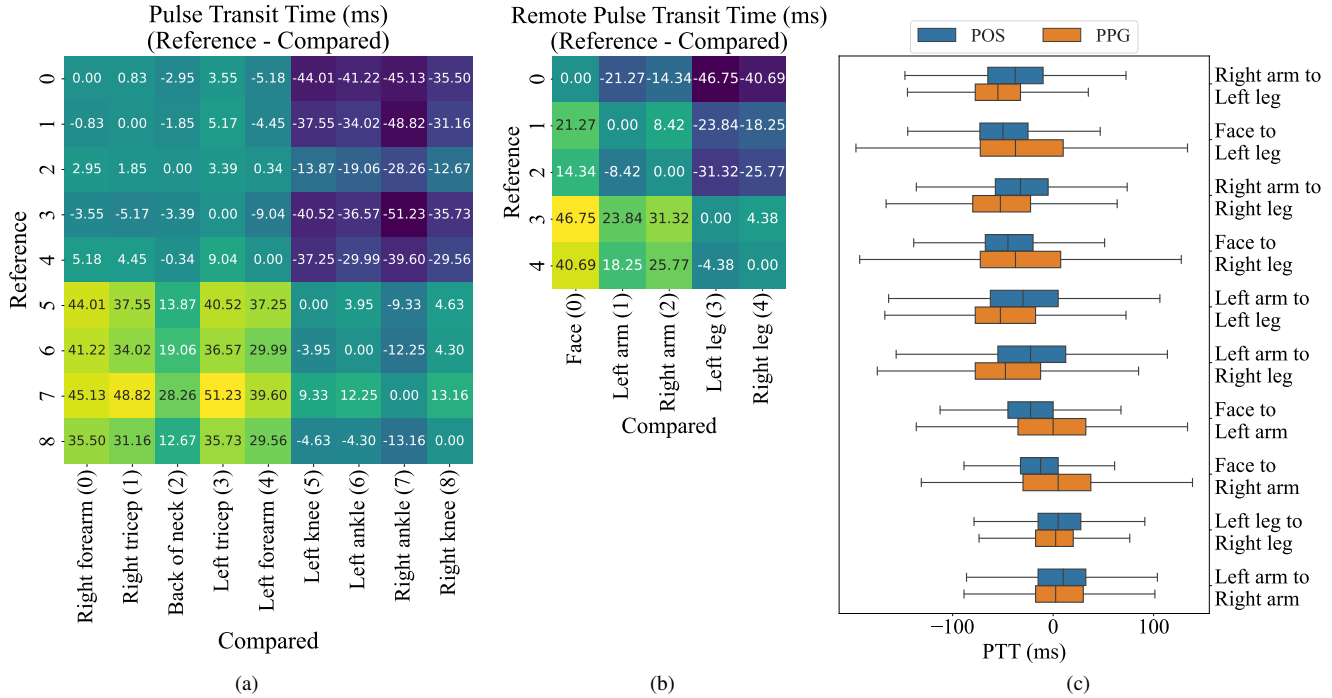
Figure 7. Pulse transit time from both contact PPG and rPPG calculated by a sliding cross-correlation between pairs of waveforms. (a) Time lags between the MAX30101 contact PPG sensors. (b) Time lags between POS waveform predictions on the five different body regions. (c) Grouped boxplots comparing the pulse transit time from PPG and rPPG. For the PPG measurements on the arms and legs we only used the forearm and knee sensors, respectively. The PTT is calculated as $X - Y$ given a label of measurement sites "X to Y".

thermore, the sampling rate of the video is 90 fps, so large time lags such as 55 ms will only occur as 5 frames. Future work will perform more robust transit time calculations via the systolic foot and finer-grained spatial measurements of rPTT to better align with the contact-PPG sensors.

8. Conclusion

This paper offers a new Multi-Site Physiological Monitoring (MSPM) dataset with multiple contact-based PPG measurements and simultaneous video to enable rPPG from multiple parts of the human body including the face, arms, legs, and neck. We applied various rPPG methods, including CHROM [9], POS [49] and RPNet [42], to MSPM samples and performed pulse rate estimation on all face and non-face body regions. We observed that the heart rate estimation from non-face body parts, especially palm and arms, showed a promising accuracy compared to face-based rPPG. However, the deep learning-based (3D-CNN) approach evaluated in this paper, and trained entirely on facial data, learned spatial pulse estimation priors which disrupted its ability to generalize to non-facial regions. This suggests that training deep learning rPPG systems using non-facial regions may be needed to offer good fidelity for rPPG outside the face region. We also positively verified a possibility to apply rPPG to estimate pulse transit times (PTT)

between various body parts by comparing rPPG-based PTT with PPG-based PTT. We believe that these baseline experiments along with the new MSPM dataset opens an interesting research area focused on non-face multi-site rPPG.

