Cost-effective Mobile Solution for Autonomous and Continuous Vital Signs Monitoring

Hen-Wei Huang[1,2,3*], Jack Chen[1,3*], Philipp Rupp[1*], Claas Ehmke[2,4*], Peter R Chai[2,5], Riya Dhar[2], Ian Ballinger[1,3] Giovanni Traverso[1,2,3]

***Abstract*—The Covid-19 Pandemic has renewed interest in contactless vital signs monitoring using state-of-the-art computer vision, which can efficiently screen for symptoms while reducing the risk of disease transmission. Despite the promising performance, the use of static camera setups requires subjects to remain static inside a field of view (FoV) for a pre-specified duration. Due to inconsistent ambient environmental conditions, the transit of individuals through the FoV, and the time it may take to triage individuals, the widespread adoption of static camera systems to continuously monitor vital signs has had suboptimal uptake. Robotic systems enable autonomous and continuous monitoring, but these require expensive cameras, computers, and robotic platforms, limiting widespread deployment. In response, we propose a cost-effective and scalable robotic solution consisting of a suite of commercial, off-the-shelf wireless cameras for capturing photoplethysmography (PPG) on ambulatory subjects linked to a single computer that supervises the cameras to compute the vital signs of subjects. Throughout a set of careful investigations of each individual step of the wireless machine vision camera and computer, bottlenecks constraining wireless live-streaming of high-quality PPG information are identified and those are addressed by a hybrid centralized/decentralized wireless machine vision protocol. Our results demonstrate that the proposed cost-effective wireless camera achieves equivalent remote-PPG accuracy to its costly, USB3 counterparts (mean error: 5.0 BMP vs. 4.7 BPM) by means of the hybrid camera protocol which boosts the overall frame rate to 17 FPS. In contrast, using the standard method that captures the PPG with the same spatial resolution can only achieve 1 FPS. In addition, this work also elucidates how varying the distance, image pixel density, frame rate, image compression, image downsampling, and color depth affect the rPPG performance. For each of the effects, we also discuss potential solutions for the cost-effective setup.**

I. Introduction

Recent advances in computer vision and artificial intelligence have enabled contactless measurement of respiratory rate, heart rate, heart rate variability, oxygen saturation, and blood pressure via remote photoplethysmography (rPPG) [1]. These methods have shown great potential during the COVID-19 pandemic by enabling remote assessment of individuals for possible COVID-19 thereby reducing potential disease transmission and conserving personal protective equipment [2]. Despite the promise of contactless vital sign acquisition, a major drawback of camera-based setups is the restrictions on human subjects: they must be found and remain in the FoV during measurement. While this is not an insurrmountable barrier in certain situations, the majority of settings where these systems may be deployed may require screening of multiple individuals at once, or rapid assessment of individuals who may be physically unable to remain in a static FoV during the entire assessment.

To remove this restriction, we previously proposed a mobile robotic system consisting of specialized cameras mounted on a commercial quadruped robot (Boston Dynamics, Waltham MA). This system was used to facilitate vital signs monitoring for emergency department triage [3]. Throughout the deployment of the robot in the hospital during the COVID-19 pandemic, we learned several lessons that limited its applicability, such as the need for sophisticated path-planning/obstacle avoidance and the limited battery capacity, the latter of which led to frequent charging and interrupted vital signs monitoring. Furthermore, this robotic setup consisting of an all-terrain robot, high-spec cameras, and powerful computers may be

Fig. 1. Robotic blimps for vital signs monitoring. Multiple blimps connect to one external PC. The blimp consists of a 32" mylar helium balloon and a gondola.

[1]Division of Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston, MA 02115 USA
[2]The Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02142 USA
[3]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA
[4]Multi-Scale Robotics Lab, Department of Mechanical and Process Engineering, ETH Zurich, Tannenstrasse 3, CH-8092, Zurich, Switzerland
[5]Department of Emergency Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston, MA 02115 USA
*Equal Contribution

cost-prohibitive for widescale deployment in large healthcare systems, which in tandem with its bulkiness limited its scalability for continuous monitoring of multiple patients in hospitals.

Here, we demonstrate a cost-effective, aerial-robotic solution for autonomous and continuous vital signs monitoring of multiple subjects simultaneously, aiming to address the aforementioned issues. The cost-effective solution is implemented by a swarm of aerial robotic cameras wirelessly communicating with one central computer. As shown in Figure 1, the wireless robotic camera consists of a 36" helium foil balloon with a mounted gondola. An OpenMV H7 Plus (OpenMV) machine vision microcontroller serves as the robot's computation unit for capturing rPPG.

One major drawback of the cost-effective machine vision camera, such as OpenMV, is the limited computation capabilities as compared to typical computer vision systems. Thus, ROI detection and tracking as well as real-time rPPG estimation must be offloaded to an external computer, placing an unaffordable burden on wireless data transmission for real-time streaming high-quality video. To address this, we develop a hybrid centralized/decentralized camera control protocol by offloading partial computation of rPPG and region of interest tracking from the computer to the camera, thus relieving the burden on wireless data transmission.

The main contributions of this paper are two-fold. First, we describe the development of a cost-effective, easily scalable robotic platform. Second, we propose a novel machine vision protocol to enable state-of-the-art rPPG performance via wireless cameras, enabling simultaneous multi-subject monitoring.

