Camera Measurement of Physiological Vital Signs

DANIEL MCDUFF, Google

The need for remote tools for healthcare monitoring has never been more apparent. Camera measurement of vital signs leverages imaging devices to compute physiological changes by analyzing images of the human body. Building on advances in optics, machine learning, computer vision, and medicine, these techniques have progressed significantly since the invention of digital cameras. This article presents a comprehensive survey of camera measurement of physiological vital signs, describing the vital signs that can be measured and the computational techniques for doing so. I cover both clinical and non-clinical applications and the challenges that need to be overcome for these applications to advance from proofs of concept. Finally, I describe the current resources (datasets and code) available to the research community and provide a comprehensive webpage (https://cameravitals.github.io/) with links to these resource and a categorized list of all papers referenced in this article.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing**; • **Applied computing** → **Bioinformatics**; • **Computing methodologies** → **Computer vision;**

Additional Key Words and Phrases: Physiology, signal processing, machine learning, thermal imaging

ACM Reference format:
Daniel McDuff. 2023. Camera Measurement of Physiological Vital Signs. *ACM Comput. Surv.* 55, 9, Article 176 (January 2023), 40 pages.
https://doi.org/10.1145/3558518

1 INTRODUCTION

Camera measurement of vital signs has emerged as a vibrant field within computer vision and computational photography. This work combines expertise from these domains and those of signal processing, machine learning, biomedical engineering, optics, and medicine to create technologies that enable scalable and accessible physiological monitoring. The field has grown rapidly in the past 20 years, with papers published at an exponential growth rate (Figure 1 presents an example). Using cameras for non-contact measurement has several distinct advantages and applications in a range of contexts. In telehealth, remote measurement of vital signs is an important tool in assessment and diagnosis, and cameras are a ubiquitous form of sensor available on almost every digital communications device (cell phone, PC, etc.). In inpatient ICU care, remote measurement can help protect patients and physicians while also creating a more comfortable experience for those receiving care, whether it be a baby [1] or an adult [148]. Less invasive sensing can help patients sleep and eat while still being monitored. In low-resource settings, cameras are a cost-effective and

Author's address: D. McDuff, Google, Seattle, WA; email: dmcduff@google.com.

0360-0300/2023/01-ART176
https://doi.org/10.1145/3558518

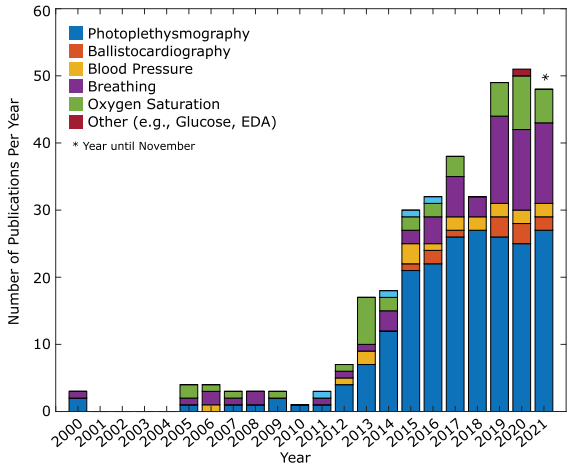

Fig. 1. The number of publications per year indexed on PubMed.gov on contactless camera measurement of physiological vital signs.

widely available form of sensor that can be easily transported and used for opportunistic measurement. Camera measurement of vitals could turn billions of devices with webcams into instruments for healthcare.

However, camera physiological measurement presents numerous challenges that must be overcome before this potential can be fully realized. Physiological changes are often subtle, are individually and contextually variable, and can easily be obscured by clothing, hair, and makeup. Other changes in a video, such as from motion and illumination, can swamp the small pixel variations that contain cardiac and pulmonary information. Last, but by no means least, are the serious ethical and privacy implications of non-contact measurement.

This article presents a survey of the field from foundations to state-of-the-art computational methods, discusses the applications of these tools, and highlights challenges and opportunities for the research community. In this survey, I focus on technologies that use visual spectrum, **near-infrared (NIR)**, and **far-infrared (FIR)** (or thermal) cameras. These are all non-ionizing regions of the electromagnetic spectrum, making imaging safe for extended periods of time and in many cases possible without an active or dedicated light source. To accompany this survey, a website has been prepared with all referenced papers categorized with key words and links to open source code repositories and datasets.

In my search of the literature, I used the following key words: 'remote', 'imaging', 'non-contact', 'camera', 'video', 'physiology', 'photoplethysmography', 'rppg', 'ippg'/'ppgi', 'ballistocardiography', 'respiration', 'breathing', 'pulse', 'blood pressure', 'electrodermal activity', 'oxygen saturation', and 'glucose'. I primarily searched on Google Scholar, Microsoft Academic, and PubMed. As an example, via PubMed, I found more than 215 papers on camera **photoplethysmography (PPG)** and **ballistocardiography (BCG)** that have been published in the past 5 years, which is an increase from approximately 60 in the previous 5 years [87]. Across all terms, I found more than 350 papers published on camera physiological measurement on PubMed alone.[1] There exist other surveys of camera methods [34, 87, 101, 131, 145] and some complementary comparative studies [101, 160]. However, some of these only focus on signal processing methods and were published before supervised learning and deep learning came to the fore [87, 145] and others focus only on deep

[1]Based on results from https://pubmed.ncbi.nlm.nih.gov/.

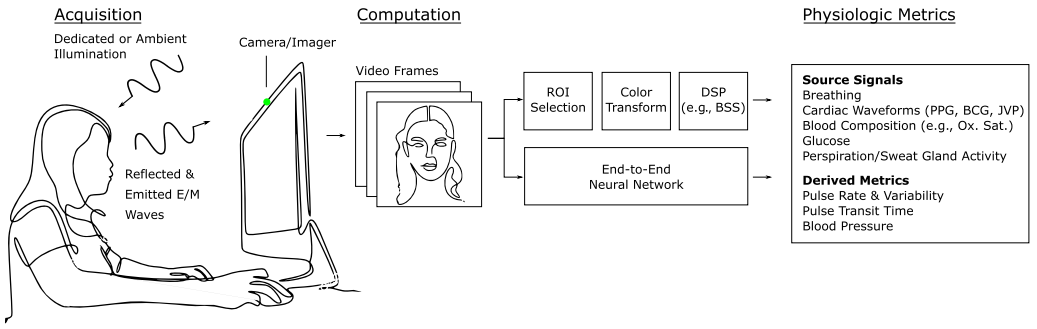

Fig. 2. Camera measurement of vital signs has emerged as a vibrant field within computer vision and computational photography. Computational methods can be used to recover a range of physiological measures from imaging in the visual, NIR, and FIR frequency ranges.

learning methods [101], whereas still others do not comprehensively cover the topic from foundations to computational methods to applications [131, 160]. I argue that given the significant advances in the research community in recent years and the interest in these tools given the growth of telehealth platforms, a systematic survey of the field is warranted. A thorough survey of the literature will help establish the current state of the art, synthesize insights from different approaches, and solidify key challenges for the community to solve.

2 FOUNDATIONS

The advent of digital cameras created new opportunities for computation analysis of human bodies and physiology. Blazek et al. [20] proposed the first imaging system for measuring cardiopulmonary signals. This computer-based CCD NIR imaging system provided evidence that peripheral blood volume could be measured without contact using an imager. Shortly after this, a similar approach was demonstrated using a visual band (RGB) camera [171], devices that are considerably more ubiquitous than NIR cameras. Successful replications of this work cemented the concept [45, 60, 146, 154] and led to the growth of a new field of non-contact camera physiological measurement. Figure 2 illustrates the typical imaging pipeline for these systems. As E/M wavelengths increase, the depth at which the waves penetrate the skin also increases (Figure 3); however, so does the amount of scattering that occurs. Depending on the signal of interest, there is a trade-off between how much light is absorbed by the body and how much is reflected. Oxygenated blood, deoxygenated blood, and skin tissue all have different absorption characteristics. Fortunately, within the visual bands close to 500 to 600 nm in the "green" color range, there is a good trade-off between light penetration depth and hemoglobin absorption, making ubiquitously available RGB cameras a valuable tool for measuring cardiac signals via PPG. Camera technology has also been improving rapidly in quality due to investment from cell phone and other smart device manufacturers. Increases in sensor resolution and frame rate and reductions in sensor noise mean that subtle changes in motion due to pulmonary and cardiac activity can be captured. Work on body motion analysis from video has found this to be a rich source of physiological information, enabling the recovery of breathing [148] and cardiac signals [11]. These methods do not require light to penetrate the skin but rather use optical flow and other motion tracking methods to measure usually very small motions.

2.1 Optical Model

Optical models serve as a principled foundation for designing computational methods for camera physiological measurement. For modeling lighting, imagers, and physiology, previous works have

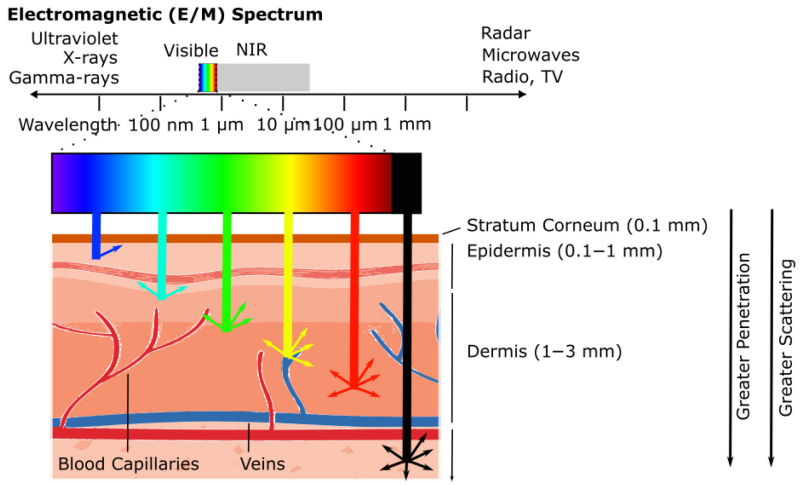

Fig. 3. Camera physiological sensing uses non-ionizing E/M wavelengths to measure vital signs. As E/M wavelengths increase, the depth at which they penetrate the skin also increases; however, so does the amount of scattering that occurs.

used the Lambert-Beer law [67] and Shafer's dichromatic reflection model [31, 76]. As an example, let us take the dichromatic reflection model as our basis. Via analysis of video pixels, we aim to capture both spatial and temporal changes and how these relate to multiple physiological processes. We start with the RGB values captured by the cameras as given by

$$\boldsymbol{C}_k(t) = I(t) \cdot (\boldsymbol{v}_s(t) + \boldsymbol{v}_d(t)) + \boldsymbol{v}_n(t), \tag{1}$$

where $I(t)$ is the luminance intensity level, modulated by the specular reflection $\boldsymbol{v}_s(t)$ and the diffuse reflection $\boldsymbol{v}_d(t)$. The quantization noise of the camera sensor is captured by $\boldsymbol{v}_n(t)$. $I(t)$ can be decomposed into two parts, $\boldsymbol{v}_s(t)$ and $\boldsymbol{v}_d(t)$, respectively [163]:

$$\boldsymbol{v}_d(t) = \boldsymbol{u}_d \cdot d_0 + \boldsymbol{u}_p \cdot p(t). \tag{2}$$

$\boldsymbol{u}_d$ is the skin-tissue unit color vector; d_0 is the reflection strength, which is stationary; $\boldsymbol{u}_p$ is the relative pulsatile strength caused by the hemoglobin and melanin absorption; and $p(t)$ represents the underlying physiological signals of interest.

$$\boldsymbol{v}_s(t) = \boldsymbol{u}_s \cdot (s_0 + \Phi(m(t), p(t))), \tag{3}$$

where $\boldsymbol{u}_s$ is the unit color vector of the light source spectrum; s_0 and $\Phi(m(t), p(t))$ denote the stationary and varying parts of specular reflections; and $m(t)$ denotes all the non-physiological variations such as changes in the illumination, head rotation, and facial expressions.

$$I(t) = I_0 \cdot (1 + \Psi(m(t), p(t))), \tag{4}$$

where I_0 is the stationary component of the luminance, and $I_0 \cdot \Psi(m(t), p(t))$ is the intensity variation as captured by the camera. As in the work of Chen and McDuff [31], we can disregard products of time-varying components, as they are relatively small, giving

$$\boldsymbol{C}_k(t) \approx \boldsymbol{u}_c \cdot I_0 \cdot c_0 + \boldsymbol{u}_c \cdot I_0 \cdot c_0 \cdot \Psi(m(t), p(t)) + \boldsymbol{u}_s \cdot I_0 \cdot \Phi(m(t), p(t)) + \boldsymbol{u}_p \cdot I_0 \cdot p(t) + \boldsymbol{v}_n(t). \tag{5}$$

Pulse and breathing signals are in fact not independent [76]. As an example, the PPG signal captures a complex combination of both pulse and breathing information. Specifically, both the specular and diffuse reflections are influenced by related physiological processes. A **respiratory sinus**

arrhythmia (RSA) is one example of this; RSA describes the rhythmical fluctuations in heart periods at the breathing frequency [13]. Another example is that breathing and cardiac pulse signals both cause observable motions of the body. We can say that the physiological process *p(t)* is a complex combination of the photoplethysmographic *ppg(t)*, the ballistocardiographic *bcg(t)*, and the breathing wave *r(t)*. Thus, $p(t) = \Theta(ppg(t), bcg(t), r(t))$ and the following equation gives a more accurate representation of the underlying process:

$$\boldsymbol{C}_k(t) \approx \boldsymbol{u}_c \cdot I_0 \cdot c_0 + \boldsymbol{u}_c \cdot I_0 \cdot c_0 \cdot \Psi(m(t), \Theta(ppg(t), bcg(t), r(t))) + \boldsymbol{u}_s \cdot I_0 \cdot \Phi(m(t), \Theta(ppg(t), bcg(t), r(t))) + \boldsymbol{u}_p \cdot I_0 \cdot p(t) + \boldsymbol{v}_n(t). \quad (6)$$

There are alternatives to this optical model, and it does not capture all physiological changes. For example, around the neck, the **jugular venous pulse (JVP)** would be observed. However, this example does provide a basis or foundation for thinking about the design of computational methods that separate the source signals of interest from noise.

3 HARDWARE

3.1 Visual Spectrum (RGB and Grayscale) Cameras

The visual light spectrum covers frequencies from 380 to 700 nm. By far, the most ubiquitous form of imager is the RGB camera. RGB imagers include webcams, cell phone cameras, and digital photography cameras (e.g., DSLRs). Cameras are even now included on some smart TVs, in-home smart devices and doorbells, refrigerators, and mirrors. These devices are typically optimized for visual clarity, creating images and videos that are clear to the human eye. Furthermore, they have often been optimized for affluent, Western, and Asian consumers, thus capturing lighter skin types more effectively then darker skin types.[2] Figure 4 shows the distribution of skin pixels for people from several countries around the world. Notice how the histogram of skin pixel values for those from Western countries (i.e., UK, Germany, Australia) fall close to the middle of the 0 to 255 pixel range with a Gaussian or normal distribution, whereas for those from African countries (Mali, Nigeria, Ivory Coast, etc.) or the Caribbean (Jamaica), the pixel values are skewed closer to 0. Pixel saturation is more likely to occur for those subjects, which would cause the changes in a video due to physiological variations to be lost. To compound this, face detection algorithms [92] and similar tools [25] that are often used in camera physiological measurement pipelines often have biases. As such, biases in the performance of physiological measurement using cameras not only stems from the optimization criteria, models, and training data that are used but also from the hardware. I will discuss attempts to characterize and correct these biases in Section 10.1. However, it should be noted that little work has attempted to address disparities in performance resulting from hardware.

Assuming uniform illumination (e.g., broad spectrum/white light), the maximum **signal-to-noise ratio (SNR)** for the **blood volume pulse (BVP)** is at approximately 570 nm [18]; this is the frequency at which the absorption of hemoglobin is greatest. However, if the illumination was particularly strong at another frequency, this could change. Surprisingly, measurements with RGB cameras can be made with reasonable precision up to 50 m from the subject [17], which highlights not only the potential for this technology in remote measurement but also the potential for it to be used for covert surveillance and other troubling applications. I will discuss the broader impacts of camera physiological measurement in Section 10.5.

3.2 NIR Cameras

NIR cameras sense light with wavelengths from 700 to 1,000 nm. Human eyes are not sensitive to this wavelength range, and therefore imaging systems can be designed with dedicated active light

[2] https://petapixel.com/2015/09/19/heres-a-look-at-how-color-film-was-originally-biased-toward-white-people/.

