

Agile mobile robotic platform for contactless vital signs monitoring

Hen-Wei Huang ¹, peter chai ¹, Claas Ehmke ¹, Gene Merewether ¹, Fara Dadabhoy ¹, Annie Feng ¹, Akhil John Thomas ¹, Canchen Li ¹, Marco da Silva ¹, Marc H. Raibert ¹, Edward W. Boyer ¹, and Giovanni Traverso ²

¹Affiliation not available

²Massachusetts Institute of Technology

October 30, 2023

Abstract

The COVID-19 pandemic has accelerated methods to facilitate contactless evaluation of patients in hospital settings. By minimizing unnecessary in-person contact with individuals who may have COVID-19 disease, healthcare workers (HCW) can prevent disease transmission, and conserve personal protective equipment. Obtaining vital signs is a ubiquitous task that is commonly done in-person. To eliminate the need for in-person contact for vital signs measurement in the hospital setting, we developed Dr. Spot, an agile quadruped robotic system that comprises a set of contactless monitoring systems for measuring vital signs and a tablet computer to enable face-to-face medical interviewing. Dr. Spot is teleoperated by trained clinical staff to facilitate enhanced telemedicine. Specifically, it has the potential to simultaneously measure skin temperature, respiratory rate, heart rate, and blood oxygen saturation simultaneously while maintaining social distancing from the patients. This is important because fluctuations in vital sign parameters are commonly used in algorithmic decisions to admit or discharge individuals with COVID-19 disease. Here, we deployed Dr. Spot in a hospital setting with the ability to measure the vital signs from healthy volunteers from which the measurements of elevated skin temperature screening, respiratory rate, heart rate, and SpO2 were carefully verified with ground-truth sensors.

Agile mobile robotic platform for contactless vital signs monitoring

Authors: Huang H*(1,2,3), Chai PR*(2,4,5,6), Ehmke C*(2), Merewether G*(7), Dadabhoy FZ (4), Feng. A (2), Li C (2), Thomas. AJ (2), da Silva M(7), Raibert M(7), Boyer EW(4,6), Traverso CG(1,3)

1. Department of Mechanical Engineering, Massachusetts Institute of Technology
2. The Koch Institute for Integrated Cancer Research, Massachusetts Institute of Technology
3. Division of Gastroenterology, Department of Medicine, Brigham and Women's Hospital
4. Department of Emergency Medicine, Brigham and Women's Hospital
5. Department of Psychosocial Oncology and Palliative Care, Dana Farber Cancer Institute
6. The Fenway Institute
7. Boston Dynamics

* Contributed equally; corresponding: cgt20@mit.edu, ctraverso@bwh.harvard.edu

Abstract:

The COVID-19 pandemic has accelerated methods to facilitate contactless evaluation of patients in hospital settings. By minimizing unnecessary in-person contact with individuals who may have COVID-19 disease, healthcare workers (HCW) can prevent disease transmission, and conserve personal protective equipment. Obtaining vital signs is a ubiquitous task that is commonly done in-person. To eliminate the need for in-person contact for vital signs measurement in the hospital setting, we developed Dr. Spot, an agile quadruped robotic system that comprises a set of contactless monitoring systems for measuring vital signs and a tablet computer to enable face-to-face medical interviewing. Dr. Spot is teleoperated by trained clinical staff to facilitate enhanced telemedicine. Specifically, it has the potential to simultaneously measure skin temperature, respiratory rate, heart rate, and blood oxygen saturation simultaneously while maintaining social distancing from the patients. This is important because fluctuations in vital sign parameters are commonly used in algorithmic decisions to admit or discharge individuals with COVID-19 disease. Here, we deployed Dr. Spot in a hospital setting with the ability to measure the vital signs from healthy volunteers from which the measurements of elevated skin temperature screening, respiratory rate, heart rate, and SpO₂ were carefully verified with ground-truth sensors.

Introduction:

Pandemic spread of the novel coronavirus, SARS-CoV-2, and resultant COVID-19 disease is changing the methods in which patients are evaluated in the hospital setting. Abnormalities in vital signs are key indicators of severity of COVID-19 disease. Physiologic findings of tachycardia, fever, tachypnea, and hypoxemia can be the first signs of clinical deterioration and play a major role in risk stratification, treatment algorithms, and disposition of COVID-19 patients [1]. Identification of changes in vital signs help healthcare workers (HCW) recognize those who are sickest and in need of urgent interventions. Obtaining in-person vital signs raises several challenges in the COVID-19 pandemic. The use of standard cutaneous monitors requires the HCWs to closely interact with patients who are potentially infected with COVID-19 and puts them at risk of acquiring SARS-CoV-2 [2], [3]. Fluctuating supplies of personal protective equipment (PPE) may limit the availability of adequate PPE for HCWs obtaining vital signs. In practice, efforts to conserve PPE may lead to vitals being documented less frequently during a patient's stay. Finally, surging infection rates among HCWs may lead to a lack of qualified personnel to perform these tasks. Cumulatively, this may result in undetected deterioration of individuals who develop vital sign abnormalities. "Could robots be effective resources in combating COVID-19?" was asked at the beginning of the pandemic as mobile robots have historically removed humans from monotonous, contaminated, and dangerous

environments [4], [5]. Our team endeavored to address several critical needs of the pandemic by mitigating the risk of exposure to HCWs by deploying an agile quadruped robot that approaches patients to acquire vital signs in a contactless manner.

Different kinds of contactless monitoring systems using radio signals [6], [7] and radar-based sensors [8] have been investigated in the past decade. These systems can easily obtain respiratory rate (RR) and heart rate (HR) from multiple people without interfering with their daily activity, but are unable to capture other vital signs relevant to COVID-19 disease, such as elevated skin temperature and blood oxygen saturation (SpO₂). In order to screen for fevers from other infectious disease epidemics, commercial infrared (IR) camera systems have been demonstrated to reliably screen individuals for fevers in indoor commercial settings like airports [9]. Similar systems using Red-Green-Blue (RGB) cameras can extract HR [10], SpO₂ [11], [12] and blood pressure [13] from skin RGB pixel changes using recorded color image of human skin surfaces. This is known as remote photoplethysmography (rPPG) and can be achieved with consumer-level cameras. Advances in computer vision (CV) and machine learning enable automatic tracking of the region of interest (ROI) even from a crowd with people wearing a mask. Combined these systems offer the ability to generate contactless vital signs measurements on a mass scale to rapidly detect abnormalities that may be consistent with COVID-19 disease. With an increasing need for solutions to screen individuals who return to work, travel from areas of high viral transmission, and participate in regional and country level reopening during the COVID-19 pandemic, contactless camera systems offer a simple, noninvasive and scalable system to obtain incontrovertible proof of vital signs abnormalities.