II. Related Works

A. Remote Photoplethysmography

Photoplethysmograhy (PPG) is an optical technique that measures changes in light absorption in the capillaries beneath the skin, which corresponds to changes in blood volume. This technique is most commonly used to measure heart rate and blood oxygen saturation, but it can also be used to measure respiration and blood pressure [4].

Remote photoplethysmography (rPPG) refers to contactless methods for PPG. The most common methods use computer vision, but methods have been developed that use radar and wireless technologies [5]. In vision-based methods, a camera records a subject and detects or segments a region of interest covering skin pixels. An rPPG algorithm then calculates the pulse signal from the variations in the region of interest. The plane-orthogonal-to-skin (POS) method is one such rPPG algorithm that is motion robust and has previously been demonstrated to correlate with ground truth heart rate estimation [1].

Despite the pragmatic benefits of leveraging CV-based methods for rPPG measurement, they importantly restrict movement of the subject. Individuals must enter and remain in the camera's field of view for the entire duration of measurement. In order to improve potential implementation of rPPG and remove restrictions from subject movement, previous work has focused on deploying rPPG on robotic platforms that may track and follow individuals. These works use a differential-wheel robot mounted with devices such as a 3D camera, laptop, and webcam [6], [7]. We previously used the quadruped Spot robot by Boston dynamics mounted with RGB and IR cameras [3].

B. Region of Interest

The region of interest (ROI) in an image is the part of an image with interesting information. For instance, the region of interest for object detection is the part of an image containing the object. A common trade-off for cameras is between image resolution and frame rate; increasing one necessarily decreases the other. Our previous work has developed a closed-loop ROI algorithm that allows a high-resolution camera to only capture and transmit what is interesting while ignoring redundant information to boost the overall frame rate; furthermore, this works enables a wide FoV to be maintained for tracking fast-moving objects [8], [9]

For rPPG, common ROIs include the forehead, cheeks, and face, which contain large regions of unobstructed skin. Previous work has investigated ROI detection and skin segmentation methods for rPPG [10], [11]. Others have assessed different ROIs and their suitability for rPPG algorithms [12], [13]. We selected the forehead as our ROI because with increasing mask use globally due to the COVID-19 pandemic, other ROI regions conductive to measuring rPPG are frequently obscured. Fortunately, the forehead has been confirmed as an excellent ROI for rPPG, with a high signal-to-noise ratio [14], [15].

C. Miniature Robotic Blimps

Miniature robotic blimps are a well-researched robotic platform consisting of a helium balloon with an attached gondola. Small motors and propellers control the blimp's motion, while micro-controllers and sensors enable robotic functionalities.

Miniature robotic blimps offer a range of benefits over other aerial robots, including lower cost, lower noise, and increased collision tolerance [16], [17]. Moreover, the robotic blimp does not consume energy to hover, thus resulting in extended battery life. However, blimps are limited by their lower speed, lower durability, and smaller payload [16]. The latter is a particular challenge, as robotic blimps can only carry a very limited set of sensors, batteries, motors, etc. As a result, some work have focused on the optimization of parameters such as number of actuators, [18], gondola placement [19], and communication interface [20].

Previous works have studied and simulated the dynamics of miniature robotic blimps to inform the design of control algorithms [17]. Controlling these blimps is made difficult due to perturbations caused by airflow and buoyancy changes due to leakage and variations in ambient conditions. Some works have designed adaptive motion control algorithms [21], while others have used reinforcement learning to achieve robust control policies capable of withstanding wind disturbance [22], [23].

Functionalities such as localization, navigation, path planning, person following, and teleoperation have all been explored on the robotic blimp platform [24], [25]. Commonly

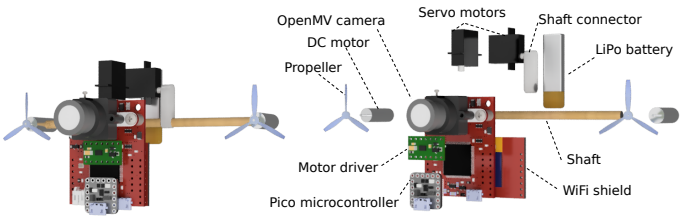


Fig. 2. CAD model of robotic blimp gondola showing the assembled and exploded views.

proposed applications for this platform include search and rescue, surveillance, and video broadcasting [26]–[28].

III. Methods

For vital signs monitoring in a hospital filled with moving persons and machinery, an overhead aerial robot may be more suitable as it can more easily navigate chaotic environments compared to grounded robots. Although drones are the more common aerial robot, they are loud and pose safety hazards especially when operated indoors. In contrast, miniature robotic blimps make very little noise and are collision-tolerant, and have zero energy consumption-hovering. The main benefits of the blimp are its safety, long battery life, and low noise, but it has low speed and limited payload mass. We use a swarm of low-cost robotic cameras that on each blimp coordinates with one central PC; this enables the cost-effective and easily-scalable solution capable of simultaneous multi-subject monitoring.