References

[1] John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3), 2007. 1, 2

[2] Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, and Matthias Grundmann. Blazepose: On-device real-time body pose tracking. *CoRR*, abs/2006.10204, 2020. 3, 4

[3] Michael Bentham, Gerard Stansby, and John Allen. Innovative multi-site photoplethysmography analysis for quantifying pulse amplitude and timing variability characteristics in peripheral arterial disease. *Diseases*, 6(3), Sept. 2018. 1

[4] Robert C. Block, Mohammad Yavarimanesh, Keerthana Natarajan, Andrew Carek, Azin Mousavi, Anand Chandrasekhar, Chang Sei Kim, Junxi Zhu, Giovanni Schifitto, Lalit K. Mestha, Omer T. Inan, Jin Oh Hahn, and Ramakrishna Mukkamala. Conventional pulse transit times as markers of blood pressure changes in humans. *Scientific Reports*, 10(1):1–9, 2020. 1, 2, 7

[5] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, and Julien Dubois. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, 124:82–90, 2019. 1

[6] Meiyun Cao, Timothy Burton, Gennadi Saiko, and Alexandre Douplik. Remote photoplethysmography with a high-speed camera reveals temporal and amplitude differences between glabrous and non-glabrous skin. *Sensors*, 23(2), 2023. 3, 6

[7] Gabriel Chan, Rachel Cooper, Manish Hosanee, Kaylie Welykholowa, Panayiotis A Kyriacou, Dingchang Zheng, John Allen, Derek Abbott, Nigel H Lovell, Richard Fletcher, and Mohamed Elgendi. Multi-site photoplethysmography technology for blood pressure assessment: Challenges and recommendations. *J. Clin. Med.*, 8(11):1827, Nov. 2019. 1

[8] Weixuan Chen and Daniel McDuff. DeepPhys: Video-based physiological measurement using convolutional attention networks. In *European Conference on Computer Vision (ECCV)*, pages 356–373, 2018. 2

[9] G. De Haan and V. Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013. 1, 2, 4, 5, 6, 8

[10] John Gideon and Simon Stent. The way to my heart is through contrastive learning: Remote photoplethysmography from unlabelled video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3995–4004, October 2021. 2

[11] M Harford, M Villarroel, J Jorge, O Redfern, E Finnegan, S Davidson, J D Young, L Tarassenko, and P Watkinson. Contactless skin perfusion monitoring with video cameras: tracking pharmacological vasoconstriction and vasodilation using photoplethysmographic changes. *Physiological Measurement*, 43(11):115001, nov 2022. 3

[12] Yungchien Hsu, Yen Liang Lin, and Winston Hsu. Learning-based heart rate detection from remote photoplethysmography features. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4433–4437. IEEE, 2014. 2

[13] Ruqiang Huang, Weihua Su, Shiyue Zhang, and Wei Qin. Remote Measurement of Vital Signs for Unmanned Search and Rescue Vehicles. In *2020 5th International Conference on Control and Robotics Engineering (ICCRE)*, pages 164–168, 2020. 1

[14] Janne M. J. Huttunen, Leo Kärkkäinen, and Harri Lindholm. Pulse transit time estimation of aortic pulse wave velocity and blood pressure using machine learning and simulated training data. *PLOS Computational Biology*, 15(8):1–23, 08 2019. 1

[15] Kaito Iuchi, Ryogo Miyazaki, George C. Cardoso, Keiko Ogawa-Ochiai, and Norimichi Tsumura. Blood pressure estimation by spatial pulse-wave dynamics in a facial video. *Biomedical Optics Express*, 13(11):6035, Nov. 2022. 1, 2

[16] Kaito Iuchi, Ryogo Miyazaki, George C. Cardoso, Keiko Ogawa-Ochiai, and Norimichi Tsumura. Remote estimation of continuous blood pressure by a convolutional neural network trained on spatial patterns of facial pulse waves. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2139–2145, June 2022. 1

[17] In Cheol Jeong and Joseph Finkelstein. Introducing Contactless Blood Pressure Assessment Using a High Speed Video Camera. *Journal of Medical Systems*, 40(4):77, Apr. 2016. 1, 2

[18] Alexei A. Kamshilin, Serguei Miridonov, Victor Teplov, Riku Saarenheimo, and Ervin Nippolainen. Photoplethysmographic imaging of high spatial resolution. *Biomed. Opt. Express*, 2(4):996–1006, Apr 2011. 3

[19] Peike Li, Yunqiu Xu, Yunchao Wei, and Yi Yang. Self-correction for human parsing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 3, 4, 7

[20] Xiaodan Liang, Ke Gong, Xiaohui Shen, and Liang Lin. Look into person: Joint body parsing & pose estimation network and a new benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(4):871–885, 2019. 3