In this work, we developed a robotic-assisted vital sign acquisition platform to facilitate contactless vital signs in hospital settings (VitalCam). At the heart of this system is a robot-controlled IR and multi-monochrome camera setup that automatically tracks individuals and obtains their facial skin temperature, HR, RR, and SpO₂ on an operator-friendly platform that can be utilized by HCWs. We deployed VitalCam as payload on Spot, a quadruped robotic system developed by Boston Dynamics [14], and describe the generation of algorithms to simultaneously obtain vital signs central to COVID-19 disease evaluation.

Background

Medical tent for COVID-19 triage

In order to cohort and maintain distance between individuals with potential COVID-19 disease, many hospitals have developed outdoor triage tents that facilitate SARS-CoV-2 testing and evaluation of low risk individuals. We deployed a large 25 x 45 foot triage tent at a large academic, urban medical center in the greater Boston metropolitan area (Brigham and Women's Hospital) (Figure 1). Ambulatory individuals, otherwise well appearing, who presented to the emergency department (ED) with symptoms consistent with COVID-19 disease (upper respiratory infection, fevers, or other exposure to COVID-19) were triaged to the tent for initial evaluation. Participants underwent a brief-nurse driven interview after which they were seated in the tent waiting area which comprised ten chairs spaced six feet apart. Patients then proceeded to a separate, semi-private space, within the tent, where they met a clinician who conducted a brief, scripted interview regarding COVID-19 exposure and current symptoms. Additionally, the clinician gathered a full set of vital signs (body temperature, HR, RR, SpO₂, and blood pressure) using standard equipment. The clinician then decided if the patient required additional care within the ED, or if they can be tested and discharged from the tent.

Evolution of Dr. Spot

We collaborated with Boston Dynamics (Waltham, MA) to deploy the Spot robot (Dr. Spot) in the triage tent, ED waiting room, and ED rooms for individuals. Dr. Spot was initially equipped with an IR camera (FLIR A325) for the purpose of fever screening and RR measurement (Figure 1b and Figure 2a). Dr. Spot was remotely controlled (RC) by a trained HCW stationed in the ED to travel to and from the tent and within the ED. The overall objective was to use Dr. Spot to facilitate acquisition of vital signs and a brief interview in order to reduce exposure of HCWs to patients. The main advantage of a mobile robotic-controlled IR camera is the ease of transport of the camera system.

Traditionally, an IR camera is fixed on a stand and meticulously calibrated to a certain distance. The patients undergoing screening for fever must stand at a specified location and directly face the camera. Instead of asking patients to adapt to a static camera, a robotic system with agile movement can adjust its distance and angle of view to obtain adequate frames of reference to obtain vital signs from patients. Moreover, an iPad mounted on Dr. Spot allowed clinicians to interview patients via secure video conferencing (Figure 2b). In this paper, we focus on the evolution of algorithms to facilitate contactless vital signs using Dr. Spot.

In order to simultaneously measure heart rate, respiratory rate, skin temperature oxygen saturation, we added three monochrome cameras with filters of different wavelengths to enable motion robust vital sign collection (Figure 2c). These three monochrome cameras were arranged together and outfitted with different optical filters with the wavelength of 660 nm, 810 nm, and 880 nm. As a result, the the VitalCam carried by Dr. Spot was able to simultaneously obtain skin temperature, RR, HR, and SpO₂ from an individual patient. Two types of user-friendly remote-control interfaces were developed to facilitate acquisition and display of vital signs. One was an RQT-GUI of ROS-based interface featuring video streams from the IR camera and monochrome cameras, signal tracing curves of the RR and HR, and the values of each vital sign (Figure 3a). The other was presented on a handheld controller with a touch screen which displayed the IR streaming video and the vital sign values (Figure 3b).

IR camera for detecting vital signs relevant temperature variation

Fever screening

Infrared (IR) thermography can detect elevated skin temperature which may indicate the presence of a fever. Because fever is a characteristic vital sign abnormality found in individuals with COVID-19 disease, thermal imaging systems are frequently used to screen large crowds for temperature derangements. There are several weaknesses in this approach. First, IR cameras measure the skin temperature distribution, but not the body temperature. There may be a wide variation of skin temperature distribution compared to the core body temperature which is used to define fever. Second, the correlation between skin temperature and core body temperature is poorly understood [15], [16]. It is well established that temperature of the inner canthus of the eye (tear duct) correlates strongly to core body temperature and is described as the most reliable location for temperature measurement on the face [17]. Yet accurate acquisition of temperature in this region of the body requires high resolution camera systems to impart adequate pixels for image processing. Unlike RGB cameras, most IR cameras deployed for elevated body temperature screening have a resolution around 320 x 240 pixels or 640 x 512 which is suboptimal for measuring temperature at the ocular canthus, especially at a distance greater than 2 meters. An IR camera has to be placed very close, less than 1 meter, to a measuring subject to ensure there are enough pixels to cover this area. We observed that there is about 1 °C drop on the highest face skin temperature when the subject is moved from 0.6 m from the camera to 2 m from the camera at the ambient temperature of 19 °C (Figure 4a). This

temperature drop is not only a function of the distance but also the ambient condition. The lower the ambient temperature and the farther the measuring subject are, the higher the variation in observed skin temperature.

As it is almost impossible to capture accurate ear duct temperature for fever screening while adhering to social distancing guidelines, the IR camera manufacturer FLIR suggests establishing a baseline by scanning and saving readings from ten known healthy individuals coming from similar ambient conditions [18]. Readings from all future subjects at the same ambient temperature scanned will be compared to this population baseline. Subjects with facial skin temperature higher than the baseline will be asked to undergo further diagnostic evaluation. The success of this approach relies on a calibrated temperature reference that needs to be placed in the field of view during the measurement. The reference source, also referred to as a “black body” allows the imaging software to calibrate the scene to a more precise value by providing a higher accuracy reference, than the camera is capable of. In addition, the black body needs to be located at the same distance from the camera, as the subject’s face that is being imaged. Although a black body addresses the inaccuracy of skin temperature readings using IR cameras at a fixed distance, implementation may be challenging in a dynamic changing indoor environment, such as an emergency department of a hospital. Moreover, the need to maintain a specific, static distance between the camera and the black body constrains the mobility of the robot. To remove the black body from the fever screening setup, we proposed a compensation algorithm by counting the distance and ambient temperature effect in the real-time measurement. As shown in Figure 4a, we observed that decrease in temperature with increase in distance follows a linear trend. The slope of the linear trends is also a function of ambient temperature (Figure 4b). Thus, we hypothesized that the difference in skin temperature reading at various distances could be rectified by doing inverse analyses of Figure 4.