A. Autonomous robotic blimps

The robotic blimp comprises a gondola (shown in Figure 2) attached to a mylar foil helium balloon that is 36” in diameter. The gondola is composed of the OpenMV H7 Plus (OpenMV) machine vision microcontroller, the ATWINC1500 WiFi module, a MellBell Pico microcontroller (0.6” x 0.6” package), a Polulu DRV8835 dual DC motor driver, one rechargeable LiPo battery (700 mAh), two mini DC motors with propellers, and two micro servo motors. The two DC motors controlled by the DRV8835 can adjust the speed and direction of the propellers. The two micro servo motors are used to control pitch angle of the propellers and the yaw angle of the OpenMV camera, respectively. The OpenMV is the robot’s onboard controller which is in charge of the decentralized computation for capturing high-quality rPPG information; the Pico microcontroller is in charge of blimp locomotion control. The OpenMV and Pico communicate via I2C. The OpenMV communicates via WiFi (ATWINC1500 WiFi module [29]) with an external NVIDIA Jetson AGX Xavier embedded PC (PC). The PC is the centralized computing device that provides high-level control of the robot. To acquire high-quality rPPG while keeping social distancing, the distance between the robot and a measuring subject is set at 2 meters, and the pixels covering the forehead are set to 500 x 200.

Figure 3 shows the block diagram of robotic camera operations. Each blimp begins by rotating in place, transmitting captured images over WiFi to the PC. The PC then detects

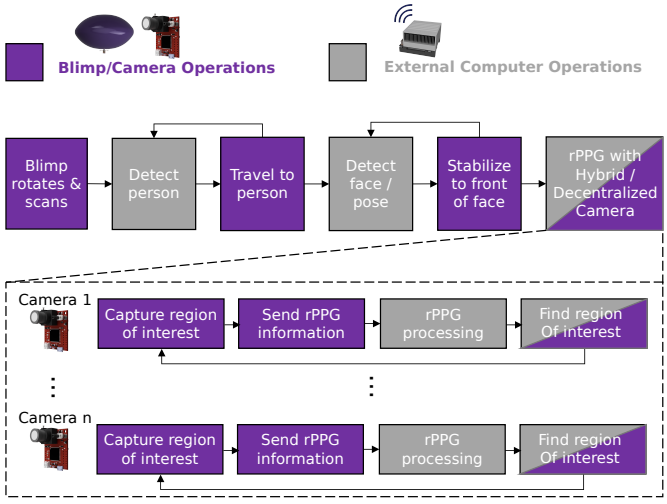


Fig. 3. Block diagram of robotic blimp platform for rPPG. The robot communicates with an external PC to perform operations.

bodies using the YoloV4 object detection model [30]. Each blimp travels to a person using the person-detection bounding box as the PID feedback. The PC then detects faces using the MTCNN object detection model [31] and Dlib face pose estimation model. Each blimp stabilizes to the person’s forehead using the face-detection bounding box and face pose as the PID feedback. Finally, each blimp uses the hybrid camera method–present in the latter subsection–to stream rPPG video for vital signs estimation.

B. rPPG recording and analyses

To test the proposed camera methods, we applied them to eleven subjects 2 meters away from the camera. FLIR Blackfly USB3 RGB camera was employed as a reference for benchmarking the performance of the cost-effective wireless camera. To test parameters that were not camera-specific, we use the UBFC datset, a public rPPG dataset with 43 subject [32].

The POS algorithm is used to analyze the rPPG information to estimate vital signs [1]. Typical rPPG algorithms measures characteristic changes in skin color caused by constriction and dilation and capillaries. The POS algorithm constructs two orthogonal signals from variations in the skin’s averaged RGB signal, from which the pulse signal is extracted.

To test the scalability of our solution, we connected five static OpenMV cameras to one external PC wirelessly using the transmission control protocol. All five OpenMV cameras run our proposed rPPG method while simultaneously coordinating with the PC. We measure the frame rate of each camera do determine the effects of multithreading.

C. Wireless machine vision protocols

Figure 4 presents the three protocols that control the wireless machine vision camera for capturing rPPG. The Standard Camera is the typical method for rPPG, which we use as a benchmark for our proposed methods. The Standard Camera captures and transmits the full-resolution image to an external

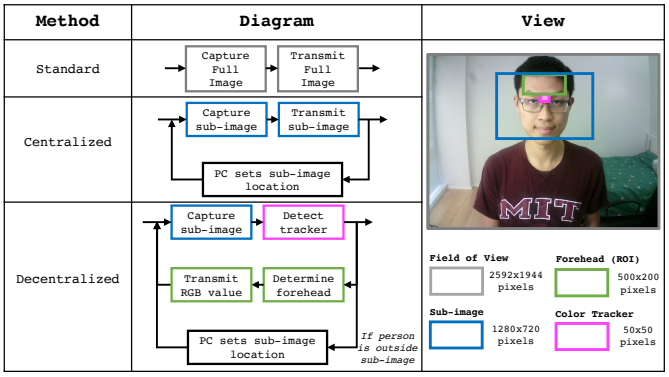


Fig. 4. Overview of the standard and our proposed camera methods for rPPG.

PC. The PC detects the forehead and computes the RGB average of the pixels covering the forehead. The averaged RGB value will then be used to analyze the rPPG information via the POS algorithm.