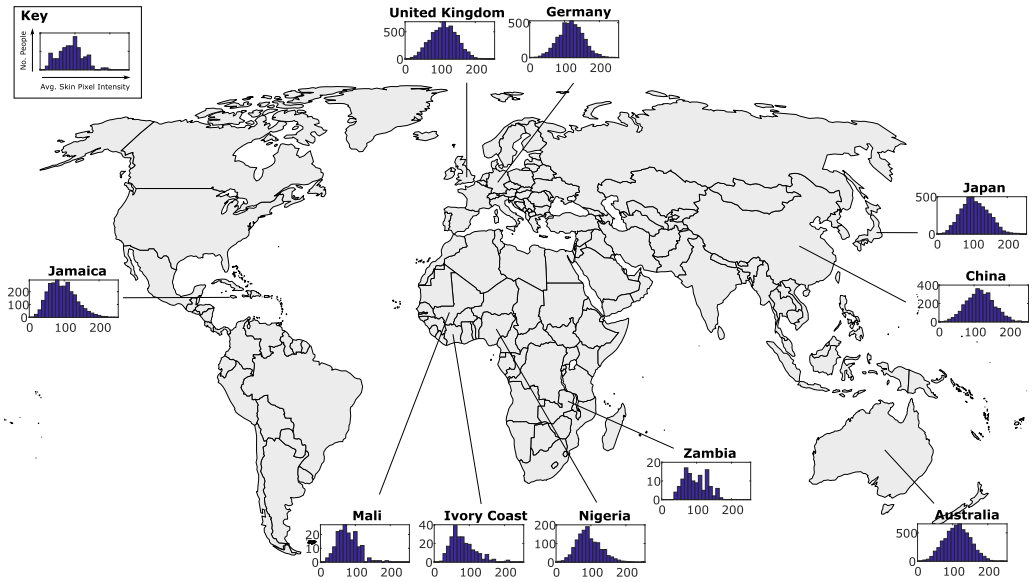

Fig. 4. Histograms of average face skin pixel intensity for RGB photographs of people from several countries around the world. Notice how for Western and East Asian countries average face skin pixels tend to be distributed in the middle of the pixel range, whereas for people from African or Caribbean countries, these distributions are skewed heavily toward zero. The fact that cameras have been optimized to capture lighter skin types has an impact on the performance of camera physiological measurement.

sources without interfering with human vision. However, hemoglobin absorption is weaker in this band compared to that of visible light, and the PPG (or blood absorption signal) will typically have a lower SNR. However, a systematic comparison of RGB and infrared camera measurement of all possible physiological parameters (including motion-based signals like BCG and breathing) has not been performed to our knowledge. Such an analysis would be a valuable contribution. The fact that infrared cameras cannot image colors in the visible range may limit the precision with which changes in motion can be measured. But it is also reasonable to think that motion-based measurement would be less affected than reflectance-based PPG measurement that captures blood volume.

NIR cameras have the distinct advantage of being able to image in low light conditions. Due to this, they are particularly suited for measuring physiological parameters during sleep and at night. As sleep studies, baby monitoring, and driving are all examples of applications that could benefit from non-contact measurement [7, 129, 153, 157], NIR cameras are attractive. Several studies have identified the possibility of detecting the effects of sleep apnea events on both the PPG and breathing signals using these devices.

Another reason that NIR cameras may be attractive for certain applications is that the photometric properties of the skin are not as strong between 700 and 1,000 nm, and therefore differences in performance by skin type might be lower. However, I am not aware of any empirical work that provides a systematic quantitative analysis of this, which is another example of a potentially valuable contribution to the research literature.

3.3 Thermal Cameras

Thermal cameras measure FIR signals, covering wavelengths from 2,000 nm (2 μm) to about 14,000 nm (14 μm). At these wavelengths, most objects even in ambient temperature radiate heat

in the form of thermal emissions. As humans are emissive sources, there is no need for "illumination," either ambient or controlled. Thermal sensors have certain capabilities that are beyond those of RGB and NIR cameras—the most obvious being that FIR signals allow for measurement of human body temperature without contact, something that is not possible with RGB and NIR cameras. However, there are other examples—for instance, imaging of sweat glands to measure transient perspiration and dermal responses is possible [134]. When activated, perspiration pores lower thermal emission, absorbing latent heat, and they appear as colder in the thermal images.

One challenge with thermal sensors is that in the FIR bands, sensor technologies are either based on micro-balometers (black-body adsorption) or materials such as indium gallium arsenide (In-Ga-As) that are much more expensive than silicon, running at several hundreds or even thousands of dollars rather than tens of dollars, which is the case with many RGB cameras. As a consequence, thermal sensors are typically of lower resolution, lower SNR, and much more expensive than RGB or NIR sensors. Some thermal sensors also require cooling, which means that they consume considerably more power than RGB and NIR imagers. Lower cost and more portable thermal cameras have been developed in recent years; these devices initially had very low spatial resolution, but that has also changed. Some "off-the-shelf" thermal cameras are now available for under $100.

3.4 Multi- and Hyper-Spectral Cameras

Studies with RGB, NIR, and FIR cameras are the most common in the literature. However, the design and use of other multi- and hyper-spectral cameras has been explored. As described previously, certain wavebands are better for PPG-derived **pulse rate (PR)** and breathing rate measurement than others, and it has been shown that imaging multiple image bands can improve the robustness of physiological signals [89, 138]. Given these results, one might reason that spectral bands common in most digital imagers may not be optimal in terms of resolution, range, and sensitivity for physiological measurement, especially when it comes to measuring absolute changes in blood composition, transient perspiration, or other physiological meaningful signals.

Although there is potential for performance gains using hyper-spectral imaging, or imaging spectroscopy, the availability of this type of novel sensors is limited in practice due to their cost, mismatched temporal resolution (push-broom vs. snapshot, or global shutter, image acquisition), and lack of ubiquity (in comparison to standard digital cameras in the visible and NIR ranges). Despite these limitations, investigations into multiband imaging, as well as imaging outside of the visible and NIR ranges, could further prove fruitful, and there are several applications of camera physiological measurement in which customized hardware might be appropriate. As an example, a five-band visible band camera, with cyan and orange pixels present (in addition to standard RGB), produced results that outperformed RGB when using the waveband combinations orange, green, and cyan [89]. Future work may seek to explore multiband, and potentially multi-imager, applications that leverage spectrally tuned approaches to integrating multiple wavebands. For example, this could be achieved through the use of specifically tuned optical filters across a small, spatially redundant array of both visible and infrared imagers.

Much of the research into multispectral imaging fuses signals from more than one camera, rather than building a new hardware, each with sensitivity in a different range (e.g., monochrome cameras with filters or NIR and FIR cameras). The predictions from these can then be fused [129] or combined at a feature [55] level to make estimates. He et al. [55] use a calibration process to decompose RGB images into multispectral cubes thereby creating a larger feature space for performing signal separation.

4 PHYSIOLOGIC MEASURES

4.1 Cardiac Measurement

Using cameras, there are several methods of cardiac pulse measurement that have been developed, via PPG, BCG, and the JVP. These are all well-established non-invasive instruments but have traditionally been measured using customized contact hardware. I will discuss each of these in detail in the following, including the pros and cons offered by each approach and how they can be fused together. It should be noted that many of the same techniques for measuring PPG also apply to BCG and the JVP. Therefore, in our discussion of algorithmic approaches to measurement, almost all can be considered as applying to all of the signals.

Photoplethysmography. PPG involves measurement of light transmitted through, or reflected from, the skin and captures changes in the subsurface BVP [5]. Non-contact camera imaging almost always leverages reflectance PPG, as imaging is generally performed from the head or other region of the body through which light will not transmit (unlike the earlobe or fingertip) and there is usually no dedicated illumination source bright enough to transmit through even thin body parts. In videos, the PPG manifests as very subtle pixel color changes of the skin. Sometimes these changes can be "subpixel," meaning that the light changes are less than one bit change per pixel per frame. However, by averaging a region of pixels together, a continuous signal can be recovered. The simplest way to think of this is as the camera as a collection of noisy pixel sensors that when aggregated together reveal the desired continuous signal. There are clinical applications for PPG measurement, as the signal contains information about the health state and risk of cardiovascular diseases [41, 115, 126]. It could be argued that the PPG has been underutilized for clinical applications, and as more methods are developed that make measurement easier, more convenient, and more accurate, we will see greater impact.

Ballistocardiography. BCG involves measurement of the mechanical motion of the body due to the cardiac pulse [139]. Although the PPG and BCG are measured via different mechanisms, they can often both be present in the same video because typically skin regions that feature PPG information also exhibit BCG motions. This can be used to help improve the estimates of downstream metrics such as PR [80]. However, it can also mean that the PPG and BCG signals are difficult to separate in a video [98]. The BCG signal can be measured from video by tracking fiducial/landmark points or via optical flow to capture the subtle motions of the body resulting from the cardiac pulse [11, 132]. The BCG provides complementary information to the PPG signal and could be used to help derive metrics related to **pulse wave velocity (PWV)** or **pulse transit time (PTT)** [132]. However, camera-based BCG measurement methods are highly sensitive to other body motions, making it difficult to design practical applications that leverage this information, except in highly controlled contexts. One advantage of the BCG measurement is that it is not dependent on the presence of skin and therefore can be used to recover cardiac signals from the back of the head or a body part covered in clothing [11].

Jugular Venous Pulse. The JVP is a diagnostic tool used to assess cardiac health. The jugular vein is an extension of the heart's right atrium, and changes in atrial pressure can be reflected in the jugular waveform. The JVP is measured by analyzing the motion of the neck, just below the chin. Specifically, distortions in the JVP waveform morphology can provide non-invasive insight into cardiac function [6]. Using camera methods, primarily capturing motion or optical flow, it has been shown that the JVP can be measured optically [7, 46]. Aiding clinicians in the observation of this signal can help with bedside examinations [2]. Again, the JVP contains complementary information to the PPG and BCG waves, meaning that combining these signals offers opportunities for additional insights into cardiac function. The fact that all three can be captured from the same

sensor data, a video, reduces the complexity of measurement and synchronization of observations. I anticipate that researchers will leverage combinations of these signals to greater effect in the future.

Pulse rate. The PR is the dominant frequency within the PPG, BCG, and JVP waveforms and is typically the simplest information to derive. If the periodic peaks corresponding to the heartbeats are not observable in the waveform, then other metrics will likely be quite difficult to measure. **Heart rate (HR)** is the number of times your heart beats, typically the gold standard for which is the ECG signal. By contrast, PR, as measured via PPG, is the number of times the peripheral BVPs due to an increase in blood flow. I am careful to use the term *pulse rate* here, as currently cameras can only measure peripheral blood flow and body motions. PR will be quite similar to HR in most cases. But in a small number of cases, a heartbeat may be weak and not be capable of pushing blood through the body, making it not possible to observe a pulse on that occasion. As such, PR can be lower than HR.

Looking for the periodic systolic peaks in the video cardiac waveforms is one way to assess their signal quality. Essentially, the BVP SNR proposed by De Haan and Van Leest [39] and frequently used in evaluating video PPG measurement captures that fact. Most camera methods concerned with cardiac measurement have evaluated performance in terms of average or instantaneous PR measurement. Although this is a logical place to start, moving forward I hope that in addition to average PR measurement, increasing emphasis is placed on other metrics as well.

Pulse rate variability. **Pulse rate variability (PRV)** captures the changes in PR over time and is a commonly used measure of **autonomic nervous system (ANS)** activity. **Heart rate variability (HRV)** is closely related to PRV and can in many cases be quite similar [89]. The two branches of the ANS are the sympathetic nervous system and the parasympathetic nervous system, which dynamically control the beat-to-beat differences of the heart. The HRV low-frequency component is modulated by baroreflex activity and contains both sympathetic and parasympathetic activity [4]. The high-frequency component reflects parasympathetic influence on the heart, and it is connected to RSA. An estimate of sympathetic modulation (the sympatho/vagal balance) can be made by considering the low-/high-frequency power ratio. PRV can be computed in several ways but generally requires detecting the pulse inter-beat intervals. The computation can be quite sensitive to the precision of the inter-beat measurement, which presents challenges for computing PRV-derived metrics in many applications. Increasingly, work in camera physiological measurement is being evaluated on the performance of inter-beat measurement rather than average PR [127]; this is an encouraging trend, as it sets a higher bar for algorithmic performance. End-to-end networks could be used to predict peak timings directly from video rather than recovering a waveform and performing peak detection. Sequence-to-sequence models might be quite effective at this task; however, a rigorous evaluation of such an approach for PRV measurement has not been performed.

Breathing. By leveraging RSA, breathing or breathing rates can be derived from the PPG, BCG, or JVP signals by analyzing the high-frequency components of the PRV [121]. However, this is not a perfect method, as some irregular breathing patterns may not be clear within the HRV, and RSA can fluctuate in intensity. When someone is under stress, it can be weaker than when they are at rest. RSA also typically decreases with age and in people suffering from diabetes or who have a pacemaker. Furthermore, since PRV itself is difficult to derive from a noisy cardiac signal, trying to measure the breathing rate via cardiac pulse variability can be unreliable. However, the principle does help motivate why multitask modeling of physiological signals may be a promising direction [76].

Pulse transit time. There are several attractive properties for imaging systems in physiological measurement. One is that imaging systems can measure signals spatially as well as temporally. Another is that cardiac pulse measurements can be made via multiple modalities (e.g., PPG, BCG, and JVP measured simultaneously). Both of these properties enable some promising opportunities for measuring PWV or PTT (i.e., the time it takes for the cardiac pulse to reach a specific part of the body and therefore the velocity of that wave). Researchers have proposed two methods for doing so using imaging systems. The first involves measurement of the PPG signal at two locations on the body from the same video sequence. Shao et al. [133] show that pulse arrival times at the palm and the face can be measured and contrasted. Other work uses the time delay between different cardiac pulse wave (e.g., BCG and PPG) major (or systolic) peaks [132]. As the BCG is similar to the seismocardiogram, it can be combined with the PPG signal to measure PTT. However, ECG reference would still be required to measure pulse arrive time [161]. PWV further requires knowing the distance between points on the body, such as the height of the subject, but it is likely that these measures capture correlated information. Building on this work, it may be possible to measure more dense or continuous spatial variation across the body using imagers rather than just two locations. Several papers have shown examples of such visualizations but have not validated this or compared it to downstream metrics such as PTT or blood pressure.

Arrhythmia. Cardiac arrhythmia, such as **atrial fibrillation (AF)**, is a predictor of serious cardiac events. More than 30% of cardioembolic strokes are directly attributable to AF [173]. Although not all forms of arrhythmia may be detectable via all cardiac signals, AF is possible to identify from the PPG signal. Studies have compared measurement using mobile phone cameras imaging of the fingertip and the face [123, 173]. **Premature ventricular contractions (PVCs)** are another form of arrhythmia that has been studied from contact measurements. Detection of PVCs from the PPG and BCG waveforms is possible, as summarized in the work of Shao et al. [131]. But there is no published camera-based measurement work to our knowledge. Qualitatively, the author has observed PVCs in camera PPG data that were validated by ECG measurements.

Morphological features. Cardiac pulse waves have interesting morphological features. Distortion of the JVP can provide information about cardiac function [6]. In the PPG signal, each pulse wave features a systolic peak and diastolic peak separated by a dicrotic notch or inflection. Fingertip analysis of PPG signals has revealed the promise of these features for downstream assessments [40]. Using these features, metrics such as the left ventricle ejection time (the time between the systolic foot and the dichotic notch) can be derived. For assessing cardiac health, morphological features could be more important than HR or HRV.

However, accurately measuring these subtle waveform dynamics is non-trivial. For example, the dicrotic notch may only manifest as an inflection in the raw PPG wave, but in the second derivative this inflection is a maximum. Computing the second derivative, or acceleration PPG, can be a useful tool for extracting waveform features. The second derivative of the PPG signal can be used as an indicator of arterial stiffness—which itself is an indicator of cardiac disease [62], and similar information contained with the wave can be used to estimate vascular aging [147], which was higher in subjects with a history of diabetes mellitus, hypertension, hypercholesterolemia, and ischemic heart disease compared to age-matched subjects without.

Under controlled conditions, camera algorithms can measure subtle morphological features. McDuff et al. [88] evaluated measurement of systolic-diastolic peak-to-peak time and Hill et al. [58] evaluated left ventricle ejection time. In the latter, it was observed that optimizing for the second derivative error directly, rather computing it from a lower-order prediction, can improve the accuracy of that measure, presumably because the dynamics of the waveform morphology are more faithfully preserved.

Blood pressure. Some of the metrics derived from camera physiological measurements are correlated with blood pressure. However, currently there is little evidence that cameras could be used to directly measure blood pressure. Morphological features in the PPG wave do contain some information about blood pressure. Using a network trained on contact sensor data and then fine-tuning that on rPPG signals, Schrumpt et al. [130] were able to show reasonable blood pressure prediction. Utilizing the spatial measurement opportunities presented by cameras, researchers have shown that non-contact PPG from the face and palm can be used to derive PTT, which has then been correlated with blood pressure [63, 133]. However, these were relatively small studies. A larger study with more than 1,300 subjects found that pulse amplitude, PR, PRV, PTT, pulse shape, and pulse energy features extracted from non-contact PPG measurements could be used to predict systolic pressure and diastolic pressure with reasonable performance [83]. Although these are promising results, their study only features normotensive subjects and not hypertensive or hypotensive patients. Further work is needed to build confidence in the potential of camera measurement of blood pressure, but the opportunities that that would present are obvious, and therefore I expect this to be an area of active research.

4.2 Pulmonary Measurement

There are many parallels in the methods used for pulmonary measurement, as for cardiac measurement. Using cameras, the most obvious method for measuring pulmonary activity is analyzing motion of the body, primarily the torso, mouth, and nostrils. In their simplest form, these algorithms uses pixel averaging to capture changes in luminosity within a video over time. As with cardiac measurement, these naive methods can be improved by segmenting a region of interest rather than using the whole frame, but neither case will typically lead to robust estimates in the presence of other motions or lighting changes. To improve upon this, two forms of motion analysis have been proposed: the first involving tracking fiducial, or landmark, points and the second involving measuring optical flow at a pixel level (i.e., dense flow). There are numerous examples of these approaches applied to camera breathing measurement from RGB [12, 31, 81, 148], NIR [12, 31], and FIR images [70, 81, 114]. The combination of modalities/sensors has also been explored [99] and methods evaluated on more than one modality within the same study [81]. End-to-end supervised neural architectures have also been used. These neural architectures are trained in a similar fashion to cardiac measurement systems with pixels forming the input and a loss computed on the predicted breathing waveform [31]. Solutions that rely on sparse landmark points will be limited in their potential, as inevitably additional information available within the video will be ignored.