Face detection with ROI

To enable real-time skin temperature calibration based on the inverse analyses, the distance between the camera and measuring subject and ambient temperature around the camera have to be acquired during the measurement. The ambient temperature was acquired by using a temperature sensor (BME280, Bosch). To accurately estimate the distance between the camera and subject, we employed a face detection algorithm known as InsightFace that continuously detects and tracks facial features through real-time IR imaging [19]. The raw thermal frame is rescaled to an 8-bit depth with the corresponding range of [0,255], then duplicated on the RGB channels before it is run through the InsightFace model. The model worked well on the IR images since it was trained on RGB images. A face bounding box is given to the detected face with the dimensions that are directly associated with the head size and the distance to the camera. Figure 5 shows that the face bounding box obtained from face detection algorithm monotonically decreases with the increase of distance between the camera and the subject. We measured the face bounding box dimensions from three different subjects with different genders and head sizes and observed the three curves overlap entirely with each other, which gave us the confidence to estimate the distances based on face bounding box dimensions. Through the inverse analyses of the results from Figure 5, we could estimate the distance by

$$D = 140.22 \cdot L^{-1.14}, \quad (1)$$

where L is the diagonal length of the detected face bounding box. With the feedback information of the distance and ambient temperature, the skin temperature reading at varying distance can be rectified (Figure 6a) and represented as

$$T_{rectified} = T_{IR} + T_{ambient}^{-2.774}(1900 \cdot D + 3700), \quad (2)$$

where T_{IR} is the temperature directly from the IR camera and $T_{ambient}$ is the ambient temperature. Figure 6b shows the real-time IR images marked with the original skin temperature reading, estimated distance, and the rectified skin temperature based on the distance compensation. The person standing at the distance of 0.7 m was measured with the original skin temperature of 35.85 °C and the rectified temperature of 35.88 °C. While the same person at a distance of 3.3 m with the skin temperature dropping to 34.77 °C that is rectified to be 35.65 °C. As a brief summary here, the VitalCam-enabled Dr. Spot can perform the elevated skin temperature screening without using a static black body setup for distance calibration and the measuring distance no longer constrains the screening. We believe the proposed algorithm could largely facilitate the multi-person screening in a crowded environment.

Tachypnea detection

Universal masking or application of a face cover that encloses the nasopharynx has been demonstrated to reduce the spread of SARS-CoV-2. Many countries and healthcare systems have implemented universal mask rules to mitigate disease transmission [20]. Figure 6c shows that there is a sharp temperature contrast in the IR images between wearing and not wearing a mask. We employed the temperature difference between the ROI of forehead and mask to identify if one is wearing a mask or not. If the temperature difference between the ROI of forehead and the ROI of mouth is higher than 3 °C, then the algorithm identifies the person is wearing a mask.

In the meantime, we also observed that wearing a mask created a periodic temperature variation in the IR images through inhalation and exhalation (Figure 7a(i)). Thus, we used this temperature variation to track the RR of a subject. First, the face bounding box was equally split into two parts, an upper region and a lower region. We then took the temperature from the central lower region of the face bounding box around the nose and the mouth to calculate RR. In each frame, our algorithm got the temperature average on the ROI, which was decided by the face entity's bounding box and corresponding facial landmarks. After accumulating enough (in our case: 128) frames, the algorithm first used cubic spline to sample (timestamp, temperature average) tuples with the same time interval from the original data, then calculates the periodogram for the samples, and estimates the RR by the frequency with the maximum power spectral density. It took about 4 s to acquire the RR with the camera which has a frame rate of 30 fps. A low pass filter was applied to remove the high frequency band noise and the waveform of breathing is shown in Figure 8a. To extract RR from the obtained temperature variation on the mask, Fast Fourier Transform (FFT) is applied onto the raw breathing signal shown in Figure 7a(i) and the frequency spectrum is shown in Figure 8b.

Monochrome cameras for detecting PPG

Tachycardia detection

rPPG is a simple yet low-cost optical technique that can be used to measure blood volume changes underneath facial skin via a consumer-level camera. The light absorption characteristics of bloodstream haemoglobin exhibit a strong peak at the wavelength between 500-600 nm, which corresponds to the frequency band of green light signal captured by an RGB camera. Normally rPPG is very sensitive to the presence of motion and noise artifacts. To enable a motion robust rPPG, De Haan et al. presented a method, the "PBV-method", which introduces the unique "signature" of the blood volume pulse signal [21]. De Haan et al. showed that the optical absorption changes induced by blood volume variations in the skin occur with a very specific vector in a normalized RGB-space (referred to as \vec{P}_{bv}). This unique blood

volume signature enables robust rPPG pulse extraction that minimizes the contribution to the pulse-signal of color variations with other signatures. In this work, instead of directly using single RGB camera to obtain the pulse signal, we used three monochrome cameras with filtered wavelength at 660 nm, 810 nm, and 880 nm, respectively, for the purpose of facilitating SpO₂ analyses. A general overview of the operating principle of the VitalCam is shown in Figure 7a. A patient is getting monitored while sitting on a chair. The four-camera system controlled by the agile mobile robot would point to the face of the subject and focus on the ROIs. The measurements are taken for a duration of about 10 s in which the subject has to sit still and should reduce the movement of his/her head. The procedures of getting RR, HR, and SpO₂ are described in Figure 7b. In the HR and SpO₂ estimation algorithm, the ROI is located at the forehead. The ROI is split into twelve equally-sized subregions to exploit the spatial redundancy of the camera sensor.

To obtain HR, our goal is to retrieve the desired pulse signal \vec{S} from C_N which is the matrix with time-variant input values from the camera channels. Here we use three camera channels, so the dimension of C_N is $3 \times N$, where N is the number of measurements in the time window. The pulse-signal \vec{S} can be constructed as

$$\vec{S} = \vec{W} C_N, \quad (3)$$

where \vec{W} , dimension 1×3 , is the weighing matrix with $\vec{W} \vec{W}^T = 1$. Given C_N as the input from the three cameras, the aim is to construct the weights \vec{W} using the PBV method to finally calculate the pulse signal. In addition, the pulse signal \vec{S} can be related to \vec{P}_{bv} by Equation (4).