The Centralized Camera is an improved method, in which the PC directly supervises the camera to only capture and transmit the ROI. The PC first takes the full-resolution image to find the location of the subject's face, it then guides the camera to capture the sub-image containing the face. The sub-image, which is the blue bounding box (1280 x 720) in the Figure 4, allows reasonable head movement. We also identify that 1280 x 720 is the minimum ROI size with the highest pixel density for capturing rPPG without losing color fidelity. Once the face moves beyond the blue bounding box, the PC will supervise the camera to reconfigure the sub-image. In this way, the camera captures and transmits a much smaller image.

The Hybrid Centralized/Decentralized Camera is a further improved method, in which the PC offloads computation of localizing forehead as well as rPPG pre-processing to the OpenMV. Algorithm 1 shows the pseudocode for the hybrid camera. In this method, a small color tracker is placed on the subject's forehead to enable the OpenMV camera to track the forehead. Though this is a rudimentary form of forehead detection, microcontrollers such as the OpenMV do not have the capacity to run more complex models at a reasonable frame rate. For instance, we tested a TensorFlow Lite model for the OpenMV using the COCO Common Objects dataset [33]; unfortunately, this model ran at below 1 FPS on the OpenMV. Color tracking is much simpler and enables real-time, frame-by-frame detection on the OpenMV.

For the Hybrid Camera, the same blue bounding box (denoting the region of interest) as the centralized method will be captured by the camera. In every frame, the OpenMV detects the color tracker and uses its location to estimate the forehead's location. Detection is performed by filtering for color blobs in the narrow range of the color tracker. Different from the Standard and Centralized Cameras that transmit the entire RGB pixels covering the ROI to the PC, the OpenMV first calculates the average RGB value of the forehead and only transmits the averaged value to the PC. The PC then directly analyzes the rPPG information.

Algorithm 1 Operations of hybrid camera for rPPG

```
camera.capture_window = (0, 0, 2592, 1944)
img = camera.capture_image()
tracker = camera.color_detection(img)
ROI_{x_loc} = (tracker.x - 1280) / 2
ROI_{y_loc} = (tracker.y - 720 ) / 2
camera.capture_window = (ROI_{x_loc}, ROI_{y_loc}, 1280, 720)
timer.start()

while True do
    img = camera.capture_image()
    tracker = camera.color_detection(img)
    forehead = (tracker.x - 250, tracker.y - 100)
    average = camera.compute_RGB_avg(forehead)
    camera.Wifi_transmit_to_PC(average)

    if timer.time() == 60 seconds then
        camera.capture_window = (0, 0, 2592, 1944)
        img = camera.capture_image()
        tracker = camera.color_detection(img)
        ROI_{x_loc} = (tracker.x - 1280) / 2
        ROI_{y_loc} = (tracker.y - 720 ) / 2
        camera.capture_window = (ROI_x, ROI_y, 1280, 720)
        timer.restart()

    end if
end while
```

The Hybrid Camera takes advantage of two factors: color detection is a simple algorithm that can run in real time even on embedded processors; rPPG does not actually require images of the skin, and only requires the average RGB values of the skin. Thus, these basic decentralized operations on the OpenMV are sufficient to enable rPPG.

IV. Results and Discussion

Leveraging rPPG could reduce the risk of spreading airborne pathogens in hospitals and other enclosed settings as it allows healthcare workers to acquire patients' vital signs remotely or with improved barrier protection. However, simply increasing the distance between the camera and a measuring subject will lead to decreased number of pixels covering the ROI, which results in increased rPPG estimation error as shown in Figure 5. One simple solution is to change the lens focal length, which then sacrifices the FoV for tracking head or body movement. The other option is to increase the camera's image resolution which then compromises the frame rate in capturing images, which again results in increased rPPG estimation error as shown in Figure 6. To obtain high-quality rPPG, it is important to have not only enough pixels covering the ROI but also a high frame rate. However, the limited wireless transmission speed impedes the OpenMV camera to capture ROI with its highest spatial resolution, by which the overall frame rate would be only 1 FPS.

To address the bottleneck in wireless data transmission of high-quality rPPG, two machine vision camera protocols, the centralized and hybrid protocols, are developed by means of

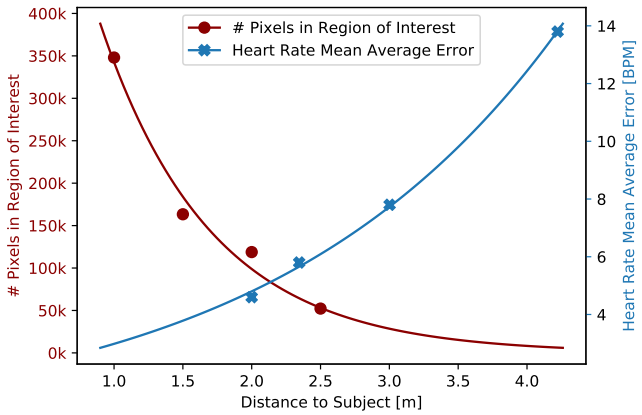


Fig. 5. Frame rate and number of region of interest pixels versus heart rate estimation of rPPG methods.

