Feasibility of Remote Blood Pressure Estimation via Narrow-band Multi-wavelength Pulse Transit Time

GAŠPER SLAPNIČAR, Jožef Stefan Institute, Slovenia
WENJIN WANG, Southern University of Science and Technology, China
MITJA LUŠTREK, Jožef Stefan Institute, Slovenia

Contact-free sensing gained much traction in the past decade. While remote monitoring of some parameters (heart rate) is approaching clinical levels of precision, others remain challenging (blood pressure). We investigated the feasibility of estimating blood pressure (BP) via pulse transit time (PTT) in a novel remote single-site manner, using a modified RGB camera. A narrow-band triple band-pass filter allowed us to measure the PTT between different skin layers, harvesting information from green and near-infrared wavelengths. The filter minimizes the inter-channel influence and band overlap, however some overlap remains within the filter bands. We further resolve this using a color-channel model and a novel channel-separation method. Using the proposed setup and algorithm, we obtained multi-wavelength (MW) PTTs in an experiment inducing BP changes to 9 subjects. The results showed good absolute Pearson's correlation coefficient between both MW PTT and systolic BP ($R = 0.61$, $p = 0.08$) as well as diastolic BP ($R = 0.54$, $p = 0.05$), pointing to feasibility of the proposed novel remote MW BP estimation via PTT. This was further confirmed in a leave-one-subject-out experiment, where a simple Random Forest regression model achieved mean absolute errors of 3.59 and 2.63 mmHg for systolic and diastolic BP respectively.

CCS Concepts: • **Applied computing** → *Health informatics*; • **Computing methodologies** → Model verification and validation; • **Hardware** → Emerging tools and methodologies.

Additional Key Words and Phrases: remote sensing, blood pressure, pulse transit time, multi-wavelength, photoplethysmography

1 INTRODUCTION

Wireless sensing has seen large development and increased interest over the past decade, especially in the fields of indoor localization [27], activity recognition [29] and physiological signal [16, 28, 31] monitoring. The initial research focus of the latter was on those physiological parameters that are most expressed and most easily detectable in a contact-free manner, such as heart rate (HR) and respiratory rate (RR). In recent years however, the field expanded rapidly towards the monitoring of more subtle parameters, such as heart rate variability (HRV), blood oxygen saturation (SpO_2) and blood pressure (BP). This was facilitated by improved accuracy of sensors going hand-in-hand with their reduced cost and subsequent widespread availability, especially in the form of RGB cameras. Alongside hardware development there have also been major algorithmic advancements, especially with the transition towards deep learning, where deep convolutional neural networks (CNNs) and transformer architectures dominate the current state of the art [5, 14]. These developments are relevant in the context of AI of Things (AIoT), as the ever more present neural networks are one of the manifestations of transition from IoT to AIoT. The latter has especially large potential in the healthcare domain [26]. Contact-free systems, similar to the one proposed in our work, can form an interconnected network of intelligent systems with many benefits.

Authors' addresses: Gašper Slapničar, gasper.slapnicar@ijs.si, Jožef Stefan Institute, Jamova cesta 39, Ljubljana, Slovenia, 1000; Wenjin Wang, wangwj3@sustech.edu.cn, Southern University of Science and Technology, 1088 Xueyuan Avenue, Shenzhen, China, 518055; Mitja Luštrek, mitja.lustrek@ijs.si, Jožef Stefan Institute, Jamova cesta 39, Ljubljana, Slovenia, 1000.

1550-4859/2023/5-ART
https://doi.org/10.1145/3597302

For example they monitor patients, give intelligent suggestions for disease management or prevention, and offer decision support, all while being non-invasive and not interfering or obstructing the patients everyday lives.

The main enabler at the forefront of contact-free physiological monitoring remains remote photoplethysmography (rPPG), which reflects cardiovascular activity and blood perfusion via color changes of the skin [2]. These can be captured using a simple setup comprising an RGB camera and a light source. State-of-the-art research uses rPPG reconstruction waveforms to estimate some highly expressed physiological parameters, such as HR, in a robust and accurate manner, achieving errors as low as 1-3 beats-per-minute (BPM) [5]. Other more subtle parameters, such as BP, remain difficult to estimate while being highly valuable for both physicians and patients. Attempts at BP estimation include either analysis of rPPG waveform morphology or measurement of pulse transit time (PTT) between multiple sites (e.g., palm and forehead) [24]. PTT is a surrogate metric often used when BP cannot be measured directly (e.g., due to sensor limitations, for instance having just optical sensors), and is known to be well-correlated with BP.

BP is the most commonly measured physiological parameter both at primary and secondary healthcare, as well as at home. Traditional wide-spread cuff-based measurement is cumbersome, induces subject stress (white-coat syndrome), requires following specific protocols and can lead to inaccurate or lack of (home) measurement entirely. Subsequently, more unobtrusive organic measurement is desirable. This is being addressed via previously mentioned contact-free BP estimation, where most commonly rPPG is reconstructed via skin color changes using an RGB camera. If only a single rPPG signal is reconstructed, it is used to compute morphological features by either defining a feature vector based on cycle reference points, or by using a black box model (e.g., neural network) to derive features internally. These are then correlated to BP by a data-driven model. When two rPPG signals are reconstructed from different locations, delay between reference points is measured, as this PTT is medically known to be correlated to BP. The latter is challenging, as two precisely synchronized sensors are required and two regions of interest must be precisely monitored at once. Most commonly one of these is the face, which also comes with privacy preservation issues. BP estimation from PPG in general remains an active research area in both contact and contact-free monitoring, and has been subject to several challenges, both technical and fundamental. The former are being addressed successfully via sensor improvements and noise removal algorithms. However, fundamentally waveform analysis remains challenging as the connection between features (often obtained from black-box models [5]) and BP is less clear [24]. Additionally, a general predictive model is difficult to obtain, as many researchers agree that hemodynamics governing BP are highly person-specific due to specifics of both blood and vascular structure, thus requiring (re)calibration of models to each individual [25]. PTT on the other hand is well-established and the connection to BP is well understood, however, requiring several measurement sites and sensors is a limitation and a burden to the user.

Instead of using traditional multi-site PTT measurement technique, we evaluate the feasibility of a single-site contact-free multi-wavelength (MW) approach in which we measure PTT between different depths of skin instead of different spatial locations. This would allow for privacy-preserving unobtrusive monitoring that minimizes the user burden and uses more information from a single site by monitoring the blood flow through the skin layers instead of between different spatial locations. It also removes the requirement for two sensors at different locations and corresponding regions with skin exposure, circumventing synchronization, or requirement for specific position of a subject (e.g., palm raised next to face) that is required for monitoring of two distant sites. Additionally, it importantly improves on existing MW contact approaches, since even slight skin contact causes some compression that influences the waveform [20, 21]. Such an approach fundamentally differs from existing work in BP estimation, which is based on either contact single-site MW PTT (traditional wearables approach) [18, 19], remote PPG morphology analysis (connection between features and BP not medically clear) [5], or remote multi-site PTT (requires several measuring sites with skin exposure, less feasible in practice) [22].

We also investigate the correlations between the observed PTTs and recorded ground-truth systolic (SBP) and diastolic (DBP) BP in an experiment involving 9 subjects in two scenarios, eliciting distinctly different BP

values. Finally, we evaluate explicit BP prediction capabilities by using a regression model to predict SBP and DBP. Additionally, we propose a vital, yet to the best of our knowledge always overlooked, color channel separation method that is crucial for the observation of PTTs, which are inherently very short in a single-site MW approach, due to the short distance the blood traverses.

Such a contact-free single-site MW approach would offer numerous advantages compared to the current "golden standard" for BP measurement, which remains the use of a cuff-based sphygmomanometer. The proposed alternative could allow for simple contact-free BP screenings, for example daily measurements when a person is in front of a bathroom mirror. It would omit the need for a trained person to place the cuff and monitor the measurement, be more comfortable, and potentially have less influence on the measurement result (e.g., white coat syndrome) – all while using just a single RGB sensor. Additionally, since the PTTs being measured are very short, the waveforms must not be influenced or distorted in any way. While subtle, such influence is present when using contact sensors due to slight skin compression [20]. Researchers in this field have also shown that leveraging information from the infrared (deeper skin layer) is valuable as pulsatility is more stable [12, 21]. Due to these shortcomings, there is an apparent need for a novel contact-free approach that also uses non-conventional infrared part of the spectrum in addition to visible light, to reconstruct PPG waveforms. The relationship between PPG, PTT and BP on the other hand is widely-researched and well-established [2, 13]. We discuss shortcomings of existing methods in greater detail in Section 2. A pilot study describing such a recording setup has been reported at BHI 2022 (Slapničar et al., citation pending upon publication), however this paper focuses on entirely different methodology validated in different experiments and does not overlap.