The relative pulsatile amplitudes in the three channels are known based on physiology and optics that construct the pulse-signal \vec{S} . There exists the following correlation between the pulse signal and the normalized channels equals \vec{P}_{bv} .

$$\vec{S} C_N^T = k \vec{P}_{bv} \Leftrightarrow \vec{W}_{PBV} C_N C_N^T = k \vec{P}_{bv}, \quad (4)$$

Thus, the weights \vec{W}_{PBV} can be calculated using:

$$\vec{W}_{PBV} = k \vec{P}_{bv} Q^{-1} \text{ with } Q = C_N C_N^T, \quad (5)$$

where scalar k is chosen to ensure that \vec{W}_{PBV} is normalized in l_2 -norm sense. Thus, to compute the pulse signal, the weights must first be determined by the vector \vec{P}_{bv} and input from the camera channels C_N .

To extract the cardiac pulse-signal \vec{S} , the blood volume pulse vector \vec{P}_{bv} has to be known. \vec{P}_{bv} with dimension of 1×3 describes the relative pulsatile amplitudes in the channels of a camera. It is defined by:

$$P_{bv}^i(SpO_2) = \int_{\lambda=400 \text{ nm}}^{1000 \text{ nm}} PPG(SpO_2, \lambda) I(\lambda) Cam^i(\lambda) Filter^i(\lambda) d\lambda \text{ for } i = 1, \dots, n, \quad (6)$$

where $I(\lambda)$, $Cam^i(\lambda)$, and $Filter^i(\lambda)$ are the illumination spectrum, the camera sensitivity and the transmission spectrum of the filter, respectively. $PPG(SpO_2, \lambda)$ is the PPG waveform, which is experimentally determined function dependent on the SpO₂ level and light wavelength. The PBV method is applied to each subregion (i) to get $n=12$ pulse signals for each PBV vector. Figure 8c shows the averaged pulse signal from a healthy volunteer after applying a bandpass filter between 50 bpm and 150 bpm. The pulse signal was recorded for one minute after vigorous activity to approximate tachycardia observed clinically in patients with COVID-19.

After applying FFT to the pulse signals, the HR was retrieved by choosing the frequency which corresponds to the highest amplitude in the frequency spectrum (Figure 8d).

Hypoxia detection

An important property of the PBV method is that it utilizes the relative pulsatile amplitudes in the camera channels to differentiate variations in blood volume from variations from other sources such as motion. We can exploit the observation that SpO₂ affects the pulsatile amplitudes of the channels (and thus the examined \vec{P}_{bv} vector) to measure blood oxygenation values. SpO₂ values depend on the absorption spectra of HbO₂ and Hb, and different SpO₂ values lead to different PPG amplitude spectra. The PPG spectra for 70, 80, 90, and 100 percent SpO₂ are visualized in Figure 9. When SpO₂ decreases, the pulse amplitude at 660 nm increases whereas the amplitude at 880 nm decreases. Since the amplitude of 810 nm does not change as it is close to the isosbestic point of (oxy-)hemoglobin, the differences in amplitude decrease between the three channels.

An adaptive PBV (APBV) method is applied to each subregion. The pulse signal from each subregion is compared to the one of other subregions for quality analysis, and pulse signals with the highest quality are further analyzed and traced to an SpO₂ level. The PPG amplitude spectrum, $PPG(\lambda)$ in Equation (6) is the SpO₂-dependent term. Thus, the \vec{P}_{bv} vector would vary with the SpO₂ values. Since the PPG spectrum is partly determined by a linear mixture of the spectra of oxygenated and deoxygenated hemoglobin, the collection of examined \vec{P}_{bv} vectors can be expressed by the APBV model:

$$\vec{P}_{bv}^e = \vec{P}_{bv}^s + \alpha \vec{P}_{bv}^\mu, \quad (7)$$

where \vec{P}_{bv}^e is the examined vector, \vec{P}_{bv}^s is the static vector corresponding with 100 percent SpO₂, $\alpha=100-\text{SpO}_2$ is the variable gain factor, and \vec{P}_{bv}^μ is the update vector, which describes the relative SpO₂ contrast, or in other words, the change in pulsatile amplitude as a function of SpO₂ value. The specific static and update vectors we used for the three monochrome cameras were:

$$\vec{P}_{bv}^s = \begin{pmatrix} 1.00 \\ 0.56 \\ 0.42 \end{pmatrix} \text{ and } \vec{P}_{bv}^\mu = \begin{pmatrix} -0.021 \\ 0.0013 \\ -0.00032 \end{pmatrix}, \quad (8)$$

As observed in Equation (6), the values of the \vec{P}_{bv}^e vector depend on the selected wavelengths and the optical characteristics of the camera, and the PPG amplitude spectra can be linearly interpolated within a clinically relevant range of blood oxygenation levels. By linear interpolation, we varied SpO₂ levels from 70 to 100 percent. These SpO₂ levels yield to \vec{P}_{bv}^e vectors that are applied to equation (5). The resulting weight matrices are then used to compute corresponding pulse signals in equation (3). To get the SpO₂ level, the signal quality of these pulse signals is examined. A quality measure, determined by cross-spectral signal-to-noise ratio (SNR) and spectral peak correspondence, for each subregion's pulse signal is calculated to prune distorted regions. Lower quality signals are discarded, and the remaining signals are further analyzed. The pulse signal with the highest SNR is traced back to a specific PBV vector, which is traced back to an SpO₂ level by equation (7). This SpO₂ level is the estimated blood oxygen saturation value. The step resolution of our current model for SpO₂ estimation is 5%. As all the tested subjects are healthy volunteers, it is not possible to induce more than 5% decrease in the SpO₂ level from the tested subjects. We obtained SpO₂ results of 100% from four healthy volunteers with the ground truth values ranging from 96% to 100%. Our next step

in the near future will focus on verifying the accuracy of the vital sign measurements from potential COVID-19 patients.

Conclusion:

We developed the VitalCam system to reliably facilitate contactless acquisition of vital sign parameters central to triaging and managing individuals with COVID-19 disease. The VitalCam system stands to not only conserve PPE but also curb transmission of infection by helping clinical staff to detect key vital sign abnormalities in a contactless manner. Our work, outlined above, demonstrates that a multicamera system comprising IR and monochrome cameras mounted on an agile mobile robot can successfully and reliably deliver vital sign measurements while navigating in complex clinical environments and maintaining safe distances. This platform can be deployed and scaled on a mobile robotic system to acquire important biometric data in various care scenarios during the COVID-19 pandemic.