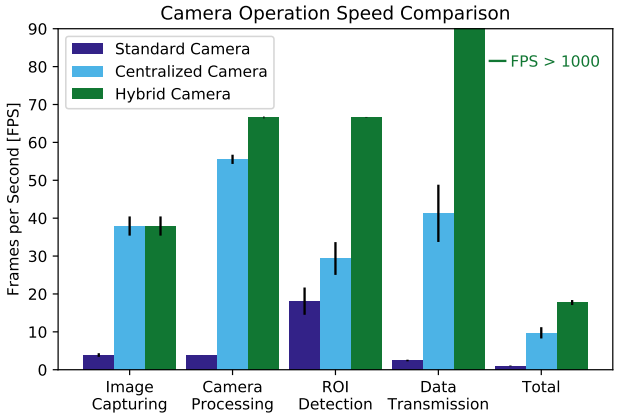


Fig. 7. Operation time and frame rate comparison between camera methods for remote rPPG. Data is collected using the OpenMV machine vision microcontroller at maximum resolution (2592x1944)

the closed-loop ROI algorithm. The centralized protocol allows the camera to only capture and transmit ROI with the highest spatial and temporal resolution. Although the centralized protocol can significantly reduce the time for capturing ROI with its highest spatial resolution, the temporal resolution is still sub-optimal. In the centralized camera protocol, the ROI should contain not only the forehead for rPPG analyses but also the entire face for ambulatory subject detection and tracking. To further boost the temporal resolution, we propose the hybrid protocol that is a combination of centralized and decentralized control of the wireless machine vision camera to allow the camera only transmit the pre-processed rPPG information that is averaged forehead RGB information with the data size of only a single pixel. As the PC does not receive images covering an entire face, the ROI tracking needs to be offloaded to the camera, thus relieving the burden on the wireless data transmission. In the hybrid protocol, the camera is partially supervised by the PC only for localizing faces. When the camera loses forehead detection, the PC takes over full control of ROI detection and rPPG processing by quickly switching to the centralized protocol. Once the face is localized again, the camera will swtich back to the hybrid protocol.

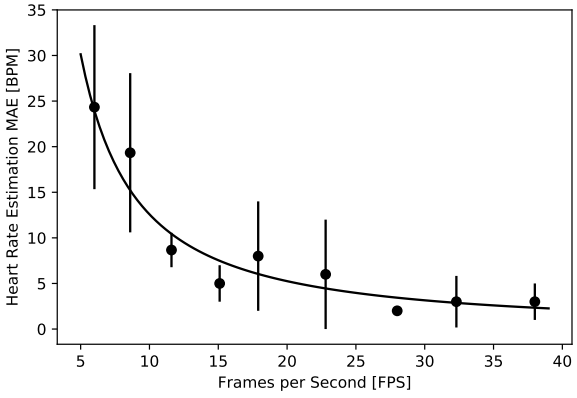


Fig. 6. Frame rate versus heart rate estimation of rPPG methods.

In the following sections, we systematically characterize the performance of the two protocols together with the standard approach in terms of their frame rate under the circumstance that the pixels covering the forehead are identical which is 500 x 200 pixels with the highest pixel density.

A. *Spatial and temporal resolution characterization*

Figure 7 shows the overall comparison of the camera methods using the OpenMV machine vision microcontroller operating at maximum resolution (2592x1944). The Standard Camera operates at 1.0±0.1 FPS, the Centralized Camera operates at 9.7±1.5 FPS, and the Hybrid Camera operates at 17.1±0.7 FPS. Although the Centralized Camera significantly improves every step of the camera operation, it can be seen that ROI detection and data transmission are the two major factors limiting its overall frame rate. As expected, the Hybrid Camerea shows that offloading the ROI detection to the OpenMV and transmitting the information only containing rPPG can greatly boost the speed in all steps.

Here, we will elucidate how each step of the camera and PC operation contributes to the operation speed in the three protocols with varying the processed image resolution which is the size of the blue bounding box shown in Figure 4. The first step for all of the camera protocols is image capturing. To ensure the forehead is covered by 500 x 200 pixels, the Standard Camera must capture the full field of view at maximum resolution (2592x1944), while the Centralized and Hybrid Cameras only need to capture a smaller region of interest containing the face (1280x720). As shown in Figure 7, this leads to a significant increase in the image capturing speed from 4 FPS to 38 FPS.

The second and third steps of the camera methods are image processing and ROI detection. Raw RGB images are too large to be wirelessly transmitted in real-time. JPEG compression significantly reduces image size, but may affect rPPG. This is shown in Figure 8. Indeed, greater image compression increases the MAE, but this effect is limited above a compression factor of 70%. Thus, when using JPEG compression

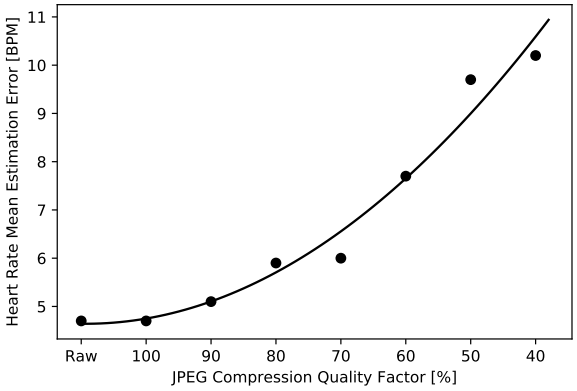


Fig. 8. Effect of JPEG image compression on rPPG accuracy. Raw refers to uncompressed RGB images. Image compression and analysis was performed using the UBFC dataset.

in our camera methods, we use a compression quality factor of 70% for the Standard and Centralized Cameras. It is worth noting that the Hybrid Camera does not require the image compression.