Breathing rate. The breathing rate is the dominant frequency within the breathing waveform and, as with HR, is typically the simplest information to derive. Breathing rates of 12 to 20 breaths per minute are normal at rest. Lower breathing rates may be observed during apnea events or exercises such as meditation. Higher breathing rates would generally be observed during physical exercise. If subjects have a chronic respiratory disease, such as asthma [27], it is not uncommon for successive breaths to vary considerably in duration from one another. Irregular breathing may mean that frequency domain analysis of the breathing waveform does not lead to one dominant peak.

Breathing rate variability. Similar to cardiac signals, the variability of breathing rates can be a useful signal about how the body is functioning. Breathing rate variability has not been studied as much as PRV or HRV.

Tidal volume. Tidal volume is the amount of air that moves in or out of the lungs with each respiratory cycle. Measuring this signal involves not only the duration and depth of each breath

but the volume of the chest [70]. This can be simplified as relative tidal volume, which requires only measuring the relative volume changes.

4.3 Electrodermal Measurement

Electrodermal activity is the change in conductance of the skin in response to sweat secretions. Sweat glands are of the order of 0.05 to 0.1 mm and are not visible to the unaided eye, nor can they typically be measured using RGB cameras. Thermal cameras are able to measure sweat gland activity via the changes in thermal emissions. Using this technique, research has revealed how to measure changes in the diameter of the gland in the perinasal region, which can be used to measure transient perspiration [134]. Using RGB cameras, it is certainly possible to measure correlates of electrodermal activity [14]. These could include BVP amplitude and vasomotion. NIR cameras may under some conditions be able to measure moisture on the surface of the body; however, there is little evidence at the moment that this would be effective at capturing a signal that correlates highly with electrodermal activity.

4.4 Blood Oxygen Saturation

The composition of the blood can be measured using cameras with multiple frequency bands, and one can think of this as a low frequency resolution form of spectroscopy. However, because of the broad frequency sensitivity of most RGB cameras, calibration can be challenging. Oxygen saturation, or the ratio between oxygenated and deoxygenated hemoglobin, is the most well studied. In non-contact camera measurement, preliminary studies have validated that oxygen saturation can be captured using RGB cameras [148]. Another method measures the total blood concentration as a function of oxygenated and deoxygenated blood [102]. This method requires calibration using a known color reference. Scattering and specular reflection are both wavelength dependent; therefore, in the presence of head or camera motion, the reflectance of light at different wavelengths will be influenced, making measurement of blood oxygen saturation even more challenging. These challenges will need to be solved before RGB cameras can be used to provide robust measurements.

4.5 Glucose

Given the significance for patients with diabetes, the non-invasive measurement of blood glucose levels is another attractive goal. However, unlike oxygen saturation, the variations in light measured via reflectance methods due to changes in glucose may be quite difficult to detect. According to modeling by Wang et al. [165], it is unlikely to detect blood glucose based on either the DC or AC component of skin-reflected light. Their model captures light in the visible to NIR range. Nevertheless, advances in the spatial, temporal, and sensitivity of imaging hardware plus additional color bands could still present opportunities for non-contact camera glucose measurement.

5 COMPUTATIONAL APPROACHES

The use of ambient illumination means that camera-based measurement is sensitive to environmental differences in the intensity and composition of the light. Camera sensor differences mean that hardware can differ in sensitivity across the frequency spectrum. Automatic camera controls can impact the image pixels before a physiologic processing pipeline, and video compression codecs can further impact pixel values. People (the subjects) exhibit large individual differences in appearance (e.g., skin type, facial hair) and physiology (e.g., pulse dynamics). Finally, contextual differences mean that motions in a video at test time might be different from those seen in the training data.

5.1 Signal Processing Methods

In the context of video-based physiological measurement, traditional signal processing techniques have several advantages. They provide simple-to-implement and often computationally efficient algorithms for the measurement of the underlying physiological signals. They are often also easy to interpret and relatively transparent. Most signal processing methods do not require training data (i.e., are unsupervised), which contributes to their simplicity and interpretability.

Early methods for PPG and breathing measurement leveraged spatial redundancy to cancel out camera quantization noise and recover the underlying waveform [146, 148, 154]. These methods work well on raw videos with limited body motion and homogeneous lighting conditions; however, the presence of motion (either from the camera or subject), illumination changes, video compression artifacts, and other sources of noise can easily corrupt the measurements. To address this, researchers proposed using blind-source signal separation techniques such as **independent component analysis (ICA)** [121, 122] and **principal component analysis (PCA)** [69, 166]. These are simple unsupervised learning techniques that can recover demixing matrices (usually linear) and optimize for certain signal properties. Typically, the demixing is performed frequently (i.e., every 30-second time window) so that the algorithm can adapt to changes in the video over time. In the case of ICA, this optimization is typically performed by maximizing the non-Gaussianity of the recovered signals. Blind-source signal separation methods often work effectively at removing noise from the waveform when it is small in amplitude or relatively periodic. However, they make naive assumptions about the properties of the underlying waveforms. Given that we have prior knowledge about the physical and optical properties of the material (skin) and the physiological waveform dynamics, it is reasonable to believe that we could leverage those to improve our signal estimates, and indeed this is what has been shown.

Chromiance-based methods [38, 39] are such an example, and these are designed with the aim of eliminating specular reflections by using specifically tuned color differences. Building two orthogonal chrominance signals from the original RGB signals (specifically, $X = R - G$ and $Y = 0.5R + 0.5G - B$) helps improve the PPG SNR. Of course, these are specific to the measurement of absorption changes and not body motions. Wang et al. [163] proposed another physically grounded demixing approach based on defining a plane orthogonal to the color space of the skin (POS), which is one of the most robust signal processing methods for PPG recovery. Another physiologically grounded approach used a physical synthetic skin model for learning demixing parameters using Monte Carlo methods [102]. Pilz et al. [119] used principles of local group invariance and then built upon this approach [118] to define a lower, or compressed, dimensional embedding of the pixel space that performed competitively for PPG signal recovery. One attractive property of these demixing and group invariance methods is that they can be quite fast to compute at test time or runtime.

For motion-based signal recovery, similar approaches have been applied. Balakrishnan et al. [11] used feature tracking to form a set of temporal signals and then PCA to recover the BCG signal. This method was adopted to compute the velocity and acceleration BCG signals in other work [132]. Hernandez et al. [56] used a similar approach applied to ego-centric videos, where the landmark tracking was applied to objects in the environment rather than points on the head or body.

For breathing, similar signal processing methods have been adopted using PCA and ICA [64]. Other related work used auto-regressive filters, averaging pixels as the first step and then performing pole selection from the auto-regressive filter model fit to the temporal pixel average signal [148]. Still another method [12] averaged pixels on only one axis (vertical) to create a 1D representation, filtering that representation and then performing correlations of these vectors across frames within a video.

Given the periodic nature of the cardiac pulse and breathing signals, filtering can significantly improve the SNR and downstream metrics. Many methods apply bandpass filtering using a

Hamming window, whereas others use methods such as continuous wavelet filtering [23]. To fairly compare computational methods, it is vital to ensure that filtering parameters are kept constant; unfortunately, there are numerous cases in the published literature in which filter cut-offs, order, and window types are not reported. For a given dataset, results can be significantly improved by tuning filter parameters, but that does not capture the performance of the underlying signal recovery algorithm.

All of these signal processing approaches have similar pitfalls. They often struggle to effectively separate noise from different sources and in most cases ignore a lot of spatial and color space information by aggressively averaging pixels early in the processing pipeline or computing the positions of a sparse set of spatial landmarks. It would seem that more complex temporal-spatial and color space representations would yield signals that more faithfully reflect the underlying physiological process—this is where supervised learning and deep neural models can offer advantages.

5.2 Supervised Learning

Convolutional models. Convolutional networks are the most common form of supervised learning used for camera physiological measurement. These networks learn representations using convolutional filters applied spatially or spatio-temporally to the input frames. DeepPhys [31] was the first to propose a convolutional attention network architecture trained using a combination of appearance frames and motion (difference) frames for physiological measurement. The two representations were processed by parallel branches with the appearance branch guiding the motion branch via a gated attention mechanism. The target signal was the first differential of the PPG wave. Špetlík et al. [137] also proposed a two-part network, but in this case the networks were applied sequentially with an "extractor" network learning representations that were then input to an "HR prediction" network. Loss was computed on the HR estimates. Liu et al. [76] extended the convolutional attention network model to include multitask prediction of both the PPG and breathing wave, thereby effectively halving the computational cost of using two networks with little reduction in accuracy.

For PPG estimation, spatial attention mechanisms essentially act as skin segmentation maps, perhaps learning to weight areas of skin with higher perfusion more heavily, although this has only been validated qualitatively. Chaichulee et al. [29] explicitly modeled skin segmentation in their network architecture before extracting the PPG and respiration signals. For breathing, the skin region may or may not be the best source of information, as in many applications the chest may be the strongest source of breathing motions but may be covered with clothing.

Using the ability of a convolutional network to perform video enhancement, essentially to remove noise, Yu et al. [175] proposed a two-stage process: the first is an encoder-decoder used to enhance the video, removing artifacts and noise, and the second is a PPG extraction network. Another method that has achieved strong results uses a different form of preprocessing. Niu et al. [104, 105] form spatio-temporal maps by computing average pixel intensities from different regions of the face and different color spaces (RGB and CYK). A convolutional network is then trained with these maps as input and the HR as the target. By preprocessing the signal in this way, the designer can incorporate prior knowledge about the spatial and color space properties of the desired signal. The trade-off is the additional computational and implementation costs that are incurred.

Given the characteristic morphology and periodicity of many physiological signals, sequence learning (e.g., via an LSTM or RNN) can help remove noise from predicted waveforms [58, 68, 79, 108, 175]. Yu et al. [175] compared a 3D-CNN architecture with a 2D-CNN + RNN architecture, finding that a 3D-CNN version was able to achieve superior PR prediction errors—suggesting that spatial-temporal modeling is more effective when information can be shared. Liu et al. [76] found

3D-CNNs to be a good solution in terms of accuracy but with a large computational overhead. Nowara et al. [108] used the inverse of an attention mask to compute a noise estimate that was also provided as input to the sequence learning step, and this noise prior helped improve PPG estimates in the presence of motions.

Researchers have attempted to build multitask models that predict cardiac and pulmonary signals [76], but while there is certainly redundancy in the representations learned that can help reduce the computational demands of running multiple models in parallel, accuracy of measurement did not improve.

Transformers. Transformers are becoming the architecture of choice for many computer vision tasks. They offer attractive trade-offs between computation and scalability with training sets. By avoiding computationally expensive convolution operations and leveraging attention mechanisms heavily, they are able to often provide a good balance between accuracy and efficiency. Preliminary work in camera physiological measurement has shown that these architectures are competitive with the state-of-the-art convolutional networks [77], but it is unclear whether with larger datasets it will be possible to exceed the performance of those convolutional baselines. Transformers have also been applied with some success for breathing measurement [66], but both of these works are early investigations and more experimentation is needed.

Support vector machines. A small number of other supervised methods have been proposed, such as using support vector machines [112]. However, they are relatively few and far between. With the dominance of neural models and importance of attention mechanisms in this task, we might infer that these other methods would be unlikely to exceed state-of-the-art performance.

5.3 Unsupervised Learning

Generative adversarial networks. Other methods have used generative adversarial networks to train models to generate realistic PPG waveforms. PulseGAN [136] is one such example, in which the authors used a chromiance signal as an intermediate representation during the training process. The DualGAN [82] method involves segmentation of multiple facial regions of interest using a set of facial landmarks. These regions of interest are then spatially averaged and transformed into both RGB and YUV color spaces. Using these data, spatio-temporal maps are constructed that form the input to a convolutional network. This method produces strong results thanks to careful segmentation and the ability to leverage multiple color space representations. However, these preprocessing steps are certainly non-trivial to implement and come at a significant computational cost [77].

Contrastive learning. Training with unlabeled videos is highly attractive in a domain of camera physiological measurement as well synchronized datasets with videos, and ground truth signals are difficult to obtain. Research has shown that training in an unsupervised fashion can be successful [51, 78]. Contrastive learning is one tool that can be used for learning from unlabeled videos. Gideon and Stent [51] present a clever self-supervised contrastive learning approach in which they resample videos using video interpolation to create positive and negative pairs. Positive pairs have a matching HR and negative pairs have a different HR as a result of the resampling. This model can then be fine-tuned in a supervised manner on a smaller dataset. This approach achieved strong results, obtaining the best performance on the Vision4Vitals Challenge [50].

In other domains of computer vision, pre-trained models have proven to be very powerful tools for many downstream tasks. They can be particularly effective when there are limited numbers of training samples for that downstream task. In the domain of camera physiological measurement, there do not currently exist any public or published models trained on very large scale data. I believe that such a set of models would be quite valuable for the community, and contrastive learning could be one approach to creating them.

5.4 Loss Functions

In the design of supervised models, the loss function used is important, as it defines what will be optimized for in the learning process. In physiological sensing models, there are typically two categories of loss function: waveform losses and metric losses. Waveform losses involve computing the error between a predicted and gold standard physiologic (e.g., cardiac or breathing) waveform, which could typically be computed for every frame. Metric losses involve computing the error between a predicted metric, such as heart *rate* or breathing *rate*, and the gold standard. This would apply for a window of time (e.g., at least one beat or breath). Because synchronization of data at the waveform level might be difficult (involving millisecond precision), often optimization is performed at the metric level, for which synchronization need not be as precise. However, the relative frequency of feedback—once per time period versus once per frame—is lower, which could impact the learning rate and the amount of training samples needed. Future work could compare these two to determine if optimization at the metric level leads to inferior recovered waveforms or conversely that it leads to more precise downstream metrics.

5.5 Meta-Learning

Given the high individual variability in both visual appearance and physiological signals, personalization or customization of models becomes attractive. Several meta-learning techniques have been proposed for camera physiological measurement. Meta-RPPG [68] was the first such approach that focuses on using transductive inference-based meta-learning. Liu et al. [78] proposed a meta-learning framework built on top of the convolutional architecture presented previously [76]. They leveraged model-agnostic meta-learning [44] and tested both unsupervised and supervised model adaptation. The unsupervised method used pseudo PPG labels generated using the plane orthogonal to the skin (POS) method [163]. Meta-learning and model personalization should receive growing interest moving forward, as it becomes possible to customize models more easily on devices.

5.6 Super Resolution and Video Enhancement

Several methods have leveraged super resolution as a means of improving the extracted physiological waveforms, especially from low-resolution input images. McDuff [84] showed that super resolution could help improve waveforms extracted from frames with resolution as low as 41×30 pixels. In this case, a neural super resolution was paired with a traditional signal processing step to extract the PPG signal. Yue et al. [176] combined a neural super resolution step with a neural PPG extraction step to create a fully supervised example. As described earlier, neural approaches have been used to enhance videos before recovering the PPG signal. Spatio-temporal video enhancement not only combats low spatial resolution but also the effects of video compression. The video enhancement network can be trained in a self-supervised manner without requiring physiologic labels and then a subsequent network fine-tuned to recover the signal itself [175].

6 MAGNIFICATION AND VISUALIZATION OF PHYSIOLOGICAL SIGNALS

Camera physiological measurement enables certain opportunities that traditional sensors do not have. Video magnification is an area of computational photography with the goal of magnifying changes of interest in a video. Magnification is helpful in cases where changes are subtle and difficult to see with the unaided eye. One application that has been used quite frequently in this field is magnification of physiological changes in a video. Early video magnification methods used Lagrangian approaches that involve estimation of motion trajectories (e.g., the motion of the chest when someone is breathing) that are then amplified [74, 162]. However, these approaches are

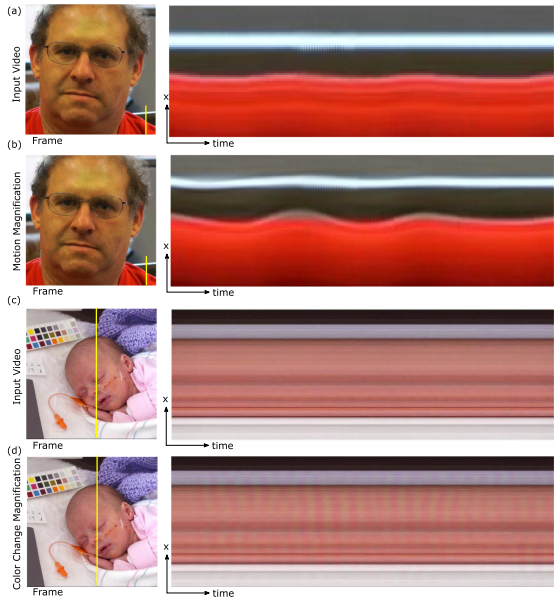

Fig. 5. Examples of video magnification of physiological signals. Scan lines for the motion (breathing) magnification method applied to the "head" video and color change (pulse) magnification applied to the "baby2" video from Wu et al. [170].

often complex to implement in practice. The neat **Eulerian video magnification (EVM)** approach proposed by Wu et al. [170] has had a significant impact on the field and raised the profile of video magnification as a whole (Figure 5). This method combines spatial decomposition with temporal filtering to reveal time varying signals without estimating motion trajectories. One drawback is that it uses linear magnification that only allows for relatively small magnifications at high spatial frequencies and cannot handle spatially variant magnification. To counter the limitation, Wadhwa et al. [158] proposed a non-linear phase-based approach, magnifying phase variations of a complex steerable pyramid over time. In general, the linear EVM technique is better at magnifying small color changes (i.e., more suitable for PPG), whereas the phase-based pipeline is better at magnifying subtle motions (i.e., more suitable for respiration, BCG, or JVP). Both the EVM and the phase-EVM techniques rely on handcrafted motion representations. To optimize the representation construction process, a supervised neural learning based method [111] was proposed, which uses a convolutional network for frame encoding and decoding. With the learned motion representation, fewer ringing artifacts and better noise characteristics have been achieved. In preliminary work, Pintea and van Gemert [120] propose the use of phase-based motion representations in a learning framework that can be applied to the transference (or magnification) of motion.