Acknowledge:

The authors would like to thank Andrew Tsang, Seth Davis, Joy Hui, Mike Grygorcewicz, Kim Ang, and Nick Sipes from Boston Dynamics.

Reference:

- [1] A. Haimovich *et al.*, “Development and validation of the COVID-19 severity index (CSI): a prognostic tool for early respiratory decompensation,” *medRxiv*, p. 2020.05.07.20094573, May 2020.
- [2] M. L. Ranney, V. Griffeth, and A. K. Jha, “Critical supply shortages - The need for ventilators and personal protective equipment during the Covid-19 pandemic,” *New England Journal of Medicine*, vol. 382, no. 18. Massachussetts Medical Society, p. E41, 30-Apr-2020.
- [3] J. D. Byrne *et al.*, “Injection Molded Autoclavable, Scalable, Conformable (iMASC) system for aerosol-based protection: a prospective single-arm feasibility study,” *BMJ Open*, vol. 10, no. 7, p. e039120, Jul. 2020.
- [4] G. Z. Yang *et al.*, “Combating COVID-19-The role of robotics in managing public health and infectious diseases,” *Science Robotics*, vol. 5, no. 40. American Association for the Advancement of Science, p. 5589, 25-Mar-2020.
- [5] F. Lanza, V. Seidita, and A. Chella, “Agents and Robots for Collaborating and Supporting Physicians in Healthcare Scenarios,” *J. Biomed. Inform.*, vol. 108, p. 103483, Jun. 2020.
- [6] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, “Smart Homes that Monitor Breathing and Heart Rate,” in *ACM Conference on Human Factors in Computing Systems (CHI)*, 2015.
- [7] S. Yue, H. He, H. Wang, H. Rahul, and D. Katabi, “Extracting Multi-Person Respiration from Entangled RF Signals,” *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 2, pp. 1–22, Jul. 2018.
- [8] M. Mercuri, I. R. Lorato, Y. H. Liu, F. Wieringa, C. Van Hoof, and T. Torfs, “Vital-sign monitoring and spatial tracking of multiple people using a contactless radar-based sensor,” *Nat. Electron.*, vol. 2, no. 6, pp. 252–262, Jun. 2019.
- [9] Y. Nakayama, G. Sun, S. Abe, and T. Matsui, “Non-contact measurement of respiratory and heart rates using a CMOS camera-equipped infrared camera for prompt infection screening at airport quarantine stations,” in *2015 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications, CIVEMSA 2015*, 2015.
- [10] M. Z. Poh, D. J. McDuff, and R. W. Picard, “Advancements in noncontact, multiparameter physiological measurements using a webcam,” *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.
- [11] M. van Gastel, W. Verkruysse, and G. de Haan, “Data-driven calibration estimation for robust remote pulse-oximetry,” *Appl. Sci.*, vol. 9, no. 18, 2019.
- [12] H. Rahman, M. U. Ahmed, and S. Begum, “Non-Contact Physiological Parameters Extraction Using Facial Video Considering Illumination, Motion, Movement and Vibration,” *IEEE Trans. Biomed. Eng.*, vol. 67, no. 1, pp. 88–98, 2020.
- [13] H. Luo *et al.*, “Smartphone-based blood pressure measurement using transdermal optical imaging technology,” *Circ. Cardiovasc. Imaging*, vol. 12, no. 8, pp. 1–10, 2019.
- [14] “Spot® | Boston Dynamics.” [Online]. Available: <https://www.bostondynamics.com/spot>. [Accessed: 14-Aug-2020].
- [15] E. Y. K. Ng, G. J. L. Kawb, and W. M. Chang, “Analysis of IR thermal imager for mass blind fever screening,” *Microvasc. Res.*, vol. 68, no. 2, pp. 104–109, Sep. 2004.

- [16] J. Mercer, J. B. Mercer, E. Francis, and J. Ring, "Fever screening and infrared thermal imaging: Concerns and guidelines."
- [17] B. B. Lahiri, S. Bagavathiappan, T. Jayakumar, and J. Philip, "Medical applications of infrared thermography: A review," *Infrared Physics and Technology*, vol. 55, no. 4. Pergamon, pp. 221–235, 01-Jul-2012.
- [18] "Use of Infrared to detect elevated Body temperatures Minimizing the spread of infections."
- [19] "GitHub - deepinsight/insightface: Face Analysis Project on MXNet." [Online]. Available: <https://github.com/deepinsight/insightface>. [Accessed: 30-Jul-2020].
- [20] X. Wang, E. G. Ferro, G. Zhou, D. Hashimoto, and D. L. Bhatt, "Association Between Universal Masking in a Health Care System and SARS-CoV-2 Positivity Among Health Care Workers.," *JAMA*, Jul. 2020.
- [21] G. De Haan and A. Van Leest, "Improved motion robustness of remote-PPG by using the blood volume pulse signature," *Physiol. Meas.*, vol. 35, no. 9, pp. 1913–1926, 2014.