The rest of the paper is organized as follows: in Section 2 we give a detailed overview of related work, focusing especially on remote and MW sensing while also looking at state-of-the-art BP estimation from PPG; we then explain the physiological background and reasoning behind MW PTT and its relationship with BP, together with the camera physics and the model of our color channels in Section 3; in Section 4 we detail our recording setup and describe the data collection protocol; in Section 5 we describe our data processing pipeline and correlation analysis; and in Section 6 we discuss our findings alongside the limitations, while also proposing some future work. We conclude our paper with a summary in Section 7.

2 RELATED WORK

2.1 MW Monitoring of Physiological Signals

A review study on MW wearable PPG was conducted recently by Ray et al. [23] in which they highlight the importance of a variety of factors for MW skin monitoring, such as melanin (skin tone) and skin temperature. They show an extensive overview of wavelengths and hardware used in these studies, showing that in the majority of cases the blue, green and infrared (IR) part of the light spectrum is used, as these wavelengths have different penetration depths in accordance with relevant vascular content at different depths of skin. Their study was broad in terms of monitored vitals, however, for BP estimation they highlight the work of Liu et al., who proposed single-site BP estimation using a custom MW contact sensing probe [19]. They showed a decent correlation between SBP and PTT using blue and red PPG trace. They later expanded their research [18] showing that MW PTT is correlated with SVR – a metric correlated to (traditional) PTT – which can in turn be used for continuous BP estimation. In a study using 20 subjects they achieved low errors of around 3 mmHg for SBP using a four-wavelength (470, 570, 590, 940 nm) sequential illumination sensor, which is not practical in a remote setting (fast-flickering lights). These errors were achieved on per-subject basis using a formula connecting PTT and BP that requires calibration values.

An early contact-free proposal of MW imaging for HR and SpO2 monitoring was given by Wieringa et al. [30]. They built a custom recording setup with a CMOS camera and light sources emitting wavelengths of 660, 810 and 940 nm. They acquired respiratory, HR and SpO2 information from PPGs at multiple wavelengths, which

correlated well with ground truth. Another study in remote MW PPG analysis was reported by Moço et al. [20] in which they highlight skin inhomogeneity as an important factor in phase differences of PPGs from different wavelengths. They suggest that the observed temporal delays reflect morphological heterogeneity of the PPG waveform due to tissue-specific factors to a larger extent than actual propagation-related properties, which can be a major challenge for PTT applications. They initially investigated only the spectrum of visible light, focusing on green and red. However, in later work they highlight NIR wavelength as being a good choice for SpO2 measurement, as it reaches the deeper pulsating arterioles instead of only the capillary bed. They express concerns about the latter exhibiting pulsatility as a consequence of elastic deformations and subsequent capillary density changes in the monitored region, rather than only blood volume changes, which they support by showing that skin compression can importantly influence the PPG morphology (e.g., when using a fingertip contact sensor). They also offer detailed models for skin layers microvasculature [21]. Their work offers important fundamentals in terms of understanding and modelling the non-trivial blood perfusion in the complex skin microvasculature.

2.2 Remote BP Estimation

There is a myriad of recent work dealing with BP estimation from rPPG. Some of it relies heavily on morphological properties of the waveform or several waveforms obtained from smaller sub-regions of the chosen measurement site, such as the forehead. Iuchi et al. [14] recently proposed one such method in which they used RGB facial videos that were converted to what they call spatial pulse wave signal. This is in essence a sequence of frames divided into small sub-regions where each sub-region shows the intensity of hemoglobin in the skin in the form of a heatmap. This sequence of frames is then fed into a branched 2D CNN to estimate SBP and DBP with mean absolute errors (MAE) as low as 6.7 mmHg for SBP and 5.4 mmHg for DBP. Importantly they also showed good correlation coefficients between prediction and ground truth – up to 0.81 for SBP and 0.84 for DBP – indicating that BP variations (even though quite limited in the dataset) were successfully modelled. In their experiment they also used personalization, as the data was split into training and testing using the 80-20 split, where data of same individual is present in both sets. Some other recent approaches are complete black-box end-to-end neural network implementations, taking a sequence of images as input and predicting BP as output, without manual data processing intermediate steps. Botina et al. [5] proposed a truly end-to-end 3D CNN architecture capable of real-time rPPG reconstructions and subsequent physiological parameter estimation, in which they reduce the dimensionality of the input frames to achieve up to 88% increase in speed. They still maintained comparable accuracy in terms of correlation coefficients to other state-of-the-art architectures. Despite good results, the challenges mentioned previously remain – personalization and calibration requirements [24], good skin exposure with enough skin pixels, lighting conditions, etc.

2.3 Synthesis and Findings

Looking at related work as well as review papers of contact-free physiological parameter estimation, we came to the following findings:

- MW measurement of PTT has seen success in contact approaches using high-frequency sensors. However, Moço et al. [21] clearly showed that compression inherently produced by contact sensors can be problematic for PTTs and the interpretation of results. Additionally they highlight the importance of NIR wavelengths which are less prone to influence of skin heterogeneity, but are typically not used, as the NIR is not available out-of-the-box in consumer cameras.
- Inter-channel influence and sensitivity of specific-wavelength pixels to other wavelengths is never discussed, despite being of vital importance when dealing with very short MW PTTs that can not be easily discerned. Without channel separation, the PTTs are extremely short and difficult to measure.

- Remote BP monitoring is a popular research topic, however, it faces many challenges like person-specific hemodynamics and tissue specifics, difficulty of modelling BP variations (e.g., often predicts around mean BP values, which results in low errors, since the majority of data is in that range) and the use of black-box models that lack explainability. Furthermore, when using traditional multi-site PTT approach, in case of multiple sensors, precise synchronization and multi-site calibration is required.

Our proposed solution aims to answer these by using a contact-free sensor in the form of a modified triple-band-pass high-frequency camera and a channel separation method preceding conventional data processing. This type of sensor modification allows us to leverage information from deeper skin layer via NIR part of the spectrum. In contrast to existing single-site approaches, which use contact PPG sensors, our proposed setup and method do not require skin contact, thus not compressing nor distorting the tissue and the waveform, meaning the precise location of systolic peaks is completely preserved. This is important for computation of short PTTs between skin layers. Usage of narrow triple-band-pass filter in combination with algorithmic channel separation resolves the inter-channel influence and overlap, allowing for robust PTT observations. Additionally, contact PPG sensors are always dependant on either battery life (consumer wearables) or cables (most often clinical devices), making them more obtrusive and not suitable in all circumstances. In different contexts, different recording setups are preferred, as each has its pros and cons. While a traditional contact approach circumvents many technical challenges such as light source and image sensor specifics, the proposed contact-free approach would be more desired in cases where wearables are not feasible due to some subject specifics such as skin conditions, allergies, preterm infants, etc. Personalization or person-specific calibration remains a fundamental challenge, which we acknowledge and work around by using a small amount of each subjects data in the training of regression models. We describe the proposed methods in detail, followed by our experimental results, in Sections 4–6.

3 PHYSIOLOGICAL AND PHYSICAL BACKGROUND

3.1 Hemodynamics and Skin Physiology

The underlying idea behind PTT is to measure the time it takes for a specific heart pulse wave to traverse a certain distance. This idea was historically most commonly implemented using pulse arrival time (PAT), where an electrocardiogram (ECG) and a distal PPG (usually at the fingertip) are synchronized and then the R peak in the QRS complex of the ECG and the corresponding systolic peak in the PPG are used as the reference points between which the PAT is measured [8].When the temporal delay is measured using two PPG signals, it is referred to as PTT [22]. This delay is shorter if vascular stiffness and BP are higher, and longer when vascular stiffness and BP are lower.