One common problem with all of the preceding methods is that they are limited to stationary subjects in which the physiological signal of interest is at another frequency (usually significantly faster) than other changes (e.g., body or camera motions), whereas many realistic physiological sensing applications would involve small changes of interest in the presence of large ones that might be at similar frequencies. For example, body motions might be at a similar frequency to the HR or breathing rate. After motion magnification, these large motions would result in large artifacts and overwhelm any smaller variations. A couple of improvements have been proposed, including a clever layer-based approach called *DVMAG* [42]. By using matting, it can amplify only a specific region of interest while maintaining the quality of nearby regions of the image. However,

the approach relies on 2D warping (either affine or translation only) to discount large motions, so it is only good at diminishing the impact of motions parallel to the camera plane and cannot deal with more complex 3D motions such as human head rotation. The other method addressing large motion interference is video acceleration magnification [177]. It assumes large motions to be linear on the temporal scale so that magnifying the motion acceleration via a second-order derivative filter will only affect small non-linear motions.

Another problem with the previous motion magnification methods is that they use frequency properties to separate target signals from noise, so they typically require the frequency of interest to be known *a priori* for the best results and, as such, have at least three parameters (the frequency bounds and a magnification factor) that need to be tuned. If there are motion signals from different sources that are at similar frequencies, it is previously not possible to isolate the different signals. Chen and McDuff [32] presented a supervised learning approach that enables learning of source signals using gradient descent and then magnification using gradient ascent.

An example of the clinical utility of magnifying physiological signals using camera measurement was provided by Abnousi et al. [2]. They used EVM to amplify videos of patients' necks and found that agreement between clinicians in the bedside assessment of the JVP was greater in the magnified condition compared to the unmagnified one. They argued that this technology could help expand the capabilities of telehealth systems.

7 CLINICAL APPLICATIONS AND VALIDATION

Neonatal monitoring. Neonates in intensive care require constant monitoring and are also active with the clinical staff interacting with them regularly [29]. The attachment of sensors can damage the skin and increase the risk of developing an infection or simply disrupt the sleep or comfort of an infant. Camera physiological measurement seems particularly well suited to this context. Numerous preliminary clinical validations studies have been conducted to access the readiness of these tools for monitoring neonates [1, 19, 29, 49, 81, 96, 114, 156]. Although the infants can and do move, they are relatively immobile (i.e., are laying down in a small incubator). Furthermore, illumination in a hospital environment can be controlled somewhat carefully. All in all, this is a promising application in which we might expect some degree of success. These initial validation studies have obtained promising results; however, further research is still needed to build confidence in the technology. There are opportunities in this context to fuse signals from multiple sensors, such as pressure-sensitive mats [81], which could offer additional benefits or help address some of the challenges of camera-based sensing, such as measurement when the body is obscured by blankets.

Kidney dialysis. Validation of camera physiological measurements has also been performed in other clinical contexts. Specifically, Tarassenko et al. [148] conducted experimentation to validate measurements on adult kidney dialysis patients. In this example, which was part of a larger-scale clinical study, 46 patients had their vital signs monitored during 133 dialysis sessions. The advantages of camera sensing in this context are similar to those in the neonatal context. Removing the need for contact sensors could increase the comfort of the subjects, helping them sleep and move more easily.

Telehealth. One natural application of these technologies is in telehealth, where platforms for video conferencing are used for remote patient care. The COVID-19 pandemic has highlighted the need for remote tools for measuring physiological states. With large numbers of telehealth visits being conducted over video conferencing platforms [9], there is still no scalable substitute for the measurements that would have traditionally been recorded at a doctor's office. Therefore, computer visions tools for physiological measurement are attractive and becoming increasingly

important [48, 128]. However, to our knowledge, there are no published results from clinical validation studies using camera physiological measurement in this context. One challenge with these studies, unlike those performed in hospitals, is how to collect gold standard sensor data while at the same time capturing videos that exhibit the natural variability that would be observed with patients joining from their home or another location. It will certainly require a great amount of work to achieve this, but the potential benefits are significant.

Sleep monitoring. Sleep studies are an important tool in diagnosing sleep disorders. **Polysomnography (PSG)** is the measurement of sleep via physiological sensing. However, the current PSG systems are cumbersome, disrupt sleep, and require specialist equipment available only at sleep laboratories. Using NIR cameras, several proof-of-concept studies have been performed demonstrating measurement of PPG [7, 157], blood oxygen saturation [153, 157], and breathing [129, 153] (Scebba et al. [129] combined NIR and FIR cameras). One study using camera vitals for sleep monitoring achieved pulse and respiratory rate detection within 2 beats/breaths per minute in more than 90% of samples and 4 percentage points error in blood oxygen saturation in 89% of samples [153]. One sleeping disorder that PSG can help identify/diagnose is sleep apnea. Amelard et al. [7] used a camera system to measure PPG and found that pulse wave amplitude decreased during obstructed breathing and recovered after inhalation with a temporal phase delay. This early study provides encouraging evidence of the potential of video measurement during sleep. Camera systems are much easier to deploy and scale in homes than the equipement currently used for PSG.

Health screening. Thermal imaging has been used for health screening at health clinics and airports for several years. Screening in this way can help limit the spread of infectious diseases and protect other people, including healthcare providers. Typically, these system measure body temperature. The limited availability of thermal cameras means that such systems cannot be deployed in every context. RGB and NIR imaging have some potential utility here; however, it is unclear if these sensors alone would be sufficient or whether they offer additional utility to thermal cameras [141].

8 NON-CLINICAL APPLICATIONS

Baby monitoring. Outside of the clinical domain, consumer baby monitors are another set of products that can leverage camera physiological measurement. Similar arguments for camera measurement apply in consumer products as in the NICU applications (i.e., less disruption to the babies' sleep and decreased risk of irritation or damage to the skin). Baby monitors that offer optical breathing measurement are already commercially available (e.g., MikuCare[3]). It is likely that the next generation of these devices will try to integrate HR measurement. Blood oxygen saturation would probably be the next most likely signal. The role of physiological sensing for infants using consumer devices has been questioned, and some argue that it could lead to increased anxiety about what these data mean.[4] As with all technologies, there is a need for user-centered design, demonstrating that the sensing and user interface solve a clear need for the consumer and minimize potential harms.

Driver monitoring. Medical conditions lead to a significant number of traffic accidents and deaths every year. In many cases, the drivers may not be aware of these underlying medical conditions [159]. Therefore, in-cabin cardiac monitoring is of interest to auto manufacturers. In-vehicle measurement could be used to help detect cardiac events (e.g., stroke) and use this information

[3] https://mikucare.com/.
[4] https://mashable.com/2017/02/18/raybaby-baby-breathing-monitor/.

to stop a vehicle safely, thus helping to prevent accidents, or for offering health monitoring as an attractive feature for customers. Demonstrations of camera physiological sensing can be found, but to our knowledge no vehicles currently on the market offer this technology. Mitsubishi Electric Research Labs (MERL) [106] and Toyota [51] have both published research on camera physiological measurement in vehicles, and this illustrates active interest from the automotive sector in these tools. Signal processing [106] and neural [51, 169] approaches have been proposed.

Biometrics. Outside of the clinical realm or applications that focus on consumer health, camera physiological measurement has received growing attention for detecting fake videos and verifying the "liveness" of a subject. Face verification tools could be fooled by a picture or a mask; however, it is quite difficult to spoof subtle physiological changes in those cases. Researchers have leveraged this to detect deep fake videos [36, 124] and propose anti-spoofing systems [75, 79]. For the latter, it is unclear if in practice these approaches would work effectively, as the recovered signals can easily be corrupted. For example, it would not be hard to introduce a periodic change that is then picked up by the camera, and whether an imaging algorithm could determine a real versus fake period change is untested. Furthermore, the motion of a subject in front of a camera or heavy makeup may obstruct measurement of the PPG signal, entirely making it appear as though a heartbeat is not present when in fact it is.

Affective measurement. The field of affective computing [117] studies technology that measures, models, and responds to affective cues. Physiological signals contain information about ANS activity. There are many other areas in which unobtrusive physiological sensing could help advance the vision of affective computing. Enabling measurement via ubiquitous devices increases opportunities to study affective signals in situ and build systems that have the potential to be deployed at scale in the real world. Two areas in which camera physiological measurement have been employed are the detection of stress and cognitive load [24, 90] and the measurement of responses to digital content [26, 116]. To build systems that respond to affective signals, researchers have developed camera sensing of parameters closely related to sympathetic and parasympathetic nervous system activity. From cardiac signals, HRV or PRV has been used as a measure to quantify changes in cognitive load or stress [24, 90]. The PPG signal contains several additional sources of information about ANS activity. BVP and vasomotion change in amplitude during stressful episodes [24, 93].

9 REGULATORY APPROVAL

Although the field is still relatively young, several products have already received regulatory approval. Oxehealth and ContinUse Biometrics/Donisi Health have both received De Novo classification for a device that uses software algorithms to analyze video and estimate PR, HR, respiratory rate, and/or breathing rates.[5, 6] These are examples of pioneering companies in the space of camera physiological measurement and are navigating relatively uncharted waters. Regulations can vary widely across the world. In this case, I will take the United States as an example for the purposes of discussion, in part because it is one of the largest markets for healthcare technologies. In the United States, usually regulatory approval is granted based on comparison with a current legally marketed device to base a "determination of substantial equivalence".[7] However, in the case of camera measurement systems, several companies have sort De Novo classification. De Novo clearance provides a marketing pathway to classify a novel medical device with a reasonable assurance of

[5] https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN200019.pdf.
[6] https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN200038.pdf.
[7] https://www.fda.gov/medical-devices/premarket-submissions-selecting-and-preparing-correct-submission/de-novo-classification-request.

safety and effectiveness for the intended use, but for which there is no legally marketed predicate device. The **U.S. Food and Drug Administration (FDA)** also recently granted market clearance for FibriCheck, a Belgium-based company whose mobile health app uses a smartphone camera or smartwatch sensors to measure a person's heartbeat.[8] Devices like the Miku baby monitor are not currently FDA approved, and although it provides feedback on a baby's breathing, the Miku Care website states that it "is not intended to diagnose, cure, treat, alleviate or prevent any disease or health condition or investigate, replace or modify any physiological process."

Challenges arise with regulatory approval when designing software-based camera measurement that will deployed on a wide variety of devices (e.g., smartphones), as it is difficult to quantify the performance on all of the different hardware configurations (cameras, lens, imaging firmware, camera controls, etc.) In most cases, it becomes necessary to limit the hardware range to only smartphones of a particular generation. Furthermore, non-contact imaging methods often aim to use ambient illumination. However, without a dedicated light source, it can be difficult to show that performance will be similarly accurate in a large range of contexts. Smartwatches, in contrast, have dedicated light-emitting diodes and are close to the skin, limiting the amount of light from the environment that might interfere with measurement. As camera physiological measurement becomes more mature, clearer precedents will likely be set for the regulatory process for these systems.

10 CHALLENGES

10.1 Fairness

In camera physiological measurement, appearance of the body or the environment is a key factor. Skin type, facial structure, facial hair, and clothing and other apparel can all affect the performance of measurement systems, as can lighting conditions.

Hardware. Starting with the hardware, all cameras are designed with certain operating criteria. Given the nature of the markets in which they are sold, these cameras have often been optimized to capture lighter skin types more effectively than dark skin types. This can introduce an inherent bias in performance even if the algorithm and training data used do not. Typically, sensitivity is greatest toward the middle of the camera's frequency range. Dark or very light skin types could be more likely to saturate the pixels, and changes due to physiological variations may be lost.

Data. Almost all datasets for camera physiological measurement have been collected in Europe, the United States, or China (see Section 12.1). As such, they predominantly contain images of lighter skin type participants. Furthermore, they generally feature younger people and often have a male bias. One challenge with constructing fair datasets in camera physiological sensing is that even the gold standard contact devices can exhibit biases [16]. Evidence of biases in SpO_2 measurement with skin type is prevalent, with three monitors tested overestimating oxygen saturation in darker skin types in adults [16] and infants [155]. But other sensors (respiration, BCG, etc.) may also introduce biases. For example, chest straps frequently used as a gold standard for measuring breathing may lead to different measurements on women than men. Further characterization of camera *and* gold standard contact devices is needed to avoid errors from propogating, or worse compounding.

Models. The design of models for camera physiological measurement may also encode bias. This type of bias is often more difficult to detect. Several of the signal processing models described in Section 5 contain hard-coded parameters but were evaluated primarily on datasets of

[8] https://www.fibricheck.com/fibricheck-receives-fda-clearance-for-its-digital-heart-rhythm-monitor/.

light skin type subjects. Some initial work has begun to characterize differences in performance of algorithms (both supervised and unsupervised) by different demographic and environmental parameters [3, 107]. From preliminary research to clinical studies and the development of products, I believe that this deserves greater attention. The development of balanced and representative datasets is one example of a significant contribution that could help toward this end. Meritable efforts toward this end have recently been published [30]. Some of this work has also included novel methods for augmenting or simulating data to help address data imbalance among other things [10]. Dasari et al. [37] investigated the estimation biases of camera PPG methods across diverse demographics. As with previous work, they observed similar biases as with contact-based devices and environmental conditions. Chari et al. [30] proposed a physics-driven approach to help mitigate the effects of skin type on PPG measurement with encouraging results. I argue that innovations in hardware, better datasets, and algorithmic contributions can all significantly improve the equitability in performance.

10.2 Motion Tolerance

The effects of subject motion have been among the most studied dynamics in camera physiological measurement. Understandably, much of the early work focused on rigid, stationary subjects [61, 146, 154, 167]. Subsequent studies allowed for limited naturalistic head motions [122], but many experimental protocols still strictly limited the amount of motion during data collection [121, 143]. Under these conditions and with reasonable image settings and illumination, recent methods will typically recover the underlying signals with high precision. For sleep measurement and in certain controlled contexts, these assumptions may not be terribly unrealistic. However, for others, such as consumer fitness applications (e.g., riding a static bike), they would be much too constrained. On this topic, work has examined PR measurement during exercise on five different fitness devices [39]. This signal processing method was constrained by optical properties of the imager and the illumination source. As performance in constrained motion conditions began to saturate, researchers started to investigate algorithm performance under greater motion (both translational [59] and rotational [43]). These approaches range from simple region of interest focused object tracking [174] to projection [12] or signal separation [121] to more complex neural networks [31]. Approaches for estimating the contribution of motion artifacts and correcting the PPG signal using adaptive filtering [28] or denoising networks [108] have also been explored.

Several systematic and carefully controlled subject motion studies have been performed. Estepp et al. [43] focused on rigid head rotations. By combining data from multiple imagers and using blind source separation, they were able to reduce the effects of rigid head motion artifact in the measurement of PR [95]. In one way, multiple imagers simply add additional spatial redundancy, and methods have utilized this reduncancy in a single camera to reduce the impact of motion-induced artifacts including translation, scaling, rotation, and talking [164]. A comparable framework has been extended to include multi-imagers in the infrared spectrum as well [152], and combining RGB and infrared imagers [53, 99, 129]. Another approach is to used motion information extracted via a body, head, or face tracking system to filter or compensate for motion [35].

Many of the aforementioned methods used signal processing approaches without leveraging supervised learning. Neural models have proved highly effective at learning spatial and temporal information. Chen and McDuff [32] illustrated this in the case of video magnification of PPG and breathing by showing how an algorithm could selectively magnify motions and color changes even in the presence of head motions at similar frequencies. All this being said, motion robustness should continue to be a focus in camera physiological measurement. Different types of motion are likely to be observed with different applications, and thus evaluation in the contexts of fitness and

exercise, human computer interaction and video conferencing, baby monitoring, and clinical care would all be quite valuable.

10.3 Ambient Lighting

Ambient lighting conditions impact camera measurement in two primary ways: *composition* and *dynamics*. Composition of the ambient light can impact the performance of computational methods, as absorption characteristics vary by frequency. The qualities and properties of constant illumination used for physiological measurement that are necessary to produce results of adequate quality have been explored [73]. Brighter, green lighting tends to give the strongest improvement in PPG measurement. This is consistent with systematic analyses that have characterized the hemoglobin absorption spectra, although less is known about how light composition might impact motion-based analysis (i.e., breathing or BCG measurement). From a practical perspective, light composition is not only tied to the absorption or reflectance properties of the body but also the image quality. As I shall discuss in the following section, lens aperture and focus, sensor sensitivity (ISO), individual frame exposure time (integration, shutter speed), and other image settings will also impact performance.

If the intensity, position, or direction of lighting changes dynamically, it will typically introduce relatively large pixel changes compared to those resulting from physiological processes. Where ambient lighting conditions can be controlled and/or held relatively constant, it can be extremely advantageous. There are several cases in which this might be true (e.g., an incubator in a hospital or in a gym) and cases in which this almost certainly will not be true (e.g., driving) Changes in illumination intensity affect absolute magnitude of camera-measured PPG waveforms [143], which can in turn impact measures of pulse wave amplitude and blood composition. However, it is still true that the effects are still largely uncharacterized for many physiological signals [87]. Several computational methods have been proposed to help combat lighting effects. Li et al. [72] used an adaptive filtering approach, with an isolated background region of interest serving as the input noise reference signal, to compensate for background illumination. Nowara et al. [108] used a similar concept, leveraging an inverse of the PPG attention mask as the background region. These methods both provided an overall reduction in HR error. However, in the case of Li et al. [72], an ablation study of the components of their multistage processing approach (region of interest detection, illumination variation correction, data pruning, and temporal filtering) was not made available to sufficiently determine the effectiveness of any single stage. Neither did Nowara et al. [108] identify if the inverse attention was primarily addressing illumination changes, body motions, or other sources of noise. Amelard et al. [8] presented results using a temporally coded light source and synchronized camera for PPG measurement in dynamic ambient lighting conditions. Novel hardware presents some interesting opportunities for combating illumination; however, understandably, most work focuses on "off-the-shelf" cameras due to the lower technical barrier and far greater availability of those devices.