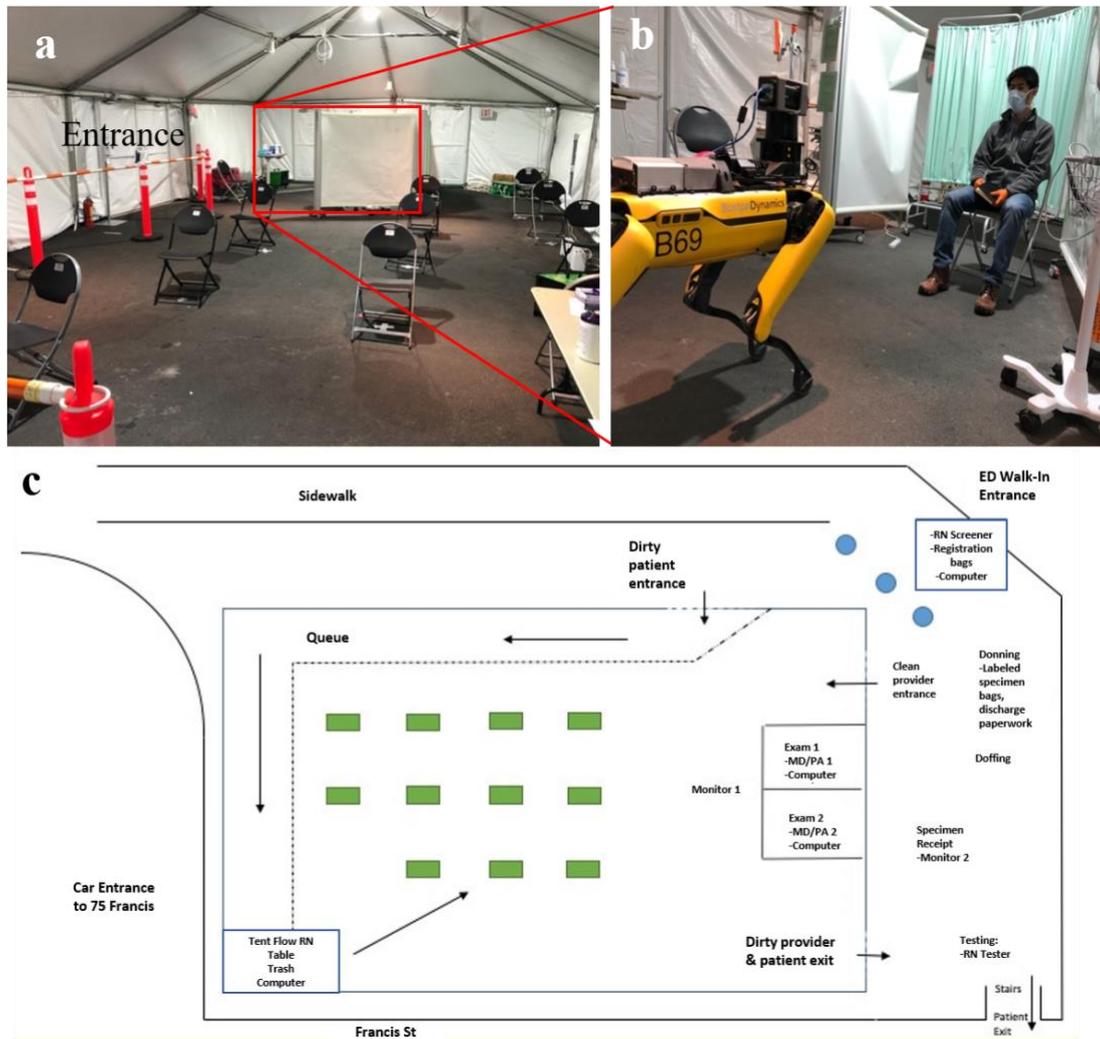


Figure 1. COVID-19 Triage Tent at Brigham and Women's Hospital. (a) medical tent for triage set outside of the emergency department of Brigham and Women's Hospital. (b) Spot with IR camera for fever screening and respiratory rate detection on a healthy volunteer. (c) floor plan of the medical tent.

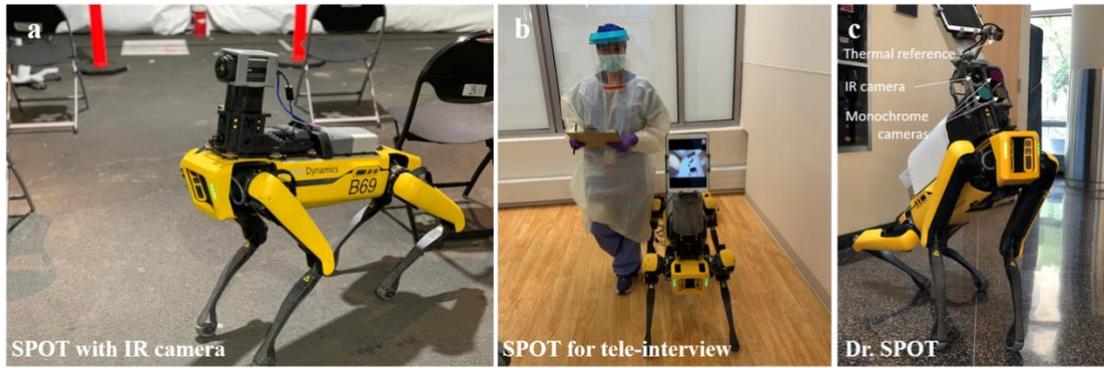


Figure 2. Boston Dynamics Spot robot. (a) Spot carrying an IR camera for fever screening and respiratory rate detection. (b) Spot carrying iPad for tele-interview. (c) Dr. Spot carrying three monochrome cameras with different filter lenses and one IR camera as well as the iPad to enable contactless vital signs monitoring and telemedicine.