While the relationship between PTT and BP seems simple and intuitive, the hemodynamics of humans are neither trivial nor general. There are two crucial variables determining systemic BP, the cardiac output (CO, the amount of blood ejected by the heart into aorta in some time, typically one minute) and the systemic vascular resistance (SVR, the fluidic resistance of blood vessels opposed to the flow of blood), as given in 1 [13].

$$BP = CO \cdot SVR \tag{1}$$

To introduce PTT into the dependancy via SVR, we rely on the Moens-Korteweg equation (2):

$$PWV = \frac{L}{PTT} = \sqrt{\frac{E_{inc} \cdot h}{2r\rho}} \tag{2}$$

where PWV is pulse wave velocity, which is the velocity at which a pulse wave propagates a certain distance L, E_{inc} is the incremental elastic modulus of the vessel wall, h is the vascular wall thickness, r is the vessel radius, and ρ is the blood density, assuming the arterial wall is isotropic, meaning uniform in all directions. SVR is related

to the elastic modulus on the right-hand side of the equation [13]. PWV is complex and varies drastically as the pulse traverses the cardiovascular system due to different types of vessels in the body. Additionally, especially in thinner vessels, wave reflection makes the PWV difficult to estimate. Literature reports different values spanning 2-3 orders of magnitude. Precise measurements in capillaries, which are the dominant vessels between different skin layers of the palm, show expected PWV in the range of 6.4 – 17.6 mm/s [4]. Assuming average palm skin thickness of 1 mm, we can expect PTTs of around 156 – 57 ms. Importantly, this depends on specific elastic properties of the arterial / capillary wall and can vary between subjects.

Modern PTT approaches simplify the traditional setup by commonly replacing the ECG with a PPG or a camera sensor, as these sensors are cheaper, simpler to use, less obtrusive and more accessible [15]. Despite this, the multi-site nature of measurement remains, and comes with its set of challenges, such as the noise on each sensor and precise synchronization required. MW approach circumvents these by using a single sensor at a single measuring site. Instead of measuring the delay of the pulse reference points between two different locations, such an approach is instead based on the idea that different wavelengths of light penetrate to different skin depths, reflecting blood perfusion as it arrives from deeper towards the upper skin layers [18, 19]. Related work established that a direct light source with wavelength 550 nm (green) can reach depths between 0.1 and 1 mm (mostly dermis layer), while wavelength 850 nm (infrared) penetrates beyond 2 mm deep (hypodermis layer), depending on skin tissue specifics [7]. The obtained waveforms from different skin depths are expected to be slightly delayed due to blood traversal. The fundamental difference between approaches is shown in Fig. 1.

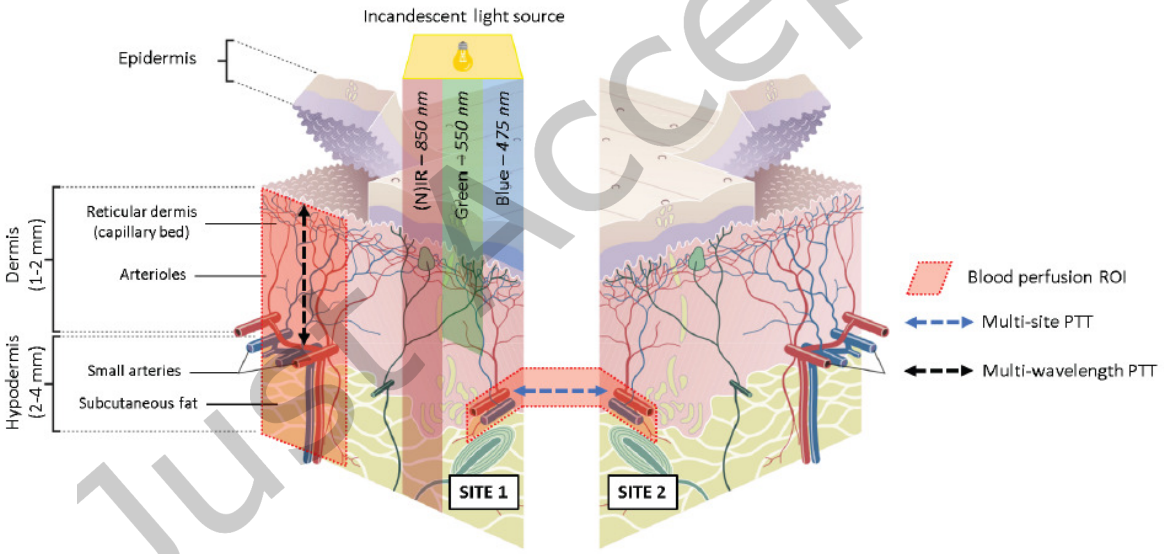

Fig. 1. The principle of the proposed multi-wavelength remote PTT monitoring compared with traditional multi-site PTT.

Fig. 1 also shows a layered structure of the skin microvasculature consisting of three layers: the topmost epidermis, which is mostly bloodless and consists of dead or dehydrated cells; the middle dermis containing capillary loops and small arterioles; followed by the deepest hypodermis containing subcutaneous fat and larger vessels. We are mainly interested in the dermis and hypodermis, as they contain majority of the vasculature that exhibits blood flow.

3.2 Camera Physics

Obtaining different wavelength information is an inherent property of RGB cameras, as each color pixel is sensitive to specific wavelengths. This is vital for obtaining delayed PPG waveforms from different skin depths. However, the image sensor of the camera is not perfect but is instead governed by the quantum efficiency of each wavelength pixels, which shows the relative amount of specific wavelength photons successfully registered by the image sensor pixels and translated to a digital signal. An example is shown in Fig. 2, which belongs to our specific image sensor, but also serves well for illustration purposes. The relative intensity of sensor response (photons registered and translated to digital signal) on the y axis is in the range [0, 1], where 0 means no photons are registered and 1 means all of them are. The x axis simply shows different wavelengths of light in the visible and near infrared (NIR) spectrum.

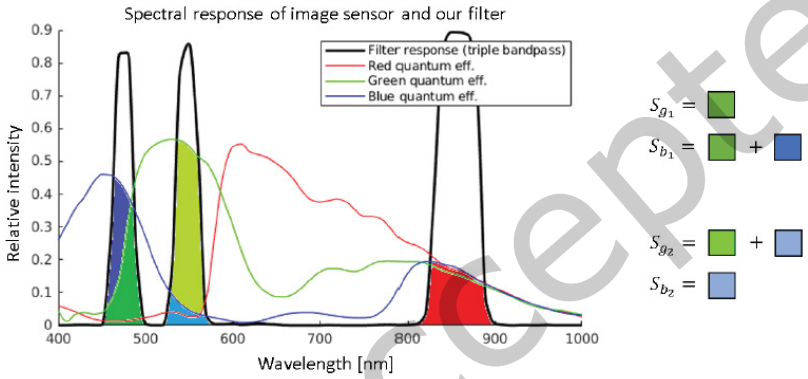

Fig. 2. Quantum efficiency of the image sensor of iDS 3040SE. Superimposed black line shows the triple band-pass filter response, which importantly helps with band overlapping. The red, green and blue lines correspond to R, G and B, while the color-coded areas under the curves S_g and S_b are used for computation of x and y in Eq. 3.

Each of the colored lines in Fig. 2 represents the relative response of the corresponding specific color pixels of the image sensor (red, green and blue) at different wavelengths. We can discern two important facts and corresponding challenges:

(1) **Channel response overlap**: Looking at each specific channel response (e.g., green), we see that its pixels also respond to wavelengths in the neighbouring bands (e.g., in blue and red wavelengths), meaning that the image and subsequent rPPG trace obtained from such a channel are an impure mixture of several channels. We can also observe that all three color-specific pixels respond in the NIR band around 850 nm. Subsequently we can hypothesise that the PPG waveforms from different wavelengths are hard to distinguish, since rPPG traces obtained from pixels of each wavelength are in fact a mixture, which must first be separated.
(2) **Lower relative response in the NIR**: The relative intensity of the pixel response falls substantially towards the longer wavelengths, reaching only around 20% quantum efficiency. This means that in order to obtain similar amplitudes and prominent systolic peaks in the NIR rPPG trace compared to other wavelengths, we must compensate by having higher energy in this part of the spectrum coming from the light source. Additionally, the band of the filter we use and describe in Section 4 is also wider in this range to compensate for the low quantum efficiency.

These challenges are fundamental to out-of-the-box cameras and vast majority of image sensors, and must therefore be understood and considered when dealing with MW approaches. We propose the following mathematical description of each color channel obtained from our camera equipped with a triple-bandpass filter, producing the quantum efficiency signature shown in Fig. 2:

$$\begin{aligned} R &= r \\ G &= g + x \cdot b + r \\ B &= b + y \cdot g + r \end{aligned} \tag{3}$$

where R, G and B are the red, green and blue values produced by the camera, while r, g and b are the corresponding channel-separated red, green and blue values. The parameters x and y are describing the ratio of the corresponding color present in mixture that must be determined.

To elaborate further, let us for example consider the G component from Fig. 2, looking only at the areas that the triple-bandpass filter passes through (under the black curve) – it comprises of the actual green response g in the band around 550 nm, but we can also observe non-trivial response of the green line in the blue band around 450 nm, as well as in the NIR band around 850 nm. The latter is (almost) equal to the $R = r$ response, while the former can be expressed as some fraction of the (channel-seperated independent) blue response in the form of $x \cdot b$. Some simplifications are made here, specifically we truncated the relative response under 0.1 to 0 (since the influence is very small) and we equalized all three responses in the NIR band, since the sensor responses are almost the same.