[21] Xin Liu, Josh Fromm, Shwetak Patel, and Daniel McDuff. Multi-task temporal shift attention networks for on-device contactless vitals measurement. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 19400–19411. Curran Associates, Inc., 2020. 2

[22] Xin Liu, Brian Hill, Ziheng Jiang, Shwetak Patel, and Daniel McDuff. Efficientphys: Enabling simple, fast and accurate camera-based cardiac measurement. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 5008–5017, January 2023. 2

[23] Gérard M. London and Alain P. Guerin. Influence of arterial pulse and reflected waves on blood pressure and cardiac function. *American Heart Journal*, 138(3):S220–S224, Sept. 1999. 5

[24] Stavros Loukogeorgakis, Rex Dawson, Nirree Phillips, Christopher N Martyn, and Stephen E Greenwald. Validation of a device to measure arterial pulse wave velocity by a photoplethysmographic method. *Physiological Measurement*, 23(3):581, jul 2002. 1

[25] M. Lubin, R. Gerbelot, R. Prada, J. Porcherot, and S. Bonnet. Evaluation of a dual-ppg system for pulse transit time moni-

toring. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1349–1352, 2021. 2

[26] Corey McCall, Rea Rostosky, Richard M. Wiard, Omer T. Inan, Laurent Giovangrandi, Charles Marsh Cuttino, and Gregory T.A. Kovacs. Noninvasive pulse transit time measurement for arterial stiffness monitoring in microgravity. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6429–6432, 2015. 1

[27] Daniel McDuff, Miah Wander, Xin Liu, Brian L. Hill, Javier Hernandez, Jonathan Lester, and Tadas Baltrusaitis. SCAMPS: Synthetics for camera measurement of physiological signals. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022. 1

[28] Yuriy Mironenko, Konstantin Kalinin, Mikhail Kopeliovich, and Mikhail Petrushan. Remote photoplethysmography: Rarely considered factors. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1197–1206, June 2020. 5

[29] Ramakrishna Mukkamala, Jin Oh Hahn, Omer T. Inan, Lalit K. Mestha, Chang Sei Kim, Hakan Toreyin, and Survi Kyal. Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice. *IEEE Transactions on Biomedical Engineering*, 62(8):1879–1901, 2015. 1, 2, 7

[30] Keerthana Natarajan, Robert C. Block, Mohammad Yavarimanesh, Anand Chandrasekhar, Lalit K. Mestha, Omer T. Inan, Jin Oh Hahn, and Ramakrishna Mukkamala. Photoplethysmography Fast Upstroke Time Intervals Can Be Useful Features for Cuff-Less Measurement of Blood Pressure Changes in Humans. *IEEE Transactions on Biomedical Engineering*, 69(1):53–62, 2022. 2

[31] Ewa M. Nowara, Daniel McDuff, and Ashok Veeraraghavan. Systematic analysis of video-based pulse measurement from compressed videos. *Biomedical Optics Express*, 12(1):494, 2021. 6

[32] M. Poh, D. J. McDuff, and R. W. Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, 58(1):7–11, 2011. 2

[33] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express*, 18(10):10762–10774, May 2010. 2

[34] Smera Premkumar and Duraisamy Jude Hemanth. Intelligent remote photoplethysmography-based methods for heart rate estimation from face videos: A survey. *Informatics*, 9(3), 2022. 1

[35] Vicent Ribas Ripoll and Alfredo Vellido. Blood Pressure Assessment with Differential Pulse Transit Time and Deep Learning: A Proof of Concept. *Kidney Diseases*, 5(1):23–27, 2019. 2

[36] Gennadi Saiko, M. Dervenis, and Alexandre Douplik. On the feasibility of pulse wave velocity imaging for remote assessment of physiological functions. *Advances in Experimental Medicine and Biology*, 1269:393–397, 05 2021. 3

[37] Lorenzo Scalise, Natascia Bernacchia, Ilaria Ercoli, and Paolo Marchionni. Heart rate measurement in neonatal patients using a webcamera. In *2012 IEEE International Symposium on Medical Measurements and Applications Proceedings*, pages 1–4, 2012. 1

[38] Dangdang Shao, Yuting Yang, Chenbin Liu, Francis Tsow, Hui Yu, and Nongjian Tao. Noncontact monitoring breathing pattern, exhalation flow rate and pulse transit time. *IEEE Transactions on Biomedical Engineering*, 61(11):2760–2767, 2014. 3

[39] Robin P Smith, Jérôme Argod, Jean-Louis Pépin, and Patrick A Lévy. Pulse transit time: an appraisal of potential clinical applications. *Thorax*, 54(5):452–457, 1999. 1