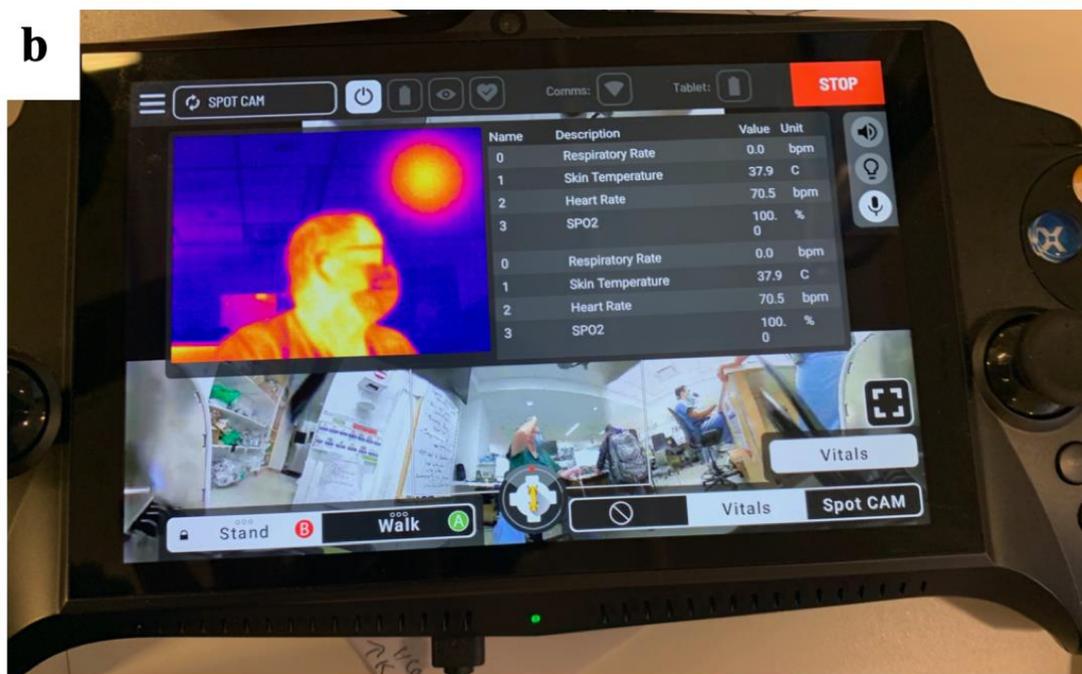
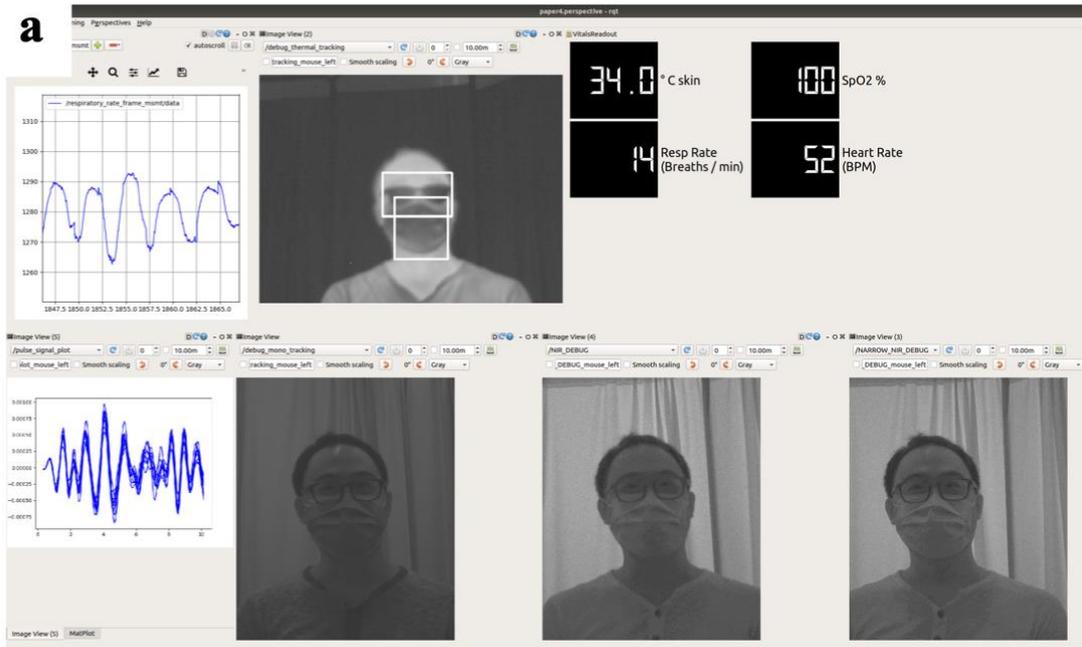


Figure 3. (a) ROS GUI of the Dr. Spot for contactless vital signs monitoring. (b) Handheld controller of the Spot with vital signs measurement results.

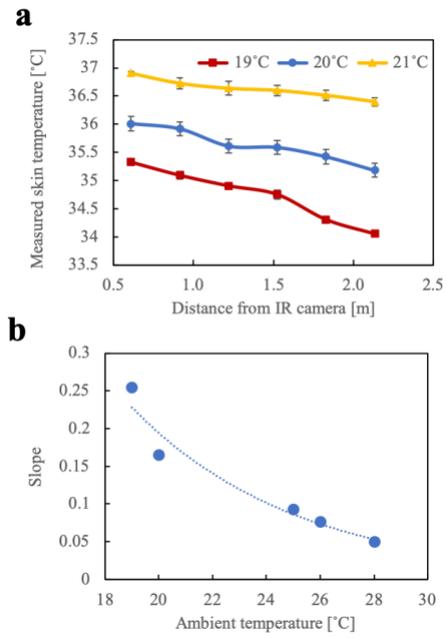


Figure 4. (a) Distance effect on the skin temperature reading using the IR camera. (b) Slope analysis of the distance effects versus ambient temperature. (c) Real-time correction of skin temperature reading.

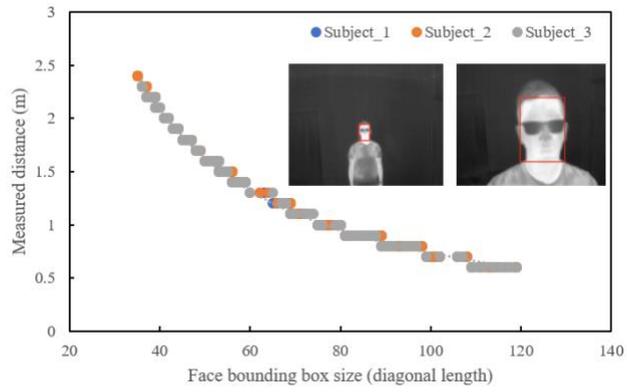


Figure 5. Face detection and tracking on IR images. Distance estimation using the variation in face bounding box dimension. The bounding box gets smaller as the distance between the measuring subject and camera increases.

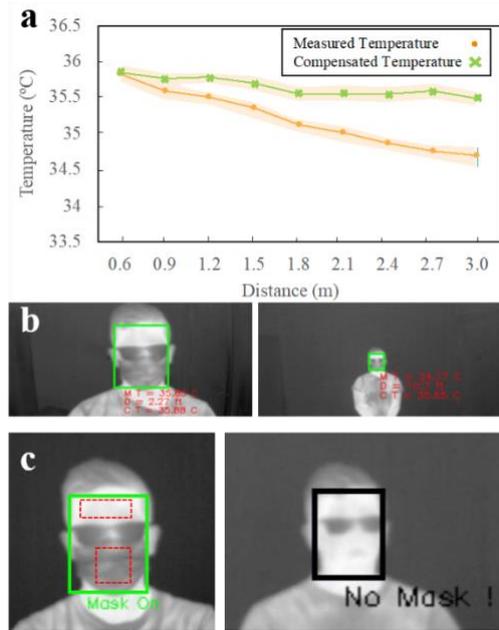


Figure 6. Real-time skin temperature compensation at various distance. (a) Measured and compensated skin temperature reading versus the distance between the camera and measuring subject. (b) IR images showing the difference in face skin temperature at 2.27 ft and 10.7 ft. (c) Mask wearing detection based on thermal images. Two ROIs, forehead and mask are taken into account.