4 RECORDING SETUP

4.1 Camera

As discussed in Section 3, skin layers are thin at 1-2 mm and blood traversal velocity in capillaries expected around 12 mm/s, so the PTT measured between different skin layers is expected to be short at around 100 ms. Knowing this, we required a camera capable of high-frequency recording. Additionally looking at related work dealing with contact MW PTT measurement [18, 19], we discerned that our camera should be capable of recording at around 250 fps. This allows measuring PTT with precision of up to 4 ms, which is more than an order of magnitude shorter than expected PTT, so it should be able to capture PTT variations related to BP. Additionally, we wanted no on-board image processing since any changes to the pixel values might influence and distort the obtained rPPG waveforms. Following these requirements we opted to use the iDS 3040SE-Q RGB camera with the Sony IMX273 1/3" CMOS image sensor and iDS-5M23-C1618 16 mm lens for our experiments. This camera allows for variable frame rates of up to hundreds of fps while also offering programmatic access to the raw images as registered by the sensor. For our experiments we chose 250 fps due to aforementioned reasons.

As seen in Fig. 1 and shown by Moço et al. [21], capillary loops dominate the papillary dermis layers reachable by shorther wavelengths, while smaller arterioles and arteries are found in the deeper layers of the skin, so we wanted to use light that also reaches those depths. NIR in the wavelengths around 850 nm seems suitable for penetrating to deeper dermis layers, however, it is not registered by the image sensor of a traditional RGB camera out of the box, since they come equipped with a default IR filter with a cutoff at 650 nm. We thus modified our camera by first removing the on-board IR filter and then replacing it with a triple band-pass MidOpt TB475/550/850 filter, which allows only light in narrow bands of 475±10 nm (blue), 550±10 nm (green) and 850±22 nm (NIR) to pass. The band is wider in the NIR due to lower sensitivity of the image sensor in the NIR range, as seen in 2. Furthermore, human skin often exhibits sweating, especially when exposed to heat or after physical activity, which can result in specular reflections on the surface. These might distort our signal reconstructions, so we additionally used the MidOpt PR1000 VIS/SWIR Wire Grid Linear Polarizer which is effective in the range of 400–2000 nm.

The combination of these modifications gives us access to NIR information from deeper arterioles in the reticular dermis while also partially addressing the first challenge mentioned in Section 3 – channel separation. As the selected filter has narrow bands, it isolates the response of the pixels producing a "cleaner" image in terms of specific wavelength fidelity, albeit this does not solve the problem entirely, as explained in Section 3.

4.2 Light Source

To further address the second challenge listed in Section 3 – poor quantum efficiency in NIR part of the spectrum – we had to use a light source that emits the full spectrum that we are interested in, as well as (ideally) increases in energy towards the NIR. Daylight has a suitable spectrum, however, we wanted good control and consistency in our light source, so we used an artificial source instead. The spectrum requirement meant that conventional LED sources are not suitable, since they have distinct peaks in the blue and green part of the spectrum while emitting virtually no energy in the NIR. An obvious initial candidate was a traditional incandescent filament bulb, which has a suitable spectrum for our use case. Upon making initial recordings with a standard 230V 50W bulb and our high-frequency camera, we observed flickering in the recordings, which is a consequence of the 50 Hz alternating current. To circumvent this we instead used a direct current power source – a 12V 650W power supply unit – alongside Osram Decostar 51ALU filament bulbs. These bulbs are coated with thin aluminum coating and unlike their dichroic counterparts, the heat is fully emitted in front, meaning we get a good amount of (N)IR emission, compensating for the lower quantum efficiency in this part of the spectrum. We decided to use two 50W bulbs to ensure enough light to get reasonable visibility, as the exposure time is very short at such high fps. The light was directed towards the region of interest (ROI) of the skin and the bulbs were placed perpendicular to one another in order to negate any shadows created by the uneven surface of the recording ROI.

4.3 Blood Pressure Monitor

Ground-truth BP measurements were obtained with a clinical-grade Omron M10-IT cuff-based digital BP monitoring device in order to obtain precise and trustworthy measurements. The cuff was placed on the upper arm in accordance with the official guidelines. The whole setup with an anonymous example subject is shown in Fig. 3.

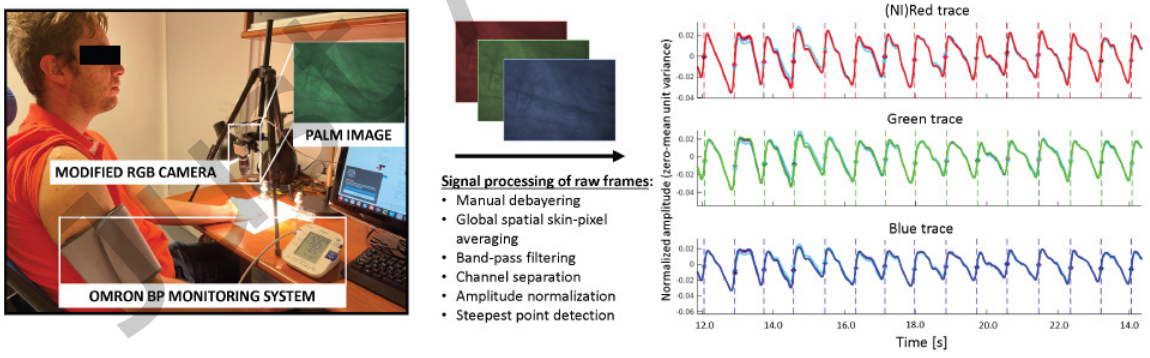

Fig. 3. Anonymous subject being recorded and video frames being processed to obtain per-channel rPPG traces. Overlapping lines of a single trace correspond to different Butterworth filter orders (1-4). Note that this photo is taken with ambient lighting to capture the recording setup, while the actual recordings for experimental data collection always take place without any ambient lighting to avoid light source interference.

4.4 Data Collection Protocol

The described recording setup was used to record the central part of the palm skin of test subjects, as seen in the example frame shown in Fig. 3. The palm was chosen for measurement since it is one of the places where skin exhibits the most pulsatility (similar to fingertip and forehead, which are the most commonly used sites) and is relatively comfortable to record. The spectral range between blue (450 nm) and NIR (850 nm) was selected as such light penetrates between 0.5 and 4 mm deep in human skin tissue, which is suitable accounting for variations in skin thickness due to calluses or other anomalies [3].

We obtained palm recordings from 9 subjects alongside their ground-truth BP measurements. Written consent for recording and data processing was obtained from all subjects. Seven were male and two were female, the total mean age was 32±3.4 years. Each subject was asked to report any potential existing health pre-conditions and 7 of them reported no existing pathologies or diseases, while also having normal body mass index. There were two outliers in terms of health status, since they were years-long cigarette smokers, while also being above the average age.

Since we were interested in very precise waveforms, and some related work shows potential influence of cuff compression on the waveform itself via blood vessel compression [20, 21], we recorded videos of the left palm while placing the cuff on the right upper arm to minimize the influence. Each recording lasted for 30 s, during which a single ground-truth measurement was taken, since the measurement itself takes more than 20 s. Thus a single SBP, DBP and HR ground truth is associated with the whole 30-sec recording, which is feasible since BP does not change a lot in 30 s outside of extreme circumstances (e.g., arterial bleeding or strong medication). Each subject was recorded several times on different days, following 2 scenarios:

- **Resting position:** In the first scenario each subject was equipped with a cuff while seated upright stationary in a chair. They were asked to close their eyes and relax for 1 minute. This allowed for the cardiovascular signals to stabilize, and afterwards they placed their palm under the camera. A video recording of their palm was then made simultaneously with the ground truth BP measurement.
- **Intense physical activity:** In the second scenario each subject was asked to perform 2 minutes of intense cardiovascular exercise consisting of jump squats followed by jumping jacks, without pause. This notably elevated their HR and was expected to raise their BP as well. Immediately after completion, they were seated in the same stable upright position with their arm immobilized on a desk support (to minimize waveform distortions and movement artefacts), the cuff was applied and the recording was made immediately. Since BP and HR do not drop that rapidly, this allowed for the capture of elevated vitals.