To the best of my knowledge, no research in camera physiological measurement has made use of computational color space calibration and white balancing methods [100]. Priors on skin color can help correct color inconsistencies in images [15], and it may be possible then that the inverse could be true—priors on scene color could be used to correct skin pixel color inconsistencies with or across videos. There are also methods for relighting faces, using as little as a single frame [142]. Both of these approaches could be helpful for relighting/altering color profiles at test time to help a model perform more accurately, or augmenting a training set to help build models that generalize better. Of course, these hypotheses need rigorous validation, but there certainly appear to be many tools in computer vision, graphics, and computational photography that could aid in camera physiological measurement.

10.4 Imaging Settings

Although cameras are ubiquitous, they vary considerably in specifications. This is one reason obtaining regulatory approval for camera-based solutions can be challenging. Determining the optimal qualities of an image sensor and characterizing how sensitive measurements are to changes in their parameters is quite valuable. These parameters include sensor type (e.g., CCD, CMOS), color filter array (e.g., Bayer, Fovenon X3, and RGBE), number and specification of frequency bands, bit depth, imager size, and number of pixels. Beyond the image sensor, there are other hardware considerations, such as lens type and quality, spectral properties of the illumination source, and image aperture/shutter speed/ISO. All of these parameters will affect the overall content of any acquired image. Then there are software properties or controls, some of which may be constrained by the hardware and others by the bandwidth or storage capabilities. These include the resolution of the video frames; the frame rate at which the video is captured; and whether white balancing, auto focus, or brightness controls are enabled and dynamically changing during video capture. Given the myriad combinations here, it is understandable that it is difficult to precisely characterize the impact of each. Needless to say, sensor quality, resolution, and frame rate all play a particular role. It may be possible to use intuition to help guide some of these judgments—for example, shutter times should avoid pixel saturation [151].

It is also important to consider the apparatus/equipment setup and context with impact signal quality for a given hardware and software configuration. The distance of the body region of interest from the camera will pack the pixel density and the additional contextual information that might be available from the image. Some datasets have characterized the face region of interest pixel density [65], and we recommend that future datasets do the same. Placing a camera very close to the body might lead to a higher number of pixels containing the signal of interest, but also potentially mean that information about other related signals, or context (e.g., body motions/activities), is lost.

Some studies exist on the comparison of multiple imagers running in parallel during data collection (e.g., [103, 143]) and offer some confidence that signal recovery can be robust over widely varying imager properties. Studies of image size (pixel density) and frame rate in single-imager [144] and multi-imager [43, 53, 129] sensor designs have shown that, as expected, these parameters do impact the performance of PPG measurement. Breathing measurement is impacted more significantly by pixel density, which is why Chen and McDuff [32] used a higher resolution for breathing model magnification than PPG magnification.

As with many of the topics discussed in this section, a great contribution would be the creation and standardization of an explicit benchmark test, and related metrics, that could be performed with a variety of imagers to better understand and compare results across studies and methods. The VIPL dataset [103] is the closest example of such a dataset (and will be described in detail in Section 12.1).

10.5 Video Compression

Video compression algorithms are designed to reduce the total number of bits needed to represent a video sequence. These algorithms have been traditionally designed to preserve video quality, characterized by scores related to human perceptual quality, such as minimizing motion artifacts and loss of clarity. Compression algorithms have not been designed directly, or indeed indirectly in most cases, for preserving physiological information with a video. Compression can impact measurements that rely on motion (e.g., breathing) less than those that rely on color changes (e.g., PPG) [110] but will to some degree impact both. The subtle changes in pixel values that are used to recover the PPG signal are often imperceptible to the eye, and these small temporal variations are often removed by compression algorithms to reduce the overall bitrate. Previous work in

systematically analyzing the impact a video compression on PPG measurement [86, 125] found a linear decrease in the PPG SNR with increasing constant rate compression factors. However, Rapczynski et al. [125] observed that performance of HR estimation was less sensitive to decreases in resolution and color subsampling, both of which can be used to reduce video bitrates.

There are other ways to reduce the impact of compression on physiological signals within a video, such as by training a model on videos at the same compression level [110]. Supervised models can learn to reverse or ignore compression artifacts to some degree. Given that video compression is necessary for many applications (i.e., in cloud-based teleconferencing systems), this insight may prove useful. It would certainly be impractical with current bandwidth limits to stream raw video at scale. Recovering the signals from heavily compressed videos is something that deserves further attention. Yu et al. [175] designed a video enhancement model that could serve this purpose and be trained in a self-supervised manner. Datasets with varying levels of video compression are somewhat easy to create, and I argue that standard versions of all public datasets could and should be created with multiple video compression levels so that researchers can report results across different compression rate factors. Zhao et al. [179] proposed such a benchmark dataset; however, access to that data is unclear.

ETHICS AND PRIVACY IMPLICATIONS

The many positive applications of the measurement of physiological signals using cameras illustrates that this technology has great potential. However, there are important risks to consider and potential mitigations that can be put in place to minimize the impact of these risks. Cameras are an unobtrusive and ubiquitous form of sensor, which are used for surveillance at scale. Using similar methods to those described for monitoring patients in intensive care, a "bad actor" could employ these tools for surveilling people. Cameras could be used to measure personal physiological information without the knowledge of the subject. Military or law enforcement bodies may try to apply this in an attempt to detect individuals who appear "nervous" via signals such as an elevated HR or irregular breathing. Or an employer may surreptitiously screen prospective employees for health conditions without their knowledge during an interview (e.g., heart arrhythmias or high blood pressure). Some may attempt to justify these applications by claiming that monitoring could also be used to screen for symptoms of a virus during a pandemic or to promote public safety.

There are several reasons that this would be irresponsible and harmful. First, there is little evidence that physiological signals would provide enough information, without additional context, for determining emotional states or job eligibility. Second, camera physiological measurement still requires significant validation in real-world settings, and it is unlikely that the current state-of-the-art camera physiological measurement systems would be accurate enough in these contexts. As described in Section 10.1, there is evidence that they currently do not perform with equal accuracy across people of all appearances and in all contexts. The populations that are subject to the worst accuracy might also be those that are already subject to disproportionate targeting and systematic negative biases [47]. Many of these issues have been discussed in the context of facial recognition, but parallels can be drawn with physiological measurement. Third, there are many possible negative social outcomes that might result even if measurement was "accurate." Normalizing covert surveillance of this kind can be dangerous and counterproductive.

As with any new technology, it is important to consider how camera physiological measurement could be applied in a negligent or irresponsible manner, whether by individuals or organizations. Application without sufficient forethought for the implications could undermine the positive applications of these methods and increase the likelihood that the public will mistrust the tools.

These applications would set a dangerous precedent and would probably be illegal. Just as is the case with traditional contact sensors, it must be made quite transparent when camera-based

physiological measurement is being used and subjects should be required to consent to data being collected. There should be no penalty for individuals who decline to be measured. Ubiquitous sensing offers the ability to measure signals in more contexts, but that does not mean that this should necessarily be acceptable. Just because cameras may be able to measure these signals in new context, or with less effort, it does not mean that they should be subject to any less regulation than existing sensors.

Although far from a solution to the challenges described earlier, researchers have proposed innovative methods for removing physiological information from videos [33] and "blocking" video-based measurement.There are also instances of more generic computer vision jamming systems [54, 168, 172] that could apply in the context of camera physiological measurement. However, we should recognize that these solutions often put the onus on the subject to opt-out and could be quite inconvenient and stigmatizing. The emphasis should be on opt-in systems that are used in well-validated and regulated contexts.

11 SOFTWARE

In this section, I highlight some of the repositories of open source code for camera physiological sensing. Unlike other domains in machine learning, there are relatively few complete repositories containing implementations of baseline methods. The research community would do well to address this.

MATLAB. For signal processing analysis, MATLAB has often been a popular language for implementation. McDuff and Blackford [85][9] implemented a set of source separation methods (Green, ICA, CHROM, POS) in MATLAB, and Pilz [118] published the PPGI-Toolbox[10] containing implementations of Green, SSR, POS, Local Group Invariance (LGI), Diffusion Process (DP), and Riemannian-PPGI (SPH) models.

Python. Increasingly, Python is becoming more popular as a language for developing camera physiological measurement methods. There are several implementations of the popular signal processing methods: Bob.rrpg.base[11] includes implementations of CHROM and SSR, and Li et al. [72] and Boccignone et al. [22] released code for Green, CHROM, ICA, LGI, PBV, PCA, POS, and SSR. Several published papers have included links to code; however, often this is only inference code and not training code.

To date, there are very few code bases that provide implementations of multiple supervised neural models, despite these being the best-performing methods. Researchers have released code for their own methods, often accompanying papers; however, a unified code base or toolbox is not available.

12 DATA

12.1 Public Datasets

Public datasets serve two important purposes for the research community (Table 1). First, they provide access to data to researchers who many not have the means to collect their own, lowering the bar to entry. Second, they provide a transparent testing set to fairly compare computational methods and set benchmarks. Descriptions of benchmark datasets should include details of the imaging device, lighting, and participant demographic information, in addition to videos and gold standard contact measurements.

[9] https://github.com/danmcduff/iphys-toolbox.
[10] https://github.com/partofthestars/PPGI-Toolbox.
[11] https://pypi.org/project/bob.rppg.base/.

Table 1. Summary of Imaging Datasets

Dataset	Context	Subjects	Videos	Imaging	Gold Standard	Other Comments
MAHNOB-HCI [135]	Implicit media tagging	27	527	Resolution: 780 × 580 Frame rate: 61 Hz	ECG, EEG, breathing	Videos are quite heavily compressed.
BP4D [178]	Multimodal affect analysis	140	1,400	Resolution: 1040 × 1392 Frame rate: 24 Hz	Blood pressure wave	There is no PPG gold standard but continuous fingertip blood pressure.
VIPL-HR [103]	Camera physiology	107	3,130	Resolution: 960 × 720/640 × 480 Frame rate: 25 Hz/20 Hz	PPG, HR, SpO_2	2,378 RGB and 752 NIR
COHFACE [57]	Camera physiology	40	160	Resolution: 640 × 480 Frame rate: 20 Hz	PPG	
UBFC-RPPG [21]	Camera physiology	42	42	Resolution: 640 × 480 Frame rate: 30 Hz	PPG, PR	
UBFC-PHYS [97]	Camera physiology	56	168	Resolution: 1024 × 1024 Frame rate: 35 Hz	PPG, EDA	Contact measurements were obtained using a wrist-worn device, and therefore the signal quality is variable.
RICE CameraHRV [113]	Camera physiology	12	60	Resolution: 1920× 1200 Frame rate: 30 Hz	PPG	
MR-NIRP [106]	Camera physiology	18	37	Resolution: 640 × 640 Frame rate: 30 Hz	PPG	It contains videos recorded during driving.
PURE [140]	Camera physiology	10	59	Resolution: 640 × 480 Frame rate: 60 Hz	PPG, SpO_2	
rPPG [65]	Camera physiology	8	52	Resolution: 1920 × 1080/640 × 480 Frame rate: 15 Hz	PR, SpO_2	
OBF [71]	Camera physiology	106	212	Resolution: 1920 × 1080 (RGB) Frame rate: 60 Hz	PPG, ECG, RR	It contains a small subset of subjects with atrial fibrillation.
PFF [59]	Camera physiology	13	85	Resolution: 1280 × 720 Frame rate: 50 Hz	PR	
VicarPPG [149]	Camera physiology	20	10	Resolution: 720 × 1280 Frame rate: 30 Hz	PPG	It features videos recorded after exercise.
VicarPPG-2/CleanerPPG [52]	Camera physiology	10	40	Resolution: 1280 × 720 Frame rate: 60 Hz	PPG, ECG	CleanerPPG is a meta-dataset associated with VicarPPG-2.
CMU [37]	Camera physiology	140	140	Resolution: 25 × 25 Frame rate: 15 Hz	PR	Videos are anonymized and only show skin regions on the forehead and cheeks, not the full face.
Scamps [94]	Camera physiology	2,800	2,800	Resolution: 320 × 240 Frame rate: 30 Hz	PR, BR, PPG, breathing	Videos are synthetic avatars.

MAHNOB-HCI [135].[12] The MAHNOB-HCI dataset was originally collected for the purposes of creating systems for implicit tagging of multimedia content. Videos of 27 participants (15 women, 12 men) were collected while they were wearing an ECG sensor. This was one of the earliest public datasets that included videos and time-synchronized physiological ground truth. One limitation of this dataset is the heavy video compression, which means that physiological information in the videos is somewhat attenuated. Videos were recorded at a resolution of 780 × 580 and 61 Hz. Most analyses [31, 72] use a 30-second clip (frames from 306 through 2,135) from 527 video sequences.

BP4D+ and MMSE-HR [178].[13] The BP4D+ dataset is a multimodal dataset containing time-synchronized 3D/2D thermal and physiological recordings. This large dataset contains videos of 140 subjects and 10 emotional sitting tasks. The videos are stored in a relatively uncompressed format, and the dataset contains a relatively broad range of ages 18 to 66 years and ethnic or racial diversity. Furthermore, unlike many other datasets, the majority of subjects were female. Of note is that this dataset does not include either PPG or ECT gold standard measures but rather contains pulse pressure waves as measured via a fingercuff. The post pressure wave is similar to but different in morphology to the PPG signal. RGB videos were recorded at a resolution of 1040 × 1392 (note: this is portrait) and 24 Hz.

VIPL-HR [103].[14] VIPL-HR is the largest multimodal dataset with videos and time-synchronized physiological recordings; it contains 2,378 RGB or visible light videos and 752 NIR videos of 107 subjects. Gold standard PPG, HR, and SpO_2 were recorded. Videos were recorded with three RGB cameras and one NIR camera: an RGB Logitech C310 at a resolution of 960 × 720 and 25 Hz; a RealSense F200 NIR camera at a resolution of 640 × 480 and an RGB camera at 1920 × 1080, both 30 Hz; and an RGB HUAWEI P9 at a resolution of 1920 × 1080 and 30 Hz.

COHFACE [57].[15] The COHFACE dataset contains RGB video recordings synchronized with cardiac (PPG) and respiratory signals. The dataset includes 160 one-minute-long video sequences of 40 subjects (12 females and 28 males). The video sequences have been recorded with a Logitech HD C525 at a resolution of 640 × 480 pixels and a frame rate of 20 Hz. Gold standard measurements were acquired using the Thought Technologies BioGraph Infiniti system.

UBFC-RPPG [21].[16] The UBFC-RPPG RGB video dataset was collected with a Logitech C920 HD Pro at 30 Hz with a resolution of 640 × 480 in uncompressed 8-bit RGB format. A CMS50E transmissive pulse oximeter was used to obtain the gold standard PPG data. During the recording, the subjects were seated 1 m from the camera. All experiments were conducted indoors with a mixture of sunlight and indoor illumination.

UBFC-PHYS [97].[17] UBFC-PHYS is another public multimodal dataset with RGB videos, in which 56 subjects (46 women and 10 men) participated in a Trier Social Stress Test (TSST)-inspired experiment. Three tasks (rest, speech, and arithmetic) were completed by each subject, resulting in 168 videos. Gold standard BVP and electrodermal activity (EDA) measurements were collected via a wristband (Empatica E4). Before and after the experiment, participants completed a form to calculate their self-reported anxiety scores. The video recordings were at a resolution of 1024 × 1024 and 35 Hz.

[12] https://mahnob-db.eu/hci-tagging/.
[13] https://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html.
[14] https://vipl.ict.ac.cn/view_database.php?id=15.
[15] https://www.idiap.ch/en/dataset/cohface.
[16] https://sites.google.com/view/ybenezeth/ubfcrppg.
[17] https://sites.google.com/view/ybenezeth/ubfc-phys.

Rice CameraHRV [113].[18] The Rice CameraHRV consists of activities with complex facial movement, containing video recordings of 14 subjects (8 male, six female) during stationary, reading, talking, video watching, and deep breathing tasks (total of 60 recordings). Each video is 2 minutes in duration. Gold standard PPG data were collected using an FDA-approved pulse oximeter. The camera recordings were made with a Blackfly BFLY-U3-23S6C (Point Grey Research) with a Sony IMX249 sensor. Frames were captured at a resolution of 1920 × 1200 and 30 Hz.

MERL-Rice NIR Pulse (MR-NIRP) [106].[19] The MR-NIRP dataset contains recordings (19) of drivers in a cockpit driving around a city and recordings (18) stationary in a garage. Each video recorded in the garage is 2 minutes in duration, and those recorded while driving are 2 to 5 minutes long. The 18 (16 male, two female) subjects were healthy, aged 25to 60 years. Four of the subjects were recorded at night and 14 during the day. Recordings were made with NIR (Point Grey Grasshopper GS3-U3-41C6NIR-C) and RGB (FLIR Grasshopper3 GS3-PGE23S6C-C) cameras mounted on the dashboard in front of the subject. The NIR camera was fitted with a 940-nm hard-coated optical density bandpass filter from Edmund Optics with a 10-nm passband. Frames were captured at a resolution of 640 × 640 and 30 Hz (no gamma correction and with fixed exposure). Gold standard PPG data were recorded with a CMS 50D+ finger pulse oximeter at 60 Hz.