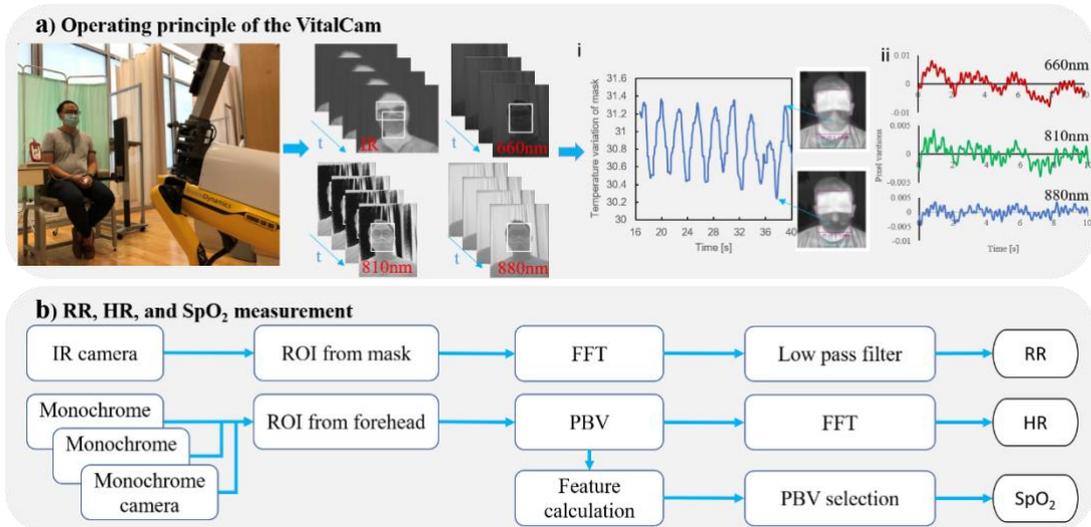


Figure 7. VitalCam Workflow. (a) Operating principle of facial video-based RR, HR, SpO₂ estimation system using IR camera and monochrome cameras with three different wavelength filter. (b) Flowchart for extracting RR, HR, and SpO₂ from facial images.

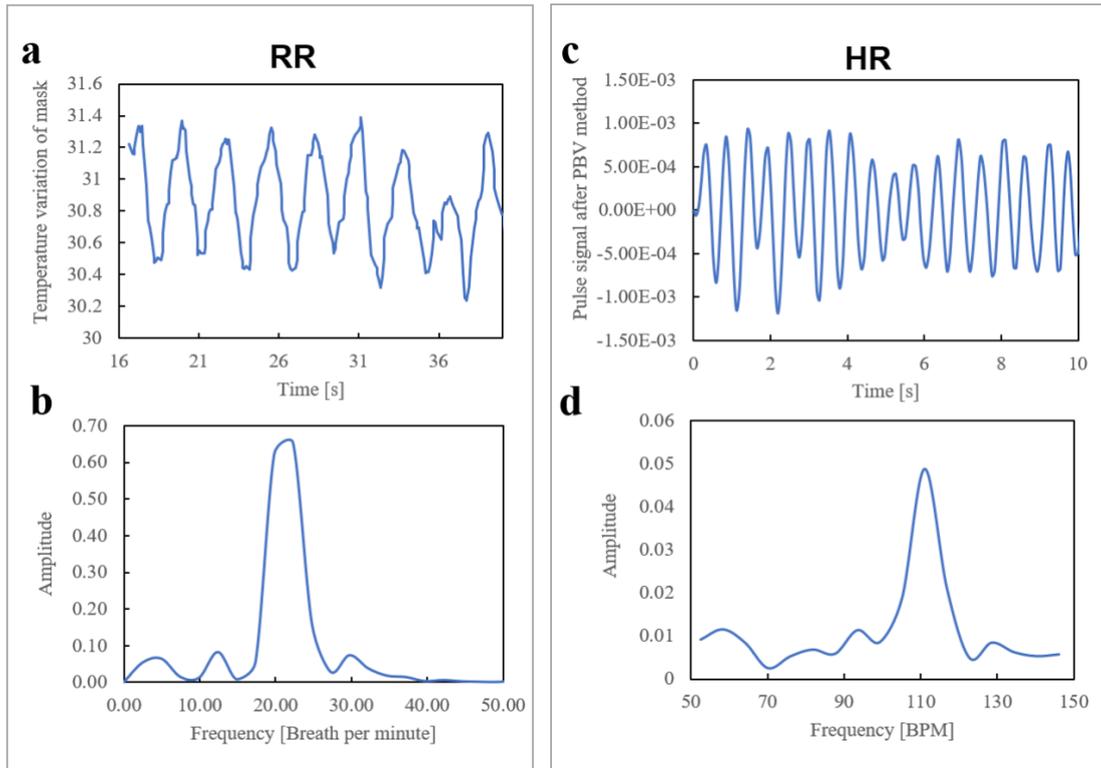


Figure 8. Respiratory and Heart Rate Representative Evaluation. (a) RR waveform after applying a low pass filter. (b) Frequency spectrum of the RR signal by FFT. (c) PPG signal after applying PBV method. (d) Frequency spectrum of the PPG signal by FFT.

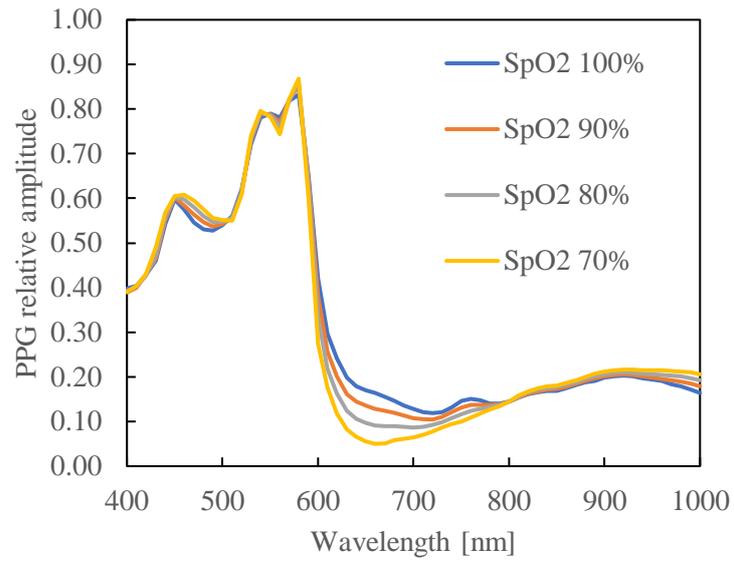


Figure 9. The modeled relative PPG spectrum using the analytic approach by Svaasand et al.[11] for four SpO₂ values.