As mentioned, we repeated this recording sequence on different days in order to avoid some daily specifics of individuals (e.g., had a stressful meeting coming up), which might not reflect their typical hemodynamic state. In total we obtained 99 recordings from 9 subjects and first checked the SBP and DBP distributions using a Shapiro–Wilk test, which returned $p > 0.05$, for both cases, confirming it to be a normal distribution. SBP and DBP were on average increased by 22.3 and 11.0 mmHg respectively between rest and activity. Since we also have dependencies between measurements (a sort of before / after scenario), as we assume that subjects with higher BP during rest will also have higher BP during activity, we used the paired T-test to check for statistically significant differences between resting and post-activity scenarios. We used the standard significance level of 0.05, and obtained p-values of 0.01 and 0.001 for DBP and SBP respectively, successfully rejecting the null hypothesis that our rest and activity BP samples come from normal distributions with the same mean, thus confirming that the two scenarios are different in a statistically significant manner.

5 DATA PROCESSING METHODOLOGY

5.1 Frame Acquisition Software

Once our recording setup was finalized at the physical level, we had to ensure that we removed any on-board default image processing. We did this by developing a piece of recording software based on the official iDS SDK together with their Peak library. In short, we first set the fps to 250, which means that we had to make a trade-off by setting the exposure time to only 4 ms. We then turned all the white-balance corrections off and set the color gains to the default value of 1.0. At this point we obtained raw 3D matrices corresponding to each frame, which we then manually Debayered as given in (4):

$$R, G, B = \begin{cases} R \\ round(\frac{G_1}{2} + \frac{G_2}{2}) \\ B \end{cases} \tag{4}$$

where R are the (nearinfra)red pixels, G_1 and G_2 are the green pixels and B are the blue pixels in a standard half green, one quarter red and one quarter blue Bayer mosaic pattern. Doing this lowered the resolution of the final frames (due to halving of the green pixels) to 540 x 720 pixels.

5.2 Signal Processing

The sequence of frames was first globally spatially averaged (taking all pixels, since only skin is present in each frame) in order to obtain three temporal RGB traces. These traces were then zero-phase filtered with a 2nd order Buttherworth band-pass filter with cutoff frequencies of [0.5, 6.0] Hz. This removed low-frequency baseline wandering and eliminated high-frequency noise. Finally we normalized our signals amplitude to a constant range of [-1, 1]. At this point we obtained relatively clean traces with obvious cardiac pulsatility, as seen in Fig. 3. Methods used in this processing pipeline were universally applied to all signals, as they were carefully designed to not influence the waveforms in any way that would distort the relationship under investigation. More specifically, both the amplitude normalization and the zero-phase filtering were chosen specifically to be generally applicable and to not influence the temporal position of the systolic peak or the steepest systolic rise point, regardless of the subject.

In order to measure PTT, we opted to use a reference-point method, meaning we used a single robustly detectable stable point per cardiac cycle, between which we measure PTT in different waveforms. Specifically we chose the steepest point on the systolic slope, as this is often more robust and stable compared to more commonly used systolic peaks. The systolic rise is more apparent even in noisy signals, while systolic peaks are sometimes difficult to precisely determine even for state-of-the-art peak detection algorithms [11]. Stability is important in this case because even a slight temporal missdetection of a few samples can represent a large part of the very short PTT. We detected these reference points by using the derivative method, taking the more clearly-expressed peaks in the first derivative. As precision and stability of these reference points is vital, we cross-checked our detections with state-of-the-art systolic rise/peak detection algorithms proposed by Elgendi et al. and Han et al. [6, 9], which were specifically developed to tackle challenging conditions and cardiac anomalies respectfully. Since we ensured our signals were as clean as possible, by both stabilizing the measurement site, as well as using aforementioned algorithmic cleaning, all algorithms detected the same locations of reference points. Detected example reference points can again be seen in Fig. 3, represented by the star markers.

At this point of the pipeline, we can make an initial observation of the PTTs by looking at the steepest points locations in each channel, and then computing the difference between these. To do this, we first took the green channel and its corresponding steepest point detections as the reference, because green has been both historically and empirically (from our observations) identified – and also widely accepted in the field – as the trace with the

most expressed and stable pulsatility. We then checked a very narrow area before and after these reference points by defining a short threshold of 40 samples and searching for steepest point detections in the other two traces (red and blue). If such points were found, we computed the temporal difference (PTT), otherwise we ignored the cycle.

Looking at early results, we had two initial important observations. First, the waveforms were relatively clean and stable in NIR and green, while the blue often expressed very shallow and noisy pulsatility. Second, the PTTs were extremely short, lasting just a sample or two. Given the former, while also knowing that blue wavelengths have the shallowest penetration mostly reaching the upper less blood-perfused skin layers, and concerns expressed in related work about lower wavelengths [21], we decided to simply remove the blue channel from further analysis, since it was too unstable. Regarding the unobservable PTTs, this was not entirely unexpected, since we already highlighted the problematic inter-channel influence in previous sections, which was only partially resolved using the narrow-band triple band-pass filter.

5.3 Channel Separation and PTT Analysis

To further reduce the inter-channel influence, we first had to determine x and y from (3), which we computed from our image sensor response as given in (5):

$$\begin{aligned} x &= \frac{S_{g_1}}{S_{b_1}} \quad in \quad [450, 500]\ nm \\ y &= \frac{S_{b_2}}{S_{g_2}} \quad in \quad [520, 570]\ nm \end{aligned} \tag{5}$$

where S_{g_1} and S_{b_1} are the areas under the curve in the blue band of the filter, while S_{b_2} and S_{g_2} are the areas under the curve in the green band of the filter. This essentially tells us the amount of one color present in the other one, where the unwanted "incorrect" color is the numerator and the dominant "true" color is the denominator as shown in Fig. 2. For our specific camera and image sensor we obtained $x = 0.61$ and $y = 0.13$. Once x and y are obtained via definite integration of the curves in Fig. 2 bound by the filter bands, we can follow (3) to express channel-separated r, g and b traces as given in (6):

$$\begin{aligned} r &= R \\ g &= G - R - x \cdot b \\ b &= B - R - y \cdot g \end{aligned} \tag{6}$$

With some expression manipulation we can express the channel-separated traces as follows. First we express b by inserting g:

$$b = B - R - y \cdot (G - R - x \cdot b) \tag{7}$$

Inserting this back into g gives us:

$$g = G - R - x \cdot b \tag{8}$$

where all the variables are known and can be obtained from the system. The factor $\frac{1}{1-x \cdot y}$ is a camera-specific constant that only influences the amplitude, which was normalized, so we can omit it, giving us the final forms as:

$$\begin{aligned} r &= R \\ g &= G - R - x \cdot (B - R) \\ b &= B - R - y \cdot (G - R) \end{aligned} \tag{9}$$

The effects of this channel separation method are summarized in Table 1. We can observe that they are positive and align with our expectations, since the PTTs become more obvious and mostly appear in the expected order [10].

5.4 BP Correlation Analysis

We analyzed the correlations between the measured ground-truth SBP and DBP values, and PTTs in both recording scenarios to investigate the feasibility of BP estimation. We used absolute Pearson's correlation coefficient as the correlation metric. As there is a general consensus in the community that BP is subject-specific, our analysis was first done on a per-subject basis, where we were interested in both the actual correlation and the detection of the two subjects who reported overall higher BP and for whom we could reasonably assume to have stiffer vessels due to a history of smoking and higher age. However, since we only have a handful of samples for each subject, we cannot make statistically significant conclusions. Subsequently we also checked the correlation between SBP / DBP and PTT in general, looking at the whole population at once. Our aim was to investigate whether we find consistent correlation patterns on both general and subject level, separating the different elicited hemodynamics and corresponding BPs. Looking at general correlations on the full population, we can also make conclusions about statistical significance of our observations. We used the MATLAB internal computation of p-values in the `corr()` function, which checks the probability of getting a correlation as large as observed by random chance, when the true correlation is zero.

5.5 BP Prediction Analysis

As the ultimate goal (that is most valuable to medical professionals) is to numerically estimate BP, we also trained a Random Forest regression model to predict SBP and DBP. We used each of the recordings as one instance with the corresponding ground-truth SBP and DBP label. We then used the average PTT in the recordings as our first feature, and then extended the feature vector with additional features describing the morphological properties of the waveform in that recording. In accordance with related work [17] we decided to use the average cycle width (T_c), time from start of cycle to systolic peak (T_s), time from systolic peak to cycle end (T_d) and areas under the curve for each of these sections (AUC_c, AUC_s and AUC_d). We did not choose any features relating to the diastolic notch, as it is difficult to precisely detect even in a traditional contact setup, while in our case the waveform is additionally distorted due to the remote nature of the measurement. Finally, we also added age and sex as demographic features describing each individual.