PURE [140]. The PURE dataset contains recordings of 10 subjects (8 male, 2 female) each during six tasks. The videos were captured with an RGB eco274CVGE camera (SVS-Vistek GmbH) at a resolution of 640 × 480 and 60 Hz. The subjects were seated in front of the camera at an average distance of 1.1 m and lit from the front with ambient natural light through a window. Gold standard measures of PPG and SpO_2 were collected with a pulse oximeter CMS50E attached to the finger. The six tasks were described as follows: (1) the subject was seated, stationary, and looking directly into the camera; (2) the subject was asked to talk while avoiding additional head motion; (3) the participant moved their head in a horizontal translational manner at an average speed proportional to the size of the face within the video; (4) this task was similar to the previous task with twice the velocity; (5) subjects were asked to orient their head toward targets placed in an arc around the camera in a predefined sequence, and the motions were designed to be random and not periodic (approximately 20° rotations); and (6) this task was similar to the previous task with larger head rotations (approximately 35° rotations).

rPPG [65].[20] The rPPG dataset includes 52 recording from three RGB cameras: a Logitech C920 webcam at a resolution of 1920 × 1080 (WMV2 video codec), a Microsoft VX800 webcam at a resolution of 640 × 480 (WMV3 video codec), and a Lenovo B590 laptop integrated webcam at a resolution of 640 × 480 pixels (WMV3 video codec). All recordings were 24-bit depth (3 × 8-bit per channel) at 15 Hz. The duration of the recordings was between 60 and 80 seconds. Between 2 and 14 videos were recorded for eight healthy subjects (seven male, one female, 24–37 years). Primary illumination was ambient daylight and indoor lighting. Subjects were seated 0.5 to 0.7 m from the camera. Gold standard PR measures were collected via a ChoiceMMed MD300C318 pulse oximeter. Participants completed a combination of stationary and head motion tasks. In the motion tasks, subjects rotated their head from right to left (with 120° amplitude), from up to down (with 100° amplitude). Subjects were also asked to speak and change facial expressions.

OBF [71]. The Oulu Bio-Face (OBF) database includes facial videos recorded from healthy subjects and from patients with AF. Recordings were made with an RGB and NIR camera. The subjects

[18] https://sh.rice.edu/camerahrv/.
[19] ftp://merl.com/pub/tmarks/MR_NIRP_dataset/README_MERL-Rice_NIR_Pulse_dataset.pdf.
[20] https://osf.io/fdrbh/wiki/home/.

were seated 1 m from the cameras. Two light sources were placed on either side of the cameras and illuminated the face at a 45°angle from a distance of 1.5 m. According to their published work, the authors plan to make this dataset publicly available; however, I was unable to find information about how to access it at the time of writing.

PFF [59].[21] The PPG From Face (PPF) database includes facial videos of 13 subjects each during five tasks (65 videos total). Each video is 3 minutes long, recorded with a resolution of 1280 × 720 at 50 Hz. Gold standard PR was collected via two Mio Alpha II wrist HR monitors (the average PR of the two readings is used). The subjects were seated in front of the camera at a distance of 0.5 m. The five tasks were as follows: (1) stationary with fluorescent illumination; (2) horizontal translational head/torso motion (right and left) with a frequency between 0.2 and 0.5 Hz (with fluorescent illumination); (3) stationary with ambient illumination primarily from windows and a computer monitor; (4) the same as task 2 with ambient illumination primarily from windows and a computer monitor; and (5) the same illumination condition as task 1, and each subject was riding an exercise bike at a constant speed.

VicarPPG [149].[22] VicarPPG includes 20 recordings of 10 subjects (aged 20–35 years). Each video is 90 seconds long, recorded at a resolution of 720 × 1280 and 30 Hz. Two videos were recorded for each subject: in the first, they were stationary after a rest period, and in the second, they were stationary after a physical exercise task. Gold standard PPG waveforms were recorded using a CMS50 pulse oximeter attached to the subject's fingertip.

VicarPPG-2/CleanerPPG [52].[23, 24] VicarPPG-2 includes 40 recordings of 10 subjects (mean age 29 years). Each video is 5 minutes long, recorded at a resolution of 1280 × 720 and 60 Hz. Four videos were recorded for each subject: in the first, they were stationary; in the second, they performed five different types of pre-planned angular body/head movements, including turning the head side to side (shaking), moving the head up and down (nodding), a combination of head shaking and nodding (round), moving the eyes while keeping head still, and naturally bobbing their heads while listening to music (dance); in the third, they played a stress-inducing game; and in the fourth, participants sat unrestrained after performing fatigue-inducing physical workouts to induce higher HRs. Gold standard PPG waveforms were recorded using a CMS50E pulse oximeter attached to the subject's fingertip. CleanerPPG is a PPG/ECG ground truth meta-dataset associated with the VicarPPG-2 dataset.

CMU [37].[25] A new CMU rPPG dataset contains videos recorded from 140 subjects subjects in India (44) and Sierra Leone (96). Three deidentified videos were generated from each face video, one each of the forehead, left cheek, and right cheek: a rectangular region with a resolution of 60 × 30 of the forehead, a square region with a resolution of 25 × 25 pixels of the left cheek, and a square region of 25 × 25 pixels of the right cheek. Videos were recorded at 15 Hz.

12.2 Synthetics and Data Augmentation

Labeled data is a limiting factor in almost all areas of computer vision and machine learning. Even large datasets can often suffer from selection bias and a lack of diversity. Although there are no easy solutions to these problems, there are methods for alleviating the problem: data augmentation and data simulation of synthesis.

[21] https://github.com/AvLab-CV/Pulse-From-Face-Database.
[22] https://docs.google.com/forms/d/e/1FAIpQLSeSXL-PHnFeI-s932qMoCAxZtFGhfbxKd9p003czGFHZJ__0Q/viewform.
[23] https://docs.google.com/forms/d/e/1FAIpQLScwnW_D5M4JVovPzpxA0Bf1ZCTaG5vh7sYu48I0MVSpgltvdw/viewform.
[24] https://docs.google.com/forms/d/e/1FAIpQLSdazfu6RTLpS20AxiqWgJOvQk5RjPMwissyAfin93GhqjYNPw/viewform.
[25] https://github.com/AiPEX-Lab/rppg_biases.

Several recent papers have proposed methods of data augmentation by creating videos with modified or augmented physiological patterns. Both Gideon and Stent [51] and Niu et al. [105] use resampling to augment the frequency of temporal information in videos. The former example performed augmentation in the video space, whereas the latter example performed the data on their spatial-temporal feature space arguing that it preserves the HR distributions. Specifically, to address the issue of the lack of representation of skin type in camera physiology dataset, Ba et al. [10] translate real videos from light-skinned subjects to dark-skinned subjects while being careful to preserve the PPG signals. A neural generator is used to simulate changes in melanin, or skin tone. This approach does not simulate other changes in appearance that might also be correlated with skin type. Nowara et al. [109] use video magnification (see Section 6 for augmenting the spatial appearance of videos and the temporal magnitude of changes in pixels). These augmentations help in the learning process, ultimately leading to the model learning better representations.

Recent work has proposed using simulations to create synthetic data for training camera physiological measurement algorithms. This can take two forms: statistical generation of videos using machine learning techniques or simulation using parameterized computer graphics models. Tsou et al. [150] used the former approach, leveraging neural models for video generation from a source image and a target PPG signal. Generative modeling definitely offers many opportunities in physiological measurement [82, 136]. Computer graphics can provide a way to create high-fidelity videos of the human body with augmented motions and skin subsurface scattering that simulate cardiac and respiratory processes [91]. Synthetics pipelines have the advantage of allowing simulation of many different combinations of appearance types, contexts, and physiological states, such as high HRs or arrhythmia states for which it may be difficult to create and gather examples in a laboratory. Research has shown that greater and greater numbers of avatars in a synthetic training set can continue to boost performance up to a point [91]. However, this is early work, and there remains a "sim-to-real" gap in performance of these systems—models trained purely on synthetic data do not generalize perfectly to real videos. Furthermore, these synthetics pipelines are typically expensive to construct, and therefore there may be limited access to them.

13 CONCLUSION

I have presented a survey of camera physiological measurement methods. These techniques have huge potential to improve the non-invasive measurement and assessment of vital signs. Camera technology and computational methods have advanced dramatically in the past 20 years, benefiting from advancements in optics, machine learning, and computer vision. With applications from consumer fitness to telehealth to neonatal monitoring to security and affective computing, there are many opportunities for these methods to have impact in the next 20 years. However, there are significant challenges that will need to be addressed to realize that vision. These include but are not limited to addressing unfair and inequitable performance, environmental robustness, the current lack of clinical validation, and privacy and ethical concerns. The ethical challenges associated with camera sensing should not be disregarded or treated lightly. Although there is a role for technological solutions that make it easier to remove physiological information from video, it is much more important to make sure that these technologies are always designed in an opt-in manner.

ACKNOWLEDGMENTS

I would like to thank all my collaborators who have contributed to work on camera physiological measurement over the past 10 years: Ming-Zher Poh, Rosalind Picard, Javier Hernandez, Sarah Gontarek, Ethan Blackford, Justin Estepp, Izumi Nishidate, Vincent Chen, Xin Liu, Ewa Nowara,

Brian Hill and Yuzhe Yang. I would also like to thank Wenjin Wang and Sander Stuijk for co-organizing the Computer Vision for Physiological Measurement (CVPM) workshops, which have helped to consolidate the research community around these methods, from which a lot of this work came.

REFERENCES

[1] Lonneke A. M. Aarts, Vincent Jeanne, John P. Cleary, C. Lieber, J. Stuart Nelson, Sidarto Bambang Oetomo, and Wim Verkruysse. 2013. Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit—A pilot study. *Early Human Development* 89, 12 (2013), 943–948.

[2] Freddy Abnousi, Guson Kang, John Giacomini, Alan Yeung, Shirin Zarafshar, Nicholas Vesom, Euan Ashley, Robert Harrington, and Celina Yong. 2019. A novel noninvasive method for remote heart failure monitoring: The Eulerian video magnification applications in heart failure study (AMPLIFY). *NPJ Digital Medicine* 2, 1 (2019), 1–6.

[3] Paul S. Addison, Dominique Jacquel, David M. H. Foo, and Ulf R. Borg. 2018. Video-based heart rate monitoring across a range of skin pigmentations during an acute hypoxic challenge. *Journal of Clinical Monitoring and Computing* 32, 5 (2018), 871–880.

[4] Solange Akselrod, David Gordon, F. Andrew Ubel, Daniel C. Shannon, A. Clifford Berger, and Richard J. Cohen. 1981. Power spectrum analysis of heart rate fluctuation: A quantitative probe of beat-to-beat cardiovascular control. *Science* 213, 4504 (1981), 220–222.

[5] John Allen. 2007. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement* 28, 3 (2007), R1.

[6] Robert Amelard, Richard L. Hughson, Danielle K. Greaves, Kaylen J. Pfisterer, Jason Leung, David A. Clausi, and Alexander Wong. 2017. Non-contact hemodynamic imaging reveals the jugular venous pulse waveform. *Scientific Reports* 7, 1 (2017), 1–10.

[7] Robert Amelard, Kaylen J. Pfisterer, Shubh Jagani, David A. Clausi, and Alexander Wong. 2018. Non-contact assessment of obstructive sleep apnea cardiovascular biomarkers using photoplethysmography imaging. In *Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics*, Vol. 10501. International Society for Optics and Photonics, 1050113.

[8] Robert Amelard, Christian Scharfenberger, Alexander Wong, and David A. Clausi. 2015. Illumination-compensated non-contact imaging photoplethysmography via dual-mode temporally coded illumination. In *Multimodal Biomedical Imaging X*, Vol. 9316. International Society for Optics and Photonics, 931607.

[9] Tucker Annis, Susan Pleasants, Gretchen Hultman, Elizabeth Lindemann, Joshua A. Thompson, Stephanie Billecke, Sameer Badlani, and Genevieve B. Melton. 2020. Rapid implementation of a COVID-19 remote patient monitoring program. *Journal of the American Medical Informatics Association* 27, 8 (Aug. 2020), 1326–1330. https://doi.org/10.1093/jamia/ocaa097

[10] Yunhao Ba, Zhen Wang, Kerim Doruk Karinca, Oyku Deniz Bozkurt, and Achuta Kadambi. 2021. Overcoming difficulty in obtaining dark-skinned subjects for remote-PPG by synthetic augmentation. *arXiv preprint arXiv:2106.06007*.

[11] Guha Balakrishnan, Fredo Durand, and John Guttag. 2013. Detecting pulse from head motions in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3430–3437.

[12] Marek Bartula, Timo Tigges, and Jens Muehlsteff. 2013. Camera-based system for contactless monitoring of respiration. In *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'13)*. IEEE, Los Alamitos, CA, 2672–2675.

[13] Gary G. Berntson, John T. Cacioppo, and Karen S. Quigley. 1993. Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology* 30, 2 (1993), 183–196.

[14] Mayur J. Bhamborae, Philipp Flotho, Adrian Mai, Elena N. Schneider, Alexander L. Francis, and Daniel J. Strauss. 2020. Towards contactless estimation of electrodermal activity correlates. In *Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'20)*. IEEE, Los Alamitos, CA, 1799–1802.

[15] Simone Bianco and Raimondo Schettini. 2014. Adaptive color constancy using faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 8 (2014), 1505–1518.

[16] Philip E. Bickler, John R. Feiner, and John W. Severinghaus. 2005. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Journal of the American Society of Anesthesiologists* 102, 4 (2005), 715–719.

[17] Ethan B. Blackford and Justin R. Estepp. 2017. Measurements of pulse rate using long-range imaging photoplethysmography and sunlight illumination outdoors. In *Optical Diagnostics and Sensing XVII: Toward Point-of-Care Diagnostics*, Vol. 10072. International Society for Optics and Photonics, 100720S.

[18] Ethan B. Blackford, Justin R. Estepp, and Daniel McDuff. 2018. Remote spectral measurements of the blood volume pulse with applications for imaging photoplethysmography. In *Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics*, Vol. 10501. International Society for Optics and Photonics, 105010Z.

[19] Nikolai Blanik, Konrad Heimann, Carina Pereira, Michael Paul, Vladimir Blazek, Boudewijn Venema, Thorsten Orlikowsky, and Steffen Leonhardt. 2016. Remote vital parameter monitoring in neonatology—Robust, unobtrusive heart rate detection in a realistic clinical scenario. *Biomedical Engineering/Biomedizinische Technik* 61, 6 (2016), 631–643.

[20] Vladimir Blazek, Ting Wu, and Dominik Hoelscher. 2000. Near-infrared CCD imaging: Possibilities for noninvasive and contactless 2D mapping of dermal venous hemodynamics. In *Optical Diagnostics of Biological Fluids V*, Vol. 3923. International Society for Optics and Photonics, 2–9.

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

[22] Giuseppe Boccignone, Donatello Conte, Vittorio Cuculo, Alessandro D'Amelio, Giuliano Grossi, and Raffaella Lanzarotti. 2020. An open framework for remote-PPG methods and their assessment. *IEEE Access* 8 (2020), 216083–216103. https://doi.org/10.1109/access.2020.3040936

[23] Frédéric Bousefsaf, Choubeila Maaoui, and Alain Pruski. 2013. Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate. *Biomedical Signal Processing and Control* 8, 6 (2013), 568–574.

[24] Frédéric Bousefsaf, Choubeila Maaoui, and Alain Pruski. 2014. Remote detection of mental workload changes using cardiac parameters assessed with a low-cost webcam. *Computers in Biology and Medicine* 53 (2014), 154–163.

[25] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. 77–91.

[26] Mihai Burzo, Daniel McDuff, Rada Mihalcea, Louis-Philippe Morency, Alexis Narvaez, and Verónica Pérez-Rosas. 2012. Towards sensing the influence of visual narratives on human affect. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*. 153–160.

[27] J. R. Catterall, P. M. A. Calverley, V. Brezinova, N. J. Douglas, H. M. Brash, C. M. Shapiro, and D. C. Flenley. 1982. Irregular breathing and hypoxaemia during sleep in chronic stable asthma. *Lancet* 319, 8267 (1982), 301–304.

[28] Giovanni Cennini, Jeremie Arguel, Kaan Akşit, and Arno van Leest. 2010. Heart rate monitoring via remote photoplethysmography with motion artifacts reduction. *Optics Express* 18, 5 (2010), 4867–4875.

[29] Sitthichok Chaichulee, Mauricio Villarroel, João Jorge, Carlos Arteta, Kenny McCormick, Andrew Zisserman, and Lionel Tarassenko. 2019. Cardio-respiratory signal extraction from video camera data for continuous non-contact vital sign monitoring using deep learning. *Physiological Measurement* 40, 11 (2019), 115001.

[30] Pradyumna Chari, Krish Kabra, Doruk Karinca, Soumyarup Lahiri, Diplav Srivastava, Kimaya Kulkarni, Tianyuan Chen, Maxime Cannesson, Laleh Jalilian, and Achuta Kadambi. 2020. Diverse R-PPG: Camera-based heart rate estimation for diverse subject skin-tones and scenes. *arXiv preprint arXiv:2010.12769*.

[31] Weixuan Chen and Daniel McDuff. 2018. DeepPhys: Video-based physiological measurement using convolutional attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV'18)*. 349–365.

[32] Weixuan Chen and Daniel McDuff. 2020. DeepMag: Source-specific change magnification using gradient ascent. *ACM Transactions on Graphics* 40, 1 (2020), 1–14.

[33] Weixuan Chen and Rosalind W. Picard. 2017. Eliminating physiological information from facial videos. In *Proceedings of the 2017 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG'17)*. IEEE, Los Alamitos, CA, 48–55.