For the Standard and Centralized protocols, data processing occurs first under which the image is processed and then wirelessly transmitted to the PC for ROI detection. For the Hybrid protocol, ROI detection occurs first under which the color marker is detected, and then the image is processed directly on the machine vision microcontroller. Figure 9 characterizes the data processing step. For the Standard and Centralized cameras, we compress the raw image in JPEG format.

The time for JPEG compression largely depends on the image size. Reducing the image resolution could greatly increase the frame rate. Unlike the Standard and Centralized Cameras, the Hybrid Camera does not require JPEG compression. Instead, it needs to compute the raw RGB average of the detected forehead with an image resolution of 500 x 200, which explains why the Centralized Camera is faster than the Hybrid Camera under the same processed image resolution (Figure 9). As a quick note, the Hybrid Camera takes 15 ms in the data processing of images with resolution of 500 x 200. While, the Centralized Camera requires 18 ms to compress images with a resolution of 1280 x 720. As the image data that the Hybrid Camera needs to process is smaller than that centralized camera does which are marked by blue and green dash lines in Figure 9, it explains the discrepancy in the camera processing speed between Figure 7 and Figure 9.

ROI detection is another step that is heavily affected by the image resolution as shown in Figure 10. For the Standard and Centralized Cameras, the PC detects the forehead ROI from the received image using the MTCNN object detection neural network. The Centralized Camera detects the ROI from a much smaller image that only contains the face; since MTCNN is faster for smaller images, the Centralized Camera achieves better performance than the standard Camera [31]. The Hybrid Camera is even faster than the Centralized one because it runs a much simpler color tracker that filters for the specified color range. This strategy allows us to increase the frame rate resolution for ROI while focusing on only the key data necessary to calculate rPPG. Under this circumstance, the Hybrid Camera achieves 38 FPS for ROI detection, which is faster than the 29 FPS for the Centralized Camera and 18 FPS for the Standard Camera.

It is worth noting that the characterization of the ROI detection step may be unfair. Indeed, the Hybrid Camera is significantly better, but it only performs color detection while the other camera methods perform face detection. However, if the Standard Camera method used color detection rather than face detection, its frame rate remains at 1.0 FPS. Similarly, if the Centralized Camera used color detection, its frame rate would be 12.0 FPS instead of 9.7 FPS. Since ROI detection is not the bottleneck, these are insignificant improvements that do not lead to higher rPPG accuracy.

The last step that would make difference in frame rate is the wireless data transmission. This is the bottleneck of

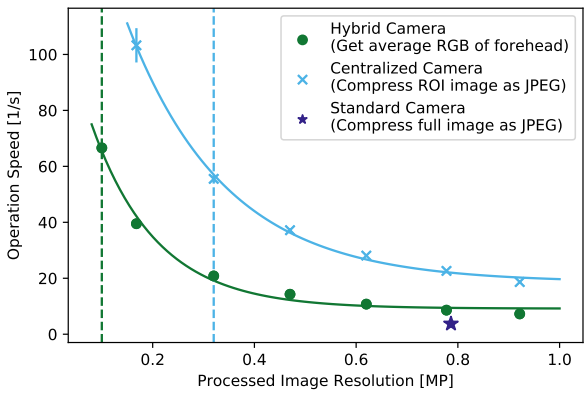


Fig. 9. Characterization of camera processing using the OpenMV machine vision microcontroller. The green and blue dashed lines indicating the exact processed image resolution in the Hybrid and Centrazlied Cameras for image formatting, respectively.

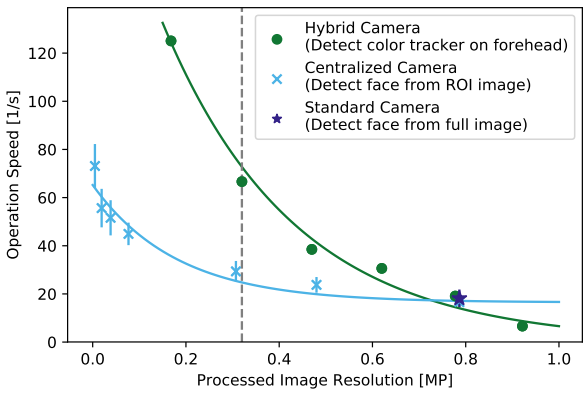


Fig. 10. Characterization of the region of interest detection using the OpenMV machine vision microcontroller (Hybrid Camera) and the PC (Standard & Centralized Cameras). The gray dashed line indicateding the exact processed image resolution in both the Hybrid and Centralized Cameras for ROI detection.