We implemented a leave-one-subject-out (LOSO) evaluation framework with personalization, where in each iteration we use the data of $n - 1$ subjects for training and the left-out subject for testing. A small part of data belonging to the test subject (in our case one instance of rest and one instance of activity scenario) is added to the training data in each iteration to personalize the model. While we are aware that this is slight overfitting to the test subject, we decided to still use personalization, as consensus in the field is that a general BP estimation model usually works much worse compared to a personalized model [24], even when the data is obtained in a highly controlled setting with contact sensors. We trained a Random Forest regression model due to relatively low time complexity and good historical performance, and evaluated the performance for SBP and DBP individually using the Mean Absolute Error (MAE) metric.

6 RESULTS AND DISCUSSION

Our initial hypothesis based on related work [10] and physiological background described in Section 3 was that the cardiac wave from the deeper vessels in the NIR range should be observed first, while those in the shallower layers like papillary dermis should be delayed. We define these times as $PTT = timestamp\_steepest_{\text{green}} - timestamp\_steepest_{\text{NIR}}$, and we expect them to be positive, if in agreement with our hypothesis. Since we have a single ground-truth SBP and DBP for many cardiac cycles in a given 30-sec recording, we computed a single PTT that corresponds to one recording, as the average of all the detected PTTs. These PTTs between NIR and green for subjects in our experiment are shown in Fig. 4

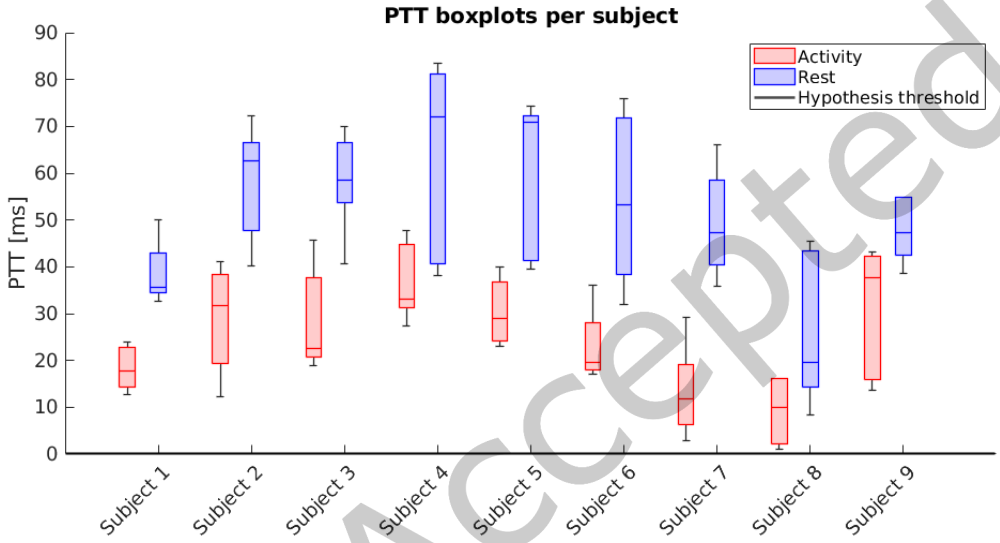

Fig. 4. Per-subject PTTs computed from reference points between NIR and green channel.

First we can observe that the PTTs are indeed consistently positive, thus in agreement with our hypothesis and related work [10]. Looking at the y axis, we can discern that the average PTT is also longer in the rest scenario, as the HR and BP are lower, SVR is higher, and blood traverses more slowly. Furthermore, these PTTs are in the expected range of 50-150 ms [4] and noticeably vary between subjects, which was not the case before applying channel separation. This confirmed that channel separation is important, both at the physical sensor level (narrow-band triple band-pass filter) as well as the signal processing level. Looking at subjects 7 and 8 – our outliers in terms of age and cardiovascular state – we can observe overall lower PTTs compared to all other subjects. This is again in agreement with our expectations of them having stiffer vessels, facilitating faster blood traversal and lower PTTs.

The second thing to validate are the effects of refined algorithmic channel separation. The average PTT between NIR and green across all subjects increases by 38.8 ms on average compared to average durations before applying our algorithm. Additionally, under the hypothesis of NIR steepest point appearing before the reference steepest point in green, the amount of cases in agreement increases by 19.3%. These statistics are summarized in Table 1.

The positive effect of channel separation is an increase in the PTTs on average, making them more easily observable and measurable. This effect can be observed when looking at the PPG waveforms in close detail (e.g., a couple of cycles), as shown in Fig. 5. The effect is especially prominent between the NIR and green waveforms, while the blue waveform has some morphological irregularities that make it challenging to extract meaningful

Table 1. Summarized statistics of refined algorithmic channel separation effects.

	Before	**After**	**Δ change**
Average PTT (green-NIR)	8.0 ms	46.8 ms	+38.8 ms
% of PTT in correct order (green-NIR)	79.4%	98.7%	+19.3%

reference points and physiological information. This is because the blue light with shorter wavelength has shallower penetration into skin tissue and the epidermis does not exhibit blood perfusion due to lack of vascular presence.

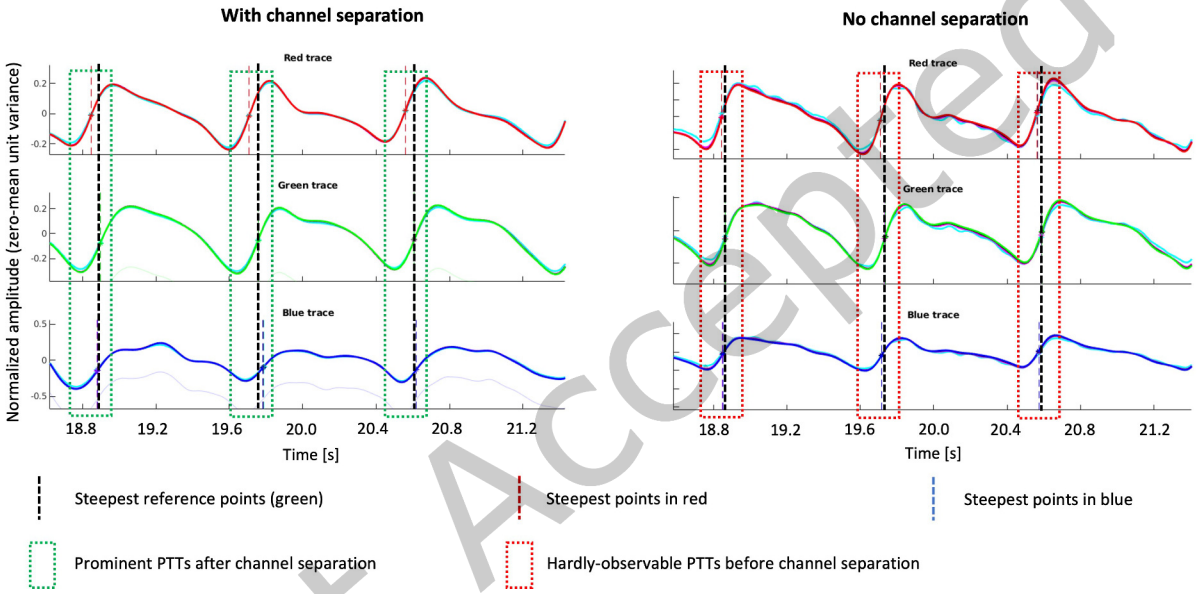

Fig. 5. Effects of channel separation on the PTTs. The left plot shows more prominent post-channel-separated PTTs observed between the dashed lines, while the right plot shows the barely noticeable PTTs before channel separation.

We then checked the general correlations between PTTs and BP on all our recorded samples across all subjects, which can be inspected in the scatter plot shown in Fig. 6. We also looked at the subject-specific correlations between PTTs and BP for both scenarios, however, even though these match in direction and value to those observed on all samples, they are statistically less significant due to the low number of samples.

In both cases, we can observe a clear separation between the two groups of measurements – red dots in the upper left corner of each plot correspond to higher BP and shorther PTTs, while blue dots in the bottom right corners correspond to lower BP and longer PTTs. The regression lines and correlation coefficients, show relatively high values of absolute correlations (R) between PTTs and BP, showing that these two biomarkers are indeed well-correlated, even when PTT is obtained in such a novel remote MW manner. If we consider the average correlation across all subjects in Fig. 6, we find higher correlations for SBP compared to DBP (0.61 vs. 0.54), which might be attributed to the fact that DBP is generally more consistent and changes less than SBP between rest and activity. For these correlations we obtained the correspoding p-values of 0.08 / 0.05 for SBP / DBP.