[34] Xun Chen, Juan Cheng, Rencheng Song, Yu Liu, Rabab Ward, and Z. Jane Wang. 2018. Video-based heart rate measurement: Recent advances and future prospects. *IEEE Transactions on Instrumentation and Measurement* 68, 10 (2018), 3600–3615.

[35] Audrey Chung, Xiao Yu Wang, Robert Amelard, Christian Scharfenberger, Joanne Leong, Jan Kulinski, Alexander Wong, and David A. Clausi. 2015. High-resolution motion-compensated imaging photoplethysmography for remote heart rate monitoring. In *Multimodal Biomedical Imaging X*, Vol. 9316. International Society for Optics and Photonics, 93160A.

[36] Umur Aybars Ciftci, Ilke Demir, and Lijun Yin. 2020. FakeCatcher: Detection of synthetic portrait videos using biological signals. *arXiv preprint arXiv:1901.02212*.

[37] Ananyananda Dasari, Sakthi Kumar Arul Prakash, László A. Jeni, and Conrad S. Tucker. 2021. Evaluation of biases in remote photoplethysmography methods. *NPJ Digital Medicine* 4, 1 (2021), 1–13.

[38] Gerard De Haan and Vincent Jeanne. 2013. Robust pulse rate from chrominance-based rPPG. *IEEE Transactions on Biomedical Engineering* 60, 10 (2013), 2878–2886.

[39] Gerard De Haan and Arno Van Leest. 2014. Improved motion robustness of remote-PPG by using the blood volume pulse signature. *Physiological Measurement* 35, 9 (2014), 1913.

[40] Mohamed Elgendi. 2012. On the analysis of fingertip photoplethysmogram signals. *Current Cardiology Reviews* 8, 1 (2012), 14–25.
[41] Mohamed Elgendi, Richard Fletcher, Yongbo Liang, Newton Howard, Nigel H. Lovell, Derek Abbott, Kenneth Lim, and Rabab Ward. 2019. The use of photoplethysmography for assessing hypertension. *npj Digital Medicine* 2, 1 (June 2019), 60. https://doi.org/10.1038/s41746-019-0136-7
[42] Mohamed Elgharib, Mohamed Hefeeda, Fredo Durand, and William T. Freeman. 2015. Video magnification in presence of large motions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4119–4127.
[43] Justin R. Estepp, Ethan B. Blackford, and Christopher M. Meier. 2014. Recovering pulse rate during motion artifact with a multi-imager array for non-contact imaging photoplethysmography. In *Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC'14)*. IEEE, Los Alamitos, CA, 1462–1469.
[44] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the International Conference on Machine Learning*. 1126–1135.
[45] Marc Garbey, Nanfei Sun, Arcangelo Merla, and Ioannis Pavlidis. 2007. Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. *IEEE Transactions on Biomedical Engineering* 54, 8 (2007), 1418–1426.
[46] Irene García-López and Esther Rodriguez-Villegas. 2020. Extracting the jugular venous pulse from anterior neck contact photoplethysmography. *Scientific Reports* 10, 1 (2020), 1–12.
[47] Clare Garvie. 2016. *The Perpetual Line-Up: Unregulated Police Face Recognition in America*. Georgetown Law, Center on Privacy & Technology.
[48] Monika Gawałko, David Duncker, Martin Manninger, Rachel M. J. van der Velden, Astrid N. L. Hermans, Dominique V. M. Verhaert, Laurent Pison, et al. 2021. The European TeleCheck-AF project on remote app-based management of atrial fibrillation during the COVID-19 pandemic: Centre and patient experiences. *Europace* 23, 7 (July 2021), 1003–1015. https://doi.org/10.1093/europace/euab050
[49] Kim Gibson, Ali Al-Naji, Julie Fleet, Mary Steen, Adrian Esterman, Javaan Chahl, Jasmine Huynh, and Scott Morris. 2019. Non-contact heart and respiratory rate monitoring of preterm infants based on a computer vision system: A method comparison study. *Pediatric Research* 86, 6 (2019), 738–741.
[50] John Gideon and Simon Stent. 2021. Estimating heart rate from unlabelled video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2743–2749.
[51] John Gideon and Simon Stent. 2021. 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*. 3995–4004.
[52] Amogh Gudi, Marian Bittner, and Jan van Gemert. 2020. Real-time webcam heart-rate and variability estimation with clean ground truth for evaluation. *Applied Sciences* 10, 23 (2020), 8630.
[53] Otkrist Gupta, Dan McDuff, and Ramesh Raskar. 2016. Real-time physiological measurement and visualization using a synchronized multi-camera system. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 46–53.
[54] A. Harvey. 2012. *CV Dazzle: Camouflage from Computer Vision*. Technical Report. New York University.
[55] Qinghua He, Zhiyuan Sun, Yuandong Li, Wendy Wang, and Ruikang K. Wang. 2021. Spatiotemporal monitoring of changes in oxy/deoxy-hemoglobin concentration and blood pulsation on human skin using smartphone-enabled remote multispectral photoplethysmography. *Biomedical Optics Express* 12, 5 (2021), 2919–2937.
[56] Javier Hernandez, Yin Li, James M. Rehg, and Rosalind W. Picard. 2014. Bioglass: Physiological parameter estimation using a head-mounted wearable device. In *Proceedings of the 2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare through Innovations in Mobile and Wireless Technologies (MOBIHEALTH'14)*. IEEE, Los Alamitos, CA, 55–58.
[57] Guillaume Heusch, André Anjos, and Sébastien Marcel. 2017. A reproducible study on remote heart rate measurement. *arXiv preprint arXiv:1709.00962*.
[58] Brian L. Hill, Xin Liu, and Daniel McDuff. 2021. Learning higher-order dynamics in video-based cardiac measurement. *arXiv preprint arXiv:2110.03690*.
[59] Gee-Sern Hsu, ArulMurugan Ambikapathi, and Ming-Shiang Chen. 2017. Deep learning with time-frequency representation for pulse estimation from facial videos. In *Proceedings of the 2017 IEEE International Joint Conference on Biometrics (IJCB'17)*. IEEE, Los Alamitos, CA, 383–389.
[60] Markus Huelsbusch and Vladimir Blazek. 2002. Contactless mapping of rhythmical phenomena in tissue perfusion using PPGI. In *Medical Imaging 2002: Physiology and Function from Multidimensional Images*, Vol. 4683. International Society for Optics and Photonics, 110–117.
[61] Kenneth Humphreys, Tomas Ward, and Charles Markham. 2007. Noncontact simultaneous dual wavelength photoplethysmography: A further step toward noncontact pulse oximetry. *Review of Scientific Instruments* 78, 4 (2007), 044304.

[62] Noriko Inoue, Hideshi Kawakami, Hideya Yamamoto, Chikako Ito, Saeko Fujiwara, Hideo Sasaki, and Yasuki Kihara. 2017. Second derivative of the finger photoplethysmogram and cardiovascular mortality in middle-aged and elderly Japanese women. *Hypertension Research* 40, 2 (Feb. 2017), 207–211. https://doi.org/10.1038/hr.2016.123

[63] In Cheol Jeong and Joseph Finkelstein. 2016. Introducing contactless blood pressure assessment using a high speed video camera. *Journal of Medical Systems* 40, 4 (2016), 77.

[64] João Jorge, Mauricio Villarroel, Sitthichok Chaichulee, Kenny McCormick, and Lionel Tarassenko. 2018. Data fusion for improved camera-based detection of respiration in neonates. In *Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics*, Vol. 10501. International Society for Optics and Photonics, 1050112.

[65] Mikhail Kopeliovich and Mikhail Petrushan. 2019. Color signal processing methods for webcam-based heart rate evaluation. In *Proceedings of the SAI Intelligent Systems Conference*. 703–723.

[66] Alicja Kwasniewska, Maciej Szankin, Jacek Ruminski, Anthony Sarah, and David Gamba. 2021. Improving accuracy of respiratory rate estimation by restoring high resolution features with transformers and recursive convolutional models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3857–3867.

[67] Antony Lam and Yoshinori Kuno. 2015. Robust heart rate measurement from video using select random patches. In *Proceedings of the IEEE International Conference on Computer Vision*. 3640–3648.

[68] Eugene Lee, Evan Chen, and Chen-Yi Lee. 2020. Meta-rPPG: Remote heart rate estimation using a transductive meta-learner. In *Proceedings of the European Conference on Computer Vision*. 392–409.

[69] Magdalena Lewandowska and Jędrzej Nowak. 2012. Measuring pulse rate with a webcam. *Journal of Medical Imaging and Health Informatics* 2, 1 (2012), 87–92.

[70] Gregory F. Lewis, Rodolfo G. Gatto, and Stephen W. Porges. 2011. A novel method for extracting respiration rate and relative tidal volume from infrared thermography. *Psychophysiology* 48, 7 (2011), 877–887.

[71] Xiaobai Li, Iman Alikhani, Jingang Shi, Tapio Seppanen, Juhani Junttila, Kirsi Majamaa-Voltti, Mikko Tulppo, and Guoying Zhao. 2018. The OBF database: A large face video database for remote physiological signal measurement and atrial fibrillation detection. In *Proceedings of the 2018 13th IEEE International Conference on Automatic Face and Gesture Recognition (FG'18)*. IEEE, Los Alamitos, CA, 242–249.

[72] Xiaobai Li, Jie Chen, Guoying Zhao, and Matti Pietikainen. 2014. Remote heart rate measurement from face videos under realistic situations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4264–4271.

[73] Yu-Chen Lin and Yuan-Hsiang Lin. 2017. A study of color illumination effect on the SNR of rPPG signals. In *Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'17)*. IEEE, Los Alamitos, CA, 4301–4304.

[74] Ce Liu, Antonio Torralba, William T. Freeman, Frédo Durand, and Edward H. Adelson. 2005. Motion magnification. *ACM Transactions on Graphics* 24, 3 (2005), 519–526.

[75] Siqi Liu, Pong C. Yuen, Shengping Zhang, and Guoying Zhao. 2016. 3D mask face anti-spoofing with remote photoplethysmography. In *Proceedings of the European Conference on Computer Vision*. 85–100.

[76] Xin Liu, Josh Fromm, Shwetak Patel, and Daniel McDuff. 2020. Multi-task temporal shift attention networks for on-device contactless vitals measurement. In *Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20)*. Article 1627, 11 pages.

[77] Xin Liu, Brian L. Hill, Ziheng Jiang, Shwetak Patel, and Daniel McDuff. 2021. EfficientPhys: Enabling simple, fast and accurate camera-based vitals measurement. *arXiv preprint arXiv:2110.04447*.

[78] Xin Liu, Ziheng Jiang, Josh Fromm, Xuhai Xu, Shwetak Patel, and Daniel McDuff. 2021. MetaPhys: Few-shot adaptation for non-contact physiological measurement. In *Proceedings of the Conference on Health, Inference, and Learning*. 154–163.

[79] Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. 2018. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 389–398.

[80] Jean-Pierre Lomaliza, Hanhoon Park, and Mark Billinghurst. 2020. Combining photoplethysmography and ballistocardiography to address voluntary head movements in heart rate monitoring. *IEEE Access* 8 (2020), 226224–226239.

[81] Ilde Lorato, Sander Stuijk, Mohammed Meftah, Deedee Kommers, Peter Andriessen, Carola van Pul, and Gerard de Haan. 2021. Towards continuous camera-based respiration monitoring in infants. *Sensors* 21, 7 (2021), 2268.

[82] Hao Lu, Hu Han, and S. Kevin Zhou. 2021. Dual-GAN: Joint BVP and noise modeling for remote physiological measurement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12404–12413.

[83] Hong Luo, Deye Yang, Andrew Barszczyk, Naresh Vempala, Jing Wei, Si Jia Wu, Paul Pu Zheng, Genyue Fu, Kang Lee, and Zhong-Ping Feng. 2019. Smartphone-based blood pressure measurement using transdermal optical imaging technology. *Circulation: Cardiovascular Imaging* 12, 8 (2019), e008857.

[84] Daniel McDuff. 2018. Deep super resolution for recovering physiological information from videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1367–1374.

[85] Daniel McDuff and Ethan Blackford. 2019. iPhys: An open non-contact imaging-based physiological measurement toolbox. In *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'19)*. IEEE, Los Alamitos, CA, 6521–6524.

[86] Daniel McDuff, Ethan B. Blackford, and Justin R. Estepp. 2017. The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography. In *Proceedings of the 2017 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG'17)*. IEEE, Los Alamitos, CA, 63–70.

[87] Daniel McDuff, Justin R. Estepp, Alyssa M. Piasecki, and Ethan B. Blackford. 2015. A survey of remote optical photoplethysmographic imaging methods. In *Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'15)*. IEEE, Los Alamitos, CA, 6398–6404.

[88] Daniel McDuff, Sarah Gontarek, and Rosalind Picard. 2014. Remote measurement of cognitive stress via heart rate variability. In *Proceedings of the 2014 36th Annual International Engineering in Medicine and Biology Science Conference of the IEEE*. IEEE, Los Alamitos, CA, 2957–2960.

[89] Daniel McDuff, Sarah Gontarek, and Rosalind W. Picard. 2014. Improvements in remote cardiopulmonary measurement using a five band digital camera. *IEEE Transactions on Biomedical Engineering* 61, 10 (2014), 2593–2601.

[90] Daniel McDuff, Javier Hernandez, Sarah Gontarek, and Rosalind W. Picard. 2016. COGCAM: Contact-free measurement of cognitive stress during computer tasks with a digital camera. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 4000–4004.

[91] Daniel McDuff, Xin Liu, Javier Hernandez, Erroll Wood, and Tadas Baltrusaitis. 2021. Synthetic data for multi-parameter camera-based physiological sensing. In *Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'17)*. IEEE, Los Alamitos, CA.

[92] Daniel McDuff, Shuang Ma, Yale Song, and Ashish Kapoor. 2019. Characterizing bias in classifiers using generative models. *Advances in Neural Information Processing Systems* 32 (2019), 5403–5414.

[93] Daniel McDuff, Izumi Nishidate, Kazuya Nakano, Hideaki Haneishi, Yuta Aoki, Chihiro Tanabe, Kyuichi Niizeki, and Yoshihisa Aizu. 2020. Non-contact imaging of peripheral hemodynamics during cognitive and psychological stressors. *Scientific Reports* 10, 1 (2020), 1–13.

[94] Daniel McDuff, Miah Wander, Xin Liu, Brian L. Hill, Javier Hernandez, Jonathan Lester, and Tadas Baltrusaitis. 2022. SCAMPS: Synthetics for camera measurement of physiological signals. *arXiv preprint arXiv:2206.04197*.

[95] Daniel J. McDuff, Ethan B. Blackford, and Justin R. Estepp. 2017. Fusing partial camera signals for noncontact pulse rate variability measurement. *IEEE Transactions on Biomedical Engineering* 65, 8 (2017), 1725–1739.

[96] Lalit K. Mestha, Survi Kyal, Beilei Xu, Leslie Edward Lewis, and Vijay Kumar. 2014. Towards continuous monitoring of pulse rate in neonatal intensive care unit with a webcam. In *Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, Los Alamitos, CA, 3817–3820.

[97] Rita Meziatisabour, Yannick Benezeth, Pierre De Oliveira, Julien Chappe, and Fan Yang. 2021. UBFC-Phys: A multimodal database for psychophysiological studies of social stress. *IEEE Transactions on Affective Computing*. Early access, February 3, 2021.

[98] Andreia Vieira Moco, Sander Stuijk, and Gerard De Haan. 2015. Ballistocardiographic artifacts in PPG imaging. *IEEE Transactions on Biomedical Engineering* 63, 9 (2015), 1804–1811.

[99] Toshiaki Negishi, Shigeto Abe, Takemi Matsui, He Liu, Masaki Kurosawa, Tetsuo Kirimoto, and Guanghao Sun. 2020. Contactless vital signs measurement system using RGB-thermal image sensors and its clinical screening test on patients with seasonal influenza. *Sensors* 20, 8 (2020), 2171.

[100] Rang M. H. Nguyen, Dilip K. Prasad, and Michael S. Brown. 2014. Training-based spectral reconstruction from a single RGB image. In *Proceedings of the European Conference on Computer Vision*. 186–201.

[101] Aoxin Ni, Arian Azarang, and Nasser Kehtarnavaz. 2021. A review of deep learning-based contactless heart rate measurement methods. *Sensors* 21, 11 (2021), 3719.

[102] Izumi Nishidate, Noriyuki Tanaka, Tatsuya Kawase, Takaaki Maeda, Tomonori Yuasa, Yoshihisa Aizu, Tetsuya Yuasa, and Kyuichi Niizeki. 2011. Noninvasive imaging of human skin hemodynamics using a digital red-green-blue camera. *Journal of Biomedical Optics* 16, 8 (2011), 086012–086012.

[103] Xuesong Niu, Hu Han, Shiguang Shan, and Xilin Chen. 2018. VIPL-HR: A multi-modal database for pulse estimation from less-constrained face video. *arXiv preprint arXiv:1810.04927*.

[104] Xuesong Niu, Zitong Yu, Hu Han, Xiaobai Li, Shiguang Shan, and Guoying Zhao. 2020. Video-based remote physiological measurement via cross-verified feature disentangling. In *Proceedings of the European Conference on Computer Vision*. 295–310.

[105] Xuesong Niu, Xingyuan Zhao, Hu Han, Abhijit Das, Antitza Dantcheva, Shiguang Shan, and Xilin Chen. 2019. Robust remote heart rate estimation from face utilizing spatial-temporal attention. In *Proceedings of the 2019 14th IEEE International Conference on Automatic Face and Gesture Recognition (FG'19)*. IEEE, Los Alamitos, CA, 1–8.