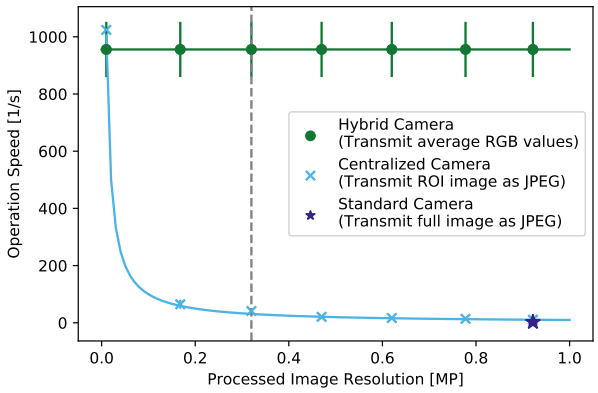


Fig. 11. Characterization of data transmission using the OpenMV machine vision microcontroller. The gray dashed line indicating the processed image resolution of both the Hybrid Camera and Centralized Camera for wireless data transmission.

the Standard Camera as it needs to transmit full resolution images. Figure 11 characterizes the data transmission step. For the Standard Camera, a full-size JPEG image (0.1 MB) must be transmitted. Since the Centralized Camera only captures a small sub-image (1280 x 720, 0.55 MB), the frame rate in data transmission would be significantly faster (41 FPS) than the Standard one (2 FPS). For the Hybrid Camera, only the RGB average value (1e-5 MB) needs to be transmitted, which can be done instantaneously by the microcontroller's WiFi module. Thus, the Hybrid Camera achieves 956 FPS for data transmission.

Since we are able to partially offload some computation from the PC to the machine vision camera using the Hybrid protocol, a swarm of robotic blimps can be employed to simultaneously monitor multiple subjects using only one central PC, thus significantly reducing the overall cost for scaling up. To test the scalability, we established TCP connections between five OpenMV machine vision cameras on the robotic blimps to one central PC via multithreading. During testing, running five OpenMVs in parallel did not cause drops in frame rate for any camera operations.

B. rPPG Estimation Accuracy of Proposed Protocols: Constraints and Evaluation

Besides the limited computation performance of the OpenMV camera, the lower CMOS color depth is another reason making it a cost-effective camera. The OpenMV camera's CMOS provides a color depth of 16-bit with RGB565 color encoding. To test the effects of color depth, we downsized 24-bit RGB888 video recordings to lower color depth and then analyzed the rPPG accuracy. As shown in Figure 12, the heart rate estimate error increases with decreased color depth. Specifically, downsizing from RGB888 to RGB565 only increases error by 10 %. While the error of RGB454 is more than 20 % than the RGB888. We then compare the accuracy of the proposed centralized and hybrid protocols for heart rate estimation with rPPG. These results support that RGB565 is still capable of capturing rPPG without losing color fidelity. Furthermore, storing RGB data as floats rather than integers greatly improves rPPG accruacy, as using integers leads to rounding errors.

To further understand constraints on rPPG, we investigated image downsampling effects on heart rate estimation accuracy. As we previously characterized, using high-resolution images places many bottlenecks on machine vision microcontrollers, from image processing to data transmission. Two common methods for image downsampling are binning (in which the values of several pixels and averaged into one larger pixel) and decimation (in which the values of periodic pixels are discarded). Figure 13 shows that image binning greatly decreases rPPG accuracy, while decimation has negligible effects. This is because the OpenMV, like other conventional machine vision microcontrollers, stores pixel values as integers; averaging pixel values then rounding to one integer causes loss in precision in the RGB value. This is directly related to the constraints shown in Figure 12, in which RGB bit depth is critical for capturing pulsatile flow. Unfortunately, the OpenMV is only capable of binning and not decimation; thus, we are currently

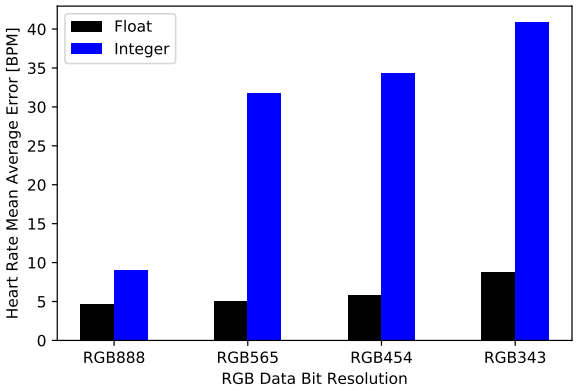


Fig. 12. Analysis of color depth bit resolution on heart rate estimation accuracy. Analysis is performed usin the UBFC dataset. The original data is downsized to different RGB resolutions and data types.

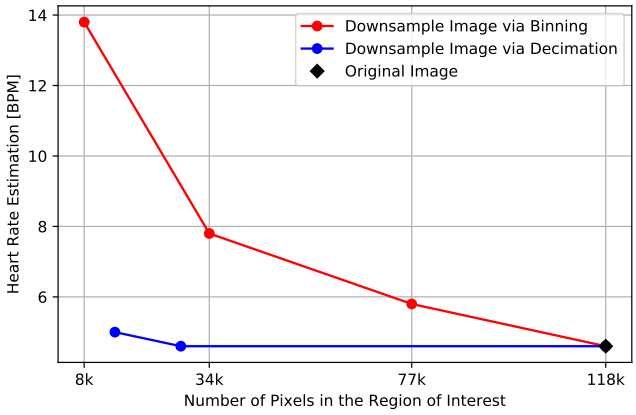


Fig. 13. Effect of downsampling image resolution on rPPG heart rate estimation accuracy.