[40] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1):42–55, 2012. 1

[41] Jeremy Speth, Nathan Vance, Adam Czajka, Kevin Bowyer, and Patrick Flynn. Unifying frame rate and temporal dilations for improved remote pulse detection. *Computer Vision and Image Understanding (CVIU)*, pages 1056–1062, 2021. 2, 4, 5

[42] Jeremy Speth, Nathan Vance, Adam Czajka, Kevin Bowyer, Diane Wright, and Patrick Flynn. Deception detection and remote physiological monitoring: A dataset and baseline experimental results. In *International Joint Conference on Biometrics (IJCB)*, pages 4264–4271, 2021. 1, 4, 5, 8

[43] Jeremy Speth, Nathan Vance, Patrick Flynn, and Adam Czajka. Non-contrastive unsupervised learning of physiological signals from video. *ArXiv*, abs/2303.07944, 2023. 2

[44] Ronny Stricker, Steffen Muller, and Horst Michael Gross. Non-contact video-based pulse rate measurement on a mobile service robot. *IEEE International Symposium on Robot and Human Interactive Communication*, pages 1056–1062, 2014. 1

[45] Zhaodong Sun and Xiaobai Li. Contrast-phys: Unsupervised video-based remote physiological measurement via spatiotemporal contrast. *European Conference on Computer Vision (ECCV)*, 2022. 2

[46] Victor Teplov, Ervin Nippolainen, Alexander A. Makarenko, Rashid Giniatullin, and Alexei A. Kamshilin. Ambiguity of mapping the relative phase of blood pulsations. *Biomedical Optics Express*, 5(9):3123, 2014. 3

[47] Nathan Vance, Jeremy Speth, Siamul Khan, Adam Czajka, Kevin W. Bowyer, Diane Wright, and Patrick Flynn. Deception detection and remote physiological monitoring: A dataset and baseline experimental results. *IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, pages 1–1, 2022. 1, 4, 5

[48] Hao Wang, Euijoon Ahn, and Jinman Kim. Self-Supervised Representation Learning Framework for Remote Physiological Measurement Using Spatiotemporal Augmentation Loss. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(2):2431–2439, June 2022. 2

[49] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan. Algorithmic principles of remote PPG. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2017. 2, 4, 5, 6, 8

[50] Wenjin Wang, Sander Stuijk, and Gerard De Haan. A Novel Algorithm for Remote Photoplethysmography: Spatial Subspace Rotation. *IEEE Transactions on Biomedical Engineering*, 63(9):1974–1984, 2016. 2

[51] Zhen Wang, Yunhao Ba, Pradyumna Chari, Oyku Deniz Bozkurt, Gianna Brown, Parth Patwa, Niranjan Vaddi, Laleh Jalilian, and Achuta Kadambi. Synthetic generation of face videos with plethysmograph physiology. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20555–20564, 2022. 1

[52] Daniel Wedekind, Alexander Trumpp, Frederik Gaetjen, Stefan Rasche M.D., Klaus Matschke M.D., Hagen Malberg, and Sebastian Zaunseder. Assessment of blind source separation techniques for video-based cardiac pulse extraction. *Journal of Biomedical Optics*, 22(3):035002, 2017. 2

[53] Yuzhe Yang, Xin Liu, Jiang Wu, Silviu Borac, Dina Katabi, Ming-Zher Poh, and Daniel McDuff. SimPer: Simple self-supervised learning of periodic targets. In *International Conference on Learning Representations (ICLR)*, 2023. 2

[54] Zitong Yu, Xiaobai Li, and Guoying Zhao. Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 1–12, 2019. 2

[55] Valeriy V. Zaytsev, Serguei V. Miridonov, Oleg V. Mamontov, and Alexei A. Kamshilin. Contactless monitoring of the blood-flow changes in upper limbs. *Biomed. Opt. Express*, 9(11):5387–5399, Nov 2018. 3

[56] Qi Zhan, Jingjing Hu, Zitong Yu, Xiaobai Li, and Wenjin Wang. Revisiting motion-based respiration measurement from videos. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2020:5909–5912, July 2020. 1

[57] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. Mediapipe hands: On-device real-time hand tracking. *CoRR*, abs/2006.10214, 2020. 3, 4

[58] Yong-Liang Zhang, Ying-Ying Zheng, Zu-Chang Ma, and Yi-Ning Sun. Radial pulse transit time is an index of arterial stiffness. *Hypertension research : official journal of the Japanese Society of Hypertension*, 34:884–7, 05 2011. 1

[59] Zheng Zhang, Jeff M. Girard, Yue Wu, Xing Zhang, Peng Liu, Umur Ciftci, Shaun Canavan, Michael Reale, Andy Horowitz, Huiyuan Yang, Jeffrey F. Cohn, Qiang Ji, and Lijun Yin. Multimodal spontaneous emotion corpus for human behavior analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 1