[106] Ewa Magdalena Nowara, Tim K. Marks, Hassan Mansour, and Ashok Veeraraghavan. 2018. SparsePPG: Towards driver monitoring using camera-based vital signs estimation in near-infrared. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'18)*.

[107] Ewa M. Nowara, Daniel McDuff, and Ashok Veeraraghavan. 2020. A meta-analysis of the impact of skin tone and gender on non-contact photoplethysmography measurements. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 284–285.

[108] Ewa M. Nowara, Daniel McDuff, and Ashok Veeraraghavan. 2021. The benefit of distraction: Denoising camera-based physiological measurements using inverse attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4955–4964.

[109] Ewa M. Nowara, Daniel McDuff, and Ashok Veeraraghavan. 2021. Combining magnification and measurement for non-contact cardiac monitoring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3810–3819.

[110] Ewa M. Nowara, Daniel McDuff, and Ashok Veeraraghavan. 2021. Systematic analysis of video-based pulse measurement from compressed videos. *Biomedical Optics Express* 12, 1 (2021), 494–508.

[111] Tae-Hyun Oh, Ronnachai Jaroensri, Changil Kim, Mohamed Elgharib, Fr'edo Durand, William T. Freeman, and Wojciech Matusik. 2018. Learning-based video motion magnification. In *Proceedings of the European Conference on Computer Vision (ECCV'18)*. 633–648.

[112] Ahmed Osman, Jay Turcot, and Rana El Kaliouby. 2015. Supervised learning approach to remote heart rate estimation from facial videos. In *Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG'15)*, Vol. 1. IEEE, Los Alamitos, CA, 1–6.

[113] Amruta Pai, Ashok Veeraraghavan, and Ashutosh Sabharwal. 2018. CameraHRV: Robust measurement of heart rate variability using a camera. In *Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics*, Vol. 10501. International Society for Optics and Photonics, 105010S.

[114] Carina Barbosa Pereira, Xinchi Yu, Tom Goos, Irwin Reiss, Thorsten Orlikowsky, Konrad Heimann, Boudewijn Venema, Vladimir Blazek, Steffen Leonhardt, and Daniel Teichmann. 2018. Noncontact monitoring of respiratory rate in newborn infants using thermal imaging. *IEEE Transactions on Biomedical Engineering* 66, 4 (2018), 1105–1114.

[115] Tania Pereira, Nate Tran, Kais Gadhoumi, Michele M. Pelter, Duc H. Do, Randall J. Lee, Rene Colorado, Karl Meisel, and Xiao Hu. 2020. Photoplethysmography based atrial fibrillation detection: A review. *npj Digital Medicine* 3, 1 (Jan. 2020), 1–12. https://doi.org/10.1038/s41746-019-0207-9

[116] Phuong Pham and Jingtao Wang. 2015. AttentiveLearner: Improving mobile MOOC learning via implicit heart rate tracking. In *Proceedings of the International Conference on Artificial Intelligence in Education*. 367–376.

[117] Rosalind W. Picard. 2000. *Affective Computing*. MIT Press, Cambridge, MA.

[118] Christian Pilz. 2019. On the vector space in photoplethysmography imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*.

[119] Christian S. Pilz, Sebastian Zaunseder, Jarek Krajewski, and Vladimir Blazek. 2018. Local group invariance for heart rate estimation from face videos in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1254–1262.

[120] Silvia L. Pintea and Jan C. van Gemert. 2016. Making a case for learning motion representations with phase. In *Proceedings of the European Conference on Computer Vision*. 55–64.

[121] Ming-Zher Poh, Daniel McDuff, and Rosalind W. Picard. 2010. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering* 58, 1 (2010), 7–11.

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

[123] Ming-Zher Poh, Yukkee Cheung Poh, Pak-Hei Chan, Chun-Ka Wong, Louise Pun, Wangie Wan-Chiu Leung, Yu-Fai Wong, Michelle Man-Ying Wong, Daniel Wai-Sing Chu, and Chung-Wah Siu. 2018. Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms. *Heart* 104, 23 (2018), 1921–1928.

[124] Hua Qi, Qing Guo, Felix Juefei-Xu, Xiaofei Xie, Lei Ma, Wei Feng, Yang Liu, and Jianjun Zhao. 2020. DeepRhythm: Exposing DeepFakes with attentional visual heartbeat rhythms. In *Proceedings of the 28th ACM International Conference on Multimedia*. 4318–4327.

[125] Michal Rapczynski, Philipp Werner, and Ayoub Al-Hamadi. 2019. Effects of video encoding on camera-based heart rate estimation. *IEEE Transactions on Biomedical Engineering* 66, 12 (2019), 3360–3370.

[126] Andrew Reisner, Phillip A. Shaltis, Devin McCombie, H. Harry Asada, David S. Warner, and Mark A. Warner. 2008. Utility of the photoplethysmogram in circulatory monitoring. *Anesthesiology* 108, 5 (May 2008), 950–958. https://doi.org/10.1097/ALN.0b013e31816c89e1

[127] Ambareesh Revanur, Zhihua Li, Umur A. Ciftci, Lijun Yin, and László A. Jeni. 2021. The first Vision for Vitals (V4V) Challenge for non-contact video-based physiological estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2760–2767.

[128] Honnesh Rohmetra, Navaneeth Raghunath, Pratik Narang, Vinay Chamola, Mohsen Guizani, and Naga Rajiv Lakkaniga. 2021. AI-enabled remote monitoring of vital signs for COVID-19: Methods, prospects and challenges. *Computing*. Epub ahead of print, March 29, 2021. https://doi.org/10.1007/s00607-021-00937-7

[129] Gaetano Scebba, Giulia Da Poian, and Walter Karlen. 2020. Multispectral video fusion for non-contact monitoring of respiratory rate and apnea. *IEEE Transactions on Biomedical Engineering* 68, 1 (2020), 350–359.
[130] Fabian Schrumpf, Patrick Frenzel, Christoph Aust, Georg Osterhoff, and Mirco Fuchs. 2021. Assessment of non-invasive blood pressure prediction from PPG and rPPG signals using deep learning. *Sensors* 21, 18 (2021), 6022.
[131] Dangdang Shao, Chenbin Liu, and Francis Tsow. 2020. Noncontact physiological measurement using a camera: A technical review and future directions. *ACS Sensors* 6, 2 (2020), 321–334.
[132] Dangdang Shao, Francis Tsow, Chenbin Liu, Yuting Yang, and Nongjian Tao. 2016. Simultaneous monitoring of ballistocardiogram and photoplethysmogram using a camera. *IEEE Transactions on Biomedical Engineering* 64, 5 (2016), 1003–1010.
[133] Dangdang Shao, Yuting Yang, Chenbin Liu, Francis Tsow, Hui Yu, and Nongjian Tao. 2014. Noncontact monitoring breathing pattern, exhalation flow rate and pulse transit time. *IEEE Transactions on Biomedical Engineering* 61, 11 (2014), 2760–2767.
[134] Dvijesh Shastri, Manos Papadakis, Panagiotis Tsiamyrtzis, Barbara Bass, and Ioannis Pavlidis. 2012. Perinasal imaging of physiological stress and its affective potential. *IEEE Transactions on Affective Computing* 3, 3 (2012), 366–378.
[135] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. 2011. A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing* 3, 1 (2011), 42–55.
[136] Rencheng Song, Huan Chen, Juan Cheng, Chang Li, Yu Liu, and Xun Chen. 2021. PulseGAN: Learning to generate realistic pulse waveforms in remote photoplethysmography. *IEEE Journal of Biomedical and Health Informatics* 25, 5 (2021), 1373–1384.
[137] Radim Špetlík, Vojtech Franc, and Jirí Matas. 2018. Visual heart rate estimation with convolutional neural network. In *Proceedings of the British Machine Vision Conference*. 3–6.
[138] Janis Spigulis. 2017. Multispectral, fluorescent and photoplethysmographic imaging for remote skin assessment. *Sensors* 17, 5 (2017), 1165.
[139] Isaac Starr, A. J. Rawson, H. A. Schroeder, and N. R. Joseph. 1939. Studies on the estimation of cardiac output in man, and of abnormalities in cardiac function, from the heart's recoil and the blood's impacts; the ballistocardiogram. *American Journal of Physiology* 127, 1 (1939), 1–28.
[140] Ronny Stricker, Steffen Müller, and Horst-Michael Gross. 2014. Non-contact video-based pulse rate measurement on a mobile service robot. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, Los Alamitos, CA, 1056–1062.
[141] Guanghao Sun, Toshiaki Negishi, Tetsuo Kirimoto, Takemi Matsui, and Shigeto Abe. 2018. Noncontact monitoring of vital signs with RGB and infrared camera and its application to screening of potential infection. In *Non-Invasive Diagnostic Methods: Image Processing*. IntechOpen.
[142] Tiancheng Sun, Jonathan T. Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul Debevec, and Ravi Ramamoorthi. 2019. Single image portrait relighting. *ACM Transactions on Graphics* 38, 4 (2019), 1–12.
[143] Yu Sun, Vicente Azorin-Peris, Roy Kalawsky, Sijung Hu, Charlotte Papin, and Stephen E. Greenwald. 2012. Use of ambient light in remote photoplethysmographic systems: Comparison between a high-performance camera and a low-cost webcam. *Journal of Biomedical Optics* 17, 3 (2012), 037005.
[144] Yu Sun, Sijung Hu, Vicente Azorin-Peris, Roy Kalawsky, and Stephen E. Greenwald. 2012. Noncontact imaging photoplethysmography to effectively access pulse rate variability. *Journal of Biomedical Optics* 18, 6 (2012), 061205.
[145] Yu Sun and Nitish Thakor. 2015. Photoplethysmography revisited: From contact to noncontact, from point to imaging. *IEEE Transactions on Biomedical Engineering* 63, 3 (2015), 463–477.
[146] Chihiro Takano and Yuji Ohta. 2007. Heart rate measurement based on a time-lapse image. *Medical Engineering & Physics* 29, 8 (2007), 853–857.
[147] Kenji Takazawa, Nobuhiro Tanaka, Masami Fujita, Osamu Matsuoka, Tokuyu Saiki, Masaru Aikawa, Sinobu Tamura, and Chiharu Ibukiyama. 1998. Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform. *Hypertension* 32, 2 (Aug. 1998), 365–370. https://doi.org/10.1161/01.HYP.32.2.365
[148] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh. 2014. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiological Measurement* 35, 5 (2014), 807.
[149] H. Emrah Tasli, Amogh Gudi, and Marten Den Uyl. 2014. Remote PPG based vital sign measurement using adaptive facial regions. In *Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP'14)*. IEEE, Los Alamitos, CA, 1410–1414.
[150] Yun-Yun Tsou, Yi-An Lee, and Chiou-Ting Hsu. 2020. Multi-task learning for simultaneous video generation and remote photoplethysmography estimation. In *Proceedings of the Asian Conference on Computer Vision*.
[151] Rik van Esch, Kambez Ebrahimkheil, Iris Cramer, Wenjin Wang, Tomas Kaandorp, Federica Sammali, A. T. M. Dierick-van Daele, et al. 2021. Remote PPG for heart rate monitoring: Lighting conditions and camera shutter time.

In *Proceedings of the 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'21)*.

[152] Mark van Gastel, Sander Stuijk, and Gerard de Haan. 2015. Motion robust remote-PPG in infrared. *IEEE Transactions on Biomedical Engineering* 62, 5 (2015), 1425–1433.

[153] Mark van Gastel, Sander Stuijk, Sebastiaan Overeem, Johannes P. van Dijk, Merel M. van Gilst, and Gerard de Haan. 2020. Camera-based vital signs monitoring during sleep—A proof of concept study. *IEEE Journal of Biomedical and Health Informatics* 25, 5 (2020), 1409–1418.

[154] Wim Verkruysse, Lars O. Svaasand, and J. Stuart Nelson. 2008. Remote plethysmographic imaging using ambient light. *Optics Express* 16, 26 (2008), 21434–21445.

[155] Zachary Vesoulis, Anna Tims, Hafsa Lodhi, Natasha Lalos, and Halana Whitehead. 2022. Racial discrepancy in pulse oximeter accuracy in preterm infants. *Journal of Perinatology* 42, 1 (2022), 79–85.

[156] Mauricio Villarroel, Sitthichok Chaichulee, João Jorge, Sara Davis, Gabrielle Green, Carlos Arteta, Andrew Zisserman, Kenny McCormick, Peter Watkinson, and Lionel Tarassenko. 2019. Non-contact physiological monitoring of preterm infants in the neonatal intensive care unit. *npj Digital Medicine* 2, 1 (2019), 1–18.

[157] Tom Vogels, Mark van Gastel, Wenjin Wang, and Gerard de Haan. 2018. Fully-automatic camera-based pulse-oximetry during sleep. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1349–1357.

[158] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T. Freeman. 2013. Phase-based video motion processing. *ACM Transactions on Graphics* 32, 4 (2013), 1–10.

[159] Julian A. Waller. 1966. Traffic accidents—C medical conditions as a cause. *California Medicine* 105, 3 (1966), 197.

[160] Chen Wang, Thierry Pun, and Guillaume Chanel. 2018. A comparative survey of methods for remote heart rate detection from frontal face videos. *Frontiers in Bioengineering and Biotechnology* 6 (2018), 33.

[161] Edward Jay Wang, Junyi Zhu, Mohit Jain, Tien-Jui Lee, Elliot Saba, Lama Nachman, and Shwetak N. Patel. 2018. Seismo: Blood pressure monitoring using built-in smartphone accelerometer and camera. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–9.

[162] Jue Wang, Steven M. Drucker, Maneesh Agrawala, and Michael F. Cohen. 2006. The cartoon animation filter. *ACM Transactions on Graphics* 25, 3 (2006), 1169–1173.

[163] Wenjin Wang, Albertus C. den Brinker, Sander Stuijk, and Gerard de Haan. 2017. Algorithmic principles of remote PPG. *IEEE Transactions on Biomedical Engineering* 64, 7 (2017), 1479–1491.

[164] Wenjin Wang, Sander Stuijk, and Gerard De Haan. 2014. Exploiting spatial redundancy of image sensor for motion robust rPPG. *IEEE Transactions on Biomedical Engineering* 62, 2 (2014), 415–425.

[165] Yiyin Wang, Wenjin Wang, Mark van Gastel, and Gerard de Haan. 2019. Modeling on the feasibility of camera-based blood glucose measurement. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*.

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

[167] Fokko P. Wieringa, Frits Mastik, and Antonius F. W. van der Steen. 2005. Contactless multiple wavelength photoplethysmographic imaging: A first step toward "SpO 2 camera" technology. *Annals of Biomedical Engineering* 33, 8 (2005), 1034–1041.

[168] Michael J. Wilber, Vitaly Shmatikov, and Serge Belongie. 2016. Can we still avoid automatic face detection? In *Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV'16)*. IEEE, Los Alamitos, CA, 1–9.

[169] Bing-Fei Wu, Yun-Wei Chu, Po-Wei Huang, and Meng-Liang Chung. 2019. Neural network based luminance variation resistant remote-photoplethysmography for driver's heart rate monitoring. *IEEE Access* 7 (2019), 57210–57225.

[170] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. 2012. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics* 31, 4 (2012), Article 65, 8 pages.

[171] Ting Wu, Vladimir Blazek, and Hans Juergen Schmitt. 2000. Photoplethysmography imaging: A new noninvasive and noncontact method for mapping of the dermal perfusion changes. In *Optical Techniques and Instrumentation for the Measurement of Blood Composition, Structure, and Dynamics*, Vol. 4163. International Society for Optics and Photonics, 62–70.

[172] Takayuki Yamada, Seiichi Gohshi, and Isao Echizen. 2013. Privacy visor: Method for preventing face image detection by using differences in human and device sensitivity. In *Proceedings of the IFIP International Conference on Communications and Multimedia Security*. 152–161.

[173] Bryan P. Yan, William H. S. Lai, Christy K. Y. Chan, Stephen Chun-Hin Chan, Lok-Hei Chan, Ka-Ming Lam, Ho-Wang Lau, et al. 2018. Contact-free screening of atrial fibrillation by a smartphone using facial pulsatile photoplethysmographic signals. *Journal of the American Heart Association* 7, 8 (2018), e008585.

[174] Sun Yu, Sijung Hu, Vicente Azorin-Peris, Jonathon A. Chambers, Yisheng Zhu, and Stephen E. Greenwald. 2011. Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise. *Journal of Biomedical Optics* 16, 7 (2011), 077010.

[175] Zitong Yu, Wei Peng, Xiaobai Li, Xiaopeng Hong, and Guoying Zhao. 2019. Remote heart rate measurement from highly compressed facial videos: An end-to-end deep learning solution with video enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 151–160.

[176] Zijie Yue, Shuai Ding, Shanlin Yang, Hui Yang, Zhili Li, Youtao Zhang, and Yinghui Li. 2021. Deep super-resolution network for rPPG information recovery and noncontact heart rate estimation. *IEEE Transactions on Instrumentation and Measurement* 70 (2021), 1–11.

[177] Yichao Zhang, Silvia L. Pintea, and Jan C. Van Gemert. 2017. Video acceleration magnification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 529–537.

[178] Zheng Zhang, Jeff M. Girard, Yue Wu, Xing Zhang, Peng Liu, Umur Ciftci, Shaun Canavan, et al. 2016. Multimodal spontaneous emotion corpus for human behavior analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3438–3446.

[179] Changchen Zhao, Chun-Liang Lin, Weihai Chen, and Zhengguo Li. 2018. A novel framework for remote photoplethysmography pulse extraction on compressed videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1299–1308.

Received 23 November 2021; revised 20 June 2022; accepted 18 July 2022