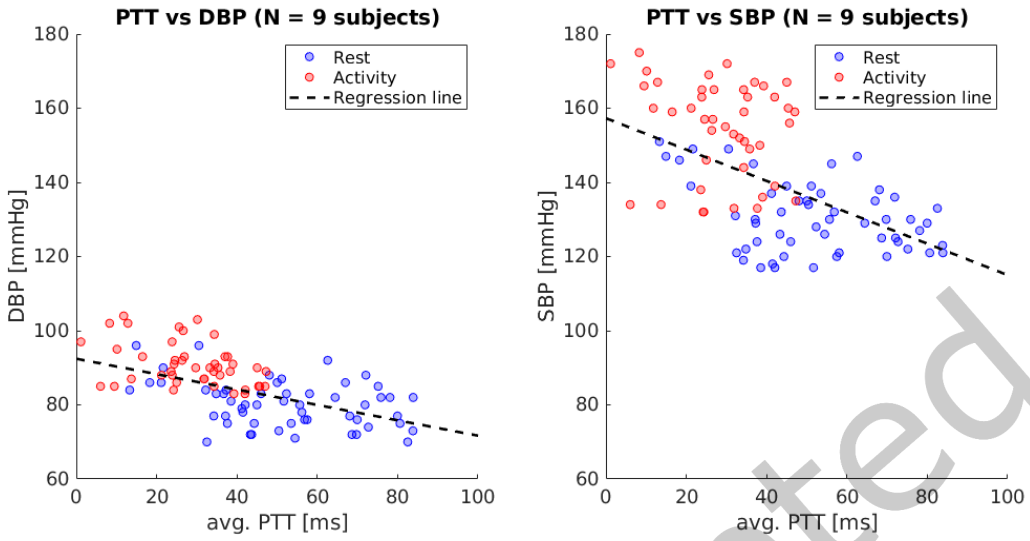

Fig. 6. Pearson's correlations and regression lines between PTT and DBP (left) / SBP (right) for all measurements of all subjects. PTT shown is obtained after the proposed channel separation.

Finally the predictive performance of our trained regression model in terms of MAE across all subjects is shown in Table 2.

Table 2. MAEs for SBP and DBP prediction using a Random Forest regression model in a leave-one-subject-out evaluation with personalization.

Subject	MAE SBP [mmHg]	MAE DBP [mmHg]
1	3.34	2.12
2	3.58	2.33
3	3.12	2.07
4	5.02	2.98
5	3.37	2.46
6	2.89	2.11
7	4.46	3.71
8	4.41	3.58
9	3.19	2.31
Avg.	**3.59**	**2.63**

We observed good stability between subjects, with the average MAEs of 3.59 mmHg and 2.63 mmHg, which surpasses the baseline that always predicts the mean by 15.61 and 9.73 mmHg for SBP and DBP respectively. The systolic error is higher, which could be attributed to larger variations in systolic pressure, while the diastolic generally changes less. We also observe a slight degradation of performance with subjects 7 and 8, which differ from others in terms of age and subsequently also their cardiovascular state (stiffer vessels). Investigating the feature importances, we found that the PTT feature dominates in importance compared to all others, achieving

mean decrease in impurity (defined as the total decrease in node impurity, weighted by the probability of reaching that node, averaged over all trees of the ensemble [1])of 0.68, while the second largest was only 0.05.

A comparison with related work is difficult as the approaches are fundamentally different (remote vs. contact, single-site vs. multi-site, PTT vs. PPG morphology), the models are often neural networks [5] that require more data and the experimental setup also differs. Despite this, our experiments confirm importance and robustness of PTT compared to morphological waveform features, show the value of personalization [24] while also highlighting that in case of PTT, very little personalization is needed for good performance, as using only 2 instances decreases the error on average by 3.52 and 2.41 mmHg for SBP and DBP respectively. Another important yet often overlooked aspect when dealing with black-box models (e.g., neural networks) is the lack of explainability or basis on known physical and physiological laws. While such models can yield lower errors, the question of overfitting to specific datasets and the reasoning for predictions is often ignored, while the PTT relationship to BP is well-established in literature [2].

7 CONCLUSION AND LIMITATIONS

A synthesis of our results brings us to the conclusion that single-site remote MW BP estimation with a modified consumer camera is feasible and could offer an alternative to current cuff-based or optical contact approaches. We list the main contribution alongside important limitations in the following sections.

7.1 Contributions

The main contributions are as follows:

- Remote MW PTT estimation is feasible and an improvement over traditional multi-site PTT monitoring, since it uses just a single site and sensor, omitting the need for precise synchronization or high-resolution cameras (e.g., when palm and forehead are captured on the same frame, which must have enough pixels of each to reconstruct the rPPG waveform). More importantly, it can circumvent problems with MW contact sensors, which were reported to influence the waveforms via skin compression [21].
- A pitfall of other work that should always be considered is the camera-specific inter-channel influence. This should be resolved both at the physical level (narrow-band filter) as well as signal processing level (the proposed channel separation method), as not accounting for this influence might lead to difficultly in detecting MW PTTs or incorrect conclusions based on their durations. Furthermore, using bands far-apart (e.g., green and NIR) produces longer and more easily observable PTTs.
- A relatively simple regression model can achieve decent performance with relatively low MAEs for both SBP and DBP, also confirming the importance of the PTT feature compared to other commonly used morphological PPG features. Additionally, only a couple of instances used for personalization improve the performance compared to a general model.

We can also conclude, in agreement with related work [21], that using the NIR band is preferred in MW PTT monitoring compared to traditional visible spectrum, since NIR light reaches deeper skin layers with arterioles, exhibiting better pulsatility compared to shallow layers reachable by blue wavelengths, while also increasing the length of the blood traversal path and delay between the waveforms. Feasibility of the proposed method was confirmed in controlled laboratory conditions, showing that usage of NIR and green wavelengths allows for acquisition of stable rPPG waveforms suitable for careful signal processing [18, 21] and subsequent PTT calculation. Despite this, some challenges remain before practical application, which are listed and discussed in detail in the following section.

7.2 Limitations

We acknowledge that a non-clinical experiment with 9 subjects cannot be used to make statistically significant general conclusions. A large-scale clinical validation would be required instead, but this requires access to more (varied) subjects and prolonged observations with continuous ground-truth BP. The former is always preferred to a cuff, as it offers much richer BP information, but is measured using an intravenous catheter, which can only be feasibly used with trained medical personnel. Such a large-scale clinical validation (in a hospital) is exceptionally challenging to organize and makes more sense for subsequent steps beyond a feasibility study. Ideally such an extensive evaluation would include hypotensive and hypertensive patients, both with mild (no medication) and more severe (using medication) cases.

We also made some simplifications in our work when doing channel separation – we approximated some signal contributions which might be refined further. Additionally, channel separation might be inaccurate when the light source spectrum does not cover the relevant spectral bands we are interested in. In future work we will consider strictly data-driven algorithms for channel separation that are independent of the quantum efficiency and light spectrum information, for instance using a genetic algorithm where the channels are initially a random linear combination of all three colors and the coefficients corresponding to each color are iteratively derived by using the error of a BP regressor as the fitness function. Another option would be to use a projection method, generating a densely populated space of candidate coefficients and try to find the best performing ones.

Furthermore, we were quite strict with removal of noisy data, since our interest was in fundamental feasibility rather than practical application. Signal noise (e.g., movement) still remains a challenge when considering widespread use. During our lab experiments, we circumvented this by ensuring a stable sitting position and fixed arm position immediately after the exercise finished. This combined with the processing pipeline allowed for very robust detection of reference points, however, the method used to detect those should always be verified for performance and stability when dealing with noisier waveforms, as it is vital for PTT computation. In practical BP screening applications it could be partially resolved with a similar setup using a foam padding on which a subject places their hand. The setup would also consist of an enclosure ensuring no interference of ambient lighting while providing a light source with the desired spectrum. Generally, sunlight has a very broad uniform spectrum containing all the bands we are interested in, so it could be used for our proposed method, however, using common LED bulbs is not suitable due to their poor spectrum lacking NIR and IR part. Still, an LED array containing diodes that also emit longer-wavelength light is feasible and relatively inexpensive to construct [12]. Such a setup could be used in a doctor's office, especially when screening many subjects (e.g., groups of school children). Even more feasible (due to required personalization) is home use for telemedicine (e.g., elderly people), as self-measurement of BP with a cuff-based device (most common in general population) is very impractical and cumbersome, potentially causing incorrect measurements or even refusal to do regular measurement.

Finally, as we decided to use the palm as the measurement site due to tissue properties, our work is palm-specific to a degree, although we expect these findings to hold for other well-perfused skin locations. This remains to be validated, as variations in both skin tissue (e.g., melanin content, thickness, etc.) and measurement location should be investigated. However, as our purpose was to validate the early feasibility of remote MW PTT measurement and subsequent BP estimation, we limited our prototype setup and scenarios to be robust and reproducible at the cost of variations in measurement sites.

We used a relatively simple regression model to check the feasibility of explicit BP prediction and get a baseline result to compare against related work. The evaluation framework was designed to use personalized calibration, meaning some small part of data of each subject was used in the training in order to personalize the regression model. While this is not ideal, the consensus in the community is that building a robust generalized model is not feasible due to subject-specific relationship between PPG features and BP [24], meaning that personalization or calibration is required (or at least desired) in all models for BP estimation, including those based on multi-site

PTT [14]. More specifically, there can be subject specifics in many biological properties, such as blood thickness, arterial wall thickness, elasticity, and other parameters related to PPG and BP. While early results seem promising, this evaluation is still in early stages and will be extended with more sophisticated models, additional features and other robust evaluation schemes.

Despite these limitations and further research opportunities, this work showed the importance of fundamental understanding of both human physiology and camera physics for contact-free single-site MW measurement, while highlighting and addressing important challenges using novel methods. We showed that it is feasible to estimate PTT in such a way and confirmed that it highly correlates with BP, offering potential for remote MW BP estimation, which could be a useful part of a potential telemedicine system, as one of the parts in the growing framework of AIoT.

ACKNOWLEDGMENTS

This work was supported by Jožef Stefan Institute, Southern University of Science and Technology, the Slovenian Research Agency (ARRS) as part of the young researcher PhD program, and by The General Program of National Natural Science Foundation of China (Grant No. 62271241). We would also like to thank all the participants in our experiments.

REFERENCES

[1] 2015. Feature importance in random forests. https://alexisperrier.com/datascience/2015/08/27/feature-importance-random-forests-gini-accuracy.html

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

[3] Caerwyn Ash, Michael Dubec, Kelvin Donne, and Tim Bashford. 2017. Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods. *Lasers in Medical Science* 32 (11 2017).

[4] Phillip Bedggood and Andrew Metha. 2021. Direct measurement of pulse wave propagation in capillaries of the human retina. *Optics Letters* 46, 18 (2021).

[5] Deivid Botina-Monsalve, Yannick Benezeth, and Johel Miteran. 2022. RTrPPG: An Ultra Light 3DCNN for Real-Time Remote Photoplethysmography. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[6] Mohamed Elgendi, Ian Norton, Matt Brearley, Derek Abbott, and Dale Schuurmans. 2013. Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions. *PloS one* 8, 10 (2013).

[7] Louise Finlayson, Isla RM Barnard, Lewis McMillan, Sally H Ibbotson, C Tom A Brown, Ewan Eadie, and Kenneth Wood. 2022. Depth penetration of light into skin as a function of wavelength from 200 to 1000 nm. *Photochemistry and Photobiology* 98, 4 (2022).

[8] Shrimanti Ghosh, Ankur Banerjee, Nilanjan Ray, Peter W Wood, Pierre Boulanger, and Raj Padwal. 2016. Continuous blood pressure prediction from pulse transit time using ECG and PPG signals. In *2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT)*. IEEE.

[9] Dong Han, Syed Khairul Bashar, Jesús Lázaro, Fahimeh Mohagheghian, Andrew Peitzsch, Nishat Nishita, Eric Ding, Emily L Dickson, Danielle DiMezza, Jessica Scott, et al. 2022. A real-time ppg peak detection method for accurate determination of heart rate during sinus rhythm and cardiac arrhythmia. *Biosensors* 12, 2 (2022).

[10] Sangjin Han, Donggeun Roh, Junyung Park, and Hangsik Shin. 2019. Design of multi-wavelength optical sensor module for depth-dependent photoplethysmography. *Sensors* 19, 24 (2019).

[11] Xiaochuan He, Rafik A Goubran, and Xiaoping P Liu. 2014. Secondary peak detection of PPG signal for continuous cuffless arterial blood pressure measurement. *IEEE Transactions on Instrumentation and Measurement* 63, 6 (2014).

[12] Jingjing Hu, Yunze He, Jie Liu, Min He, and Wenjin Wang. 2019. Illumination robust heart-rate extraction from single-wavelength infrared camera using spatial-channel expansion. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE.

[13] Josep Maria Solà i Carós. 2011. *Continuous non-invasive blood pressure estimation.* Ph. D. Dissertation. ETH.

[14] Kaito Iuchi, Ryogo Miyazaki, George C Cardoso, Keiko Ogawa-Ochiai, and Norimichi Tsumura. 2022. Remote Estimation of Continuous Blood Pressure by a Convolutional Neural Network Trained on Spatial Patterns of Facial Pulse Waves. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[15] 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).

[16] Fatema-Tuz-Zohra Khanam, Ali Al-Naji, and Javaan Chahl. 2019. Remote monitoring of vital signs in diverse non-clinical and clinical scenarios using computer vision systems: A review. *Applied Sciences* 9, 20 (2019).
[17] Yuriy Kurylyak, Francesco Lamonaca, and Domenico Grimaldi. 2013. A Neural Network-based method for continuous blood pressure estimation from a PPG signal. In *2013 IEEE International instrumentation and measurement technology conference (I2MTC)*. IEEE.
[18] Jing Liu, Bryan P Yan, Yuan-Ting Zhang, Xiao-Rong Ding, Peng Su, and Ni Zhao. 2018. Multi-wavelength photoplethysmography enabling continuous blood pressure measurement with compact wearable electronics. *IEEE Transactions on Biomedical Engineering* 66, 6 (2018).
[19] Jing Liu, Yuan-Ting Zhang, Xiao-Rong Ding, Wen-Xuan Dai, and Ni Zhao. 2016. A preliminary study on multi-wavelength PPG based pulse transit time detection for cuffless blood pressure measurement. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE.
[20] Andreia Vieira Moço, Sander Stuijk, and Gerard de Haan. 2016. Skin inhomogeneity as a source of error in remote PPG-imaging. *Biomedical optics express* 7, 11 (2016).
[21] Andreia Vieira Moço, Sander Stuijk, and Gerard de Haan. 2018. New insights into the origin of remote PPG signals in visible light and infrared. *Scientific reports* 8, 1 (2018).
[22] Tuukka Panula, Jukka-Pekka Sirkia, David Wong, and Matti Kaisti. 2022. Advances in Non-Invasive Blood Pressure Measurement Techniques. *IEEE Reviews in Biomedical Engineering* PP (2022).
[23] Daniel Ray, Tim Collins, Sandra Woolley, and Prasad Ponnapalli. 2021. A Review of Wearable Multi-wavelength Photoplethysmography. *IEEE Reviews in Biomedical Engineering* (2021).
[24] 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).
[25] Fabian Schrumpf, Paul Rudi Serdack, and Mirco Fuchs. 2022. Regression or Classification? Reflection on BP prediction from PPG data using Deep Neural Networks in the scope of practical applications. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[26] Asoke Talukder and Roland Haas. 2021. AIoT: AI meets IoT and web in smart healthcare. In *13th ACM Web Science Conference 2021*.
[27] Xuyu Wang, Xiangyu Wang, and Shiwen Mao. 2020. Deep Convolutional Neural Networks for Indoor Localization with CSI Images. *IEEE Transactions on Network Science and Engineering* 7, 1 (2020). https://doi.org/10.1109/TNSE.2018.2871165
[28] Xuyu Wang, Chao Yang, and Shiwen Mao. 2017. PhaseBeat: Exploiting CSI Phase Data for Vital Sign Monitoring with Commodity WiFi Devices. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. https://doi.org/10.1109/ICDCS.2017.206
[29] Bo Wei, Wen Hu, Mingrui Yang, and Chun Tung Chou. 2019. From real to complex: Enhancing radio-based activity recognition using complex-valued CSI. *ACM Transactions on Sensor Networks (TOSN)* 15, 3 (2019).
[30] Fokko P Wieringa, Frits Mastik, and Antonius FW van der Steen. 2005. Contactless multiple wavelength photoplethysmographic imaging: a first step toward "SpO2 camera" technology. *Annals of biomedical engineering* 33, 8 (2005).
[31] Zhicheng Yang, Parth H Pathak, Yunze Zeng, Xixi Liran, and Prasant Mohapatra. 2017. Vital sign and sleep monitoring using millimeter wave. *ACM Transactions on Sensor Networks (TOSN)* 13, 2 (2017).