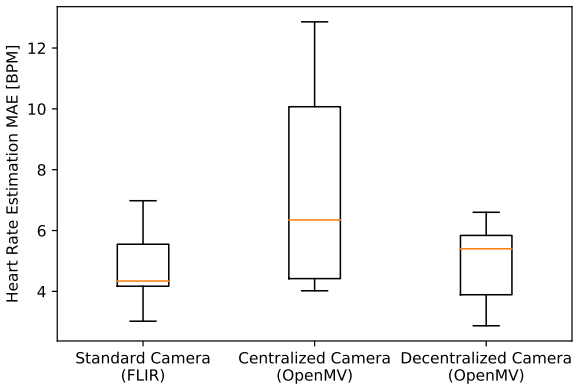


Fig. 14. Accuracy of proposed rPPG methods against standard benchmark for nine subjects. The FLIR Blackfly state-of-the-art camera is used as a benchmark. The proposed methods are deploymed on the OpenMV machine vision microcontroller.

unable to take advantage of possible savings from image downsampling via decimation.

To validate the accuracy of the proposed centralized and hybrid protocols, we employ the USB3 high resolution FLIR camera that is capable of capturing rPPG with the same spatial resolution and a significantly high frame rate (40 FPS) as the reference. Figure 14 shows the rPPG accuracy results for nine subjects using the different camera methods. The expensive Standard USB3 Camera (FLIR Blackfly with RGB888) achieves accurate performance with a mean average error of 4.7 BPM with which the overall frame rate of the FLIR Blackfly camera is about 40 FPS. The cost-effective Camera (OpenMV H7 Plus) with the Centralized protocol is unable to achieve accurate heart rate estimation due to its low frame rate. The Hybrid Camera, which addresses the bottlenecks in ROI detection and wireless data transmission, achieves comparable performance with a mean average error of 5.0 BPM due to its higher frame rate (17 FPS) while operating at maximum resolution (2592x1944). It is shown that the noticeable difference in the frame rate between the FLIR camera and the OpenMV hybrid protocol does not make big difference in the rPPG estimation error which corresponds to Figure 6.

V. Conclusion

In this paper, we propose a cost-effective robotic solution with a wireless machine vision microcontroller capable of autonomous, continuous vital sign monitoring. We develop the Hybrid Camera protocol to equip the cost-effective system with comparable performance to its expensive counterparts. Future work may focus on the deployment of the robot for clinical applications and on the development of truly motion-robust rPPG algorithms.

Acknowledgments

We thank the funding support from Karl Van Tassel (1925) Career Development Professorship, the Department of Mechanical Engineering, MIT, and the Division of Gastroenterology, Brigham and Women's Hospital. We thank the contributions from Abhijay Kemkar and Naitik Khandelwal in the development of the robotic blimp and rPPG camera methods.

References

[1] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote ppg," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, 2017. DOI: 10.1109/TBME.2016.2609282.

[2] F. Yang, S. He, S. Sadanand, A. Yusuf, and M. Bolic, "Contactless measurement of vital signs using thermal and rgb cameras: A study of covid 19-related health monitoring," *Sensors*, vol. 22, no. 2, 2022, ISSN: 1424-8220. [Online]. Available: https://www.mdpi.com/1424-8220/22/2/627.

[3] H.-W. Huang, J. Chen, P. R. Chai, *et al.*, "Mobile robotic platform for contactless vital sign monitoring," *Cyborg and Bionic Systems*, vol. 2022, 2022. DOI: 10.34133/2022/9780497.

[4] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, R1–39, Apr. 2007. DOI: 10.1088/0967-3334/28/3/R01.

[5] Y. Rong, P. C. Theofanopoulos, G. C. Trichopoulos, and D. W. Bliss, "A new principle of pulse detection based on terahertz wave plethysmography," *Scientific Reports*, vol. 12, no. 1, 2022. DOI: 10.1038/s41598-022-09801-w.

[6] H. Lee, H. Ko, H. Chung, and J. Lee, "Robot assisted instantaneous heart rate estimator using camera based remote photoplethysmograpy via plane-orthogonal-to-skin and finite state machine," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 4425–4428. DOI: 10.1109/EMBC44109.2020.9176648.

[7] H. Lee, H. Ko, H. Chung, and J. Lee, "Real-time realizable mobile imaging photoplethysmography," *Scientific Reports*, vol. 12, 2022. DOI: 10.1038/s41598-022-11265-x.

[8] J. Chen, H.-W. Huang, P. Rupp, A. Sinha, C. Ehmke, and G. Traverso, "Closed-loop region of interest enabling high spatial and temporal resolutions in object detection and tracking via wireless camera," *IEEE Access*, vol. 9, pp. 87 340–87 350, 2021. DOI: 10.1109/ACCESS.2021.3086499.

[9] H.-W. Huang, P. Rupp, J. Chen, *et al.*, "Cost-effective solution of remote photoplethysmography capable of real-time, multi-subject monitoring with social distancing," in *IEEE Sensors Conference*, 2022.

[10] P. Li, Y. Benezeth, K. Nakamura, R. Gomez, C. Li, and F. Yang, "Comparison of region of interest segmentation methods for video-based heart rate measurements," in *2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE)*, 2018, pp. 143–146. DOI: 10.1109/BIBE.2018.00034.

[11] P. Li, Y. Benezeth, K. Nakamura, R. Gomez, and F. Yang, "Model-based region of interest segmentation for remote photoplethysmography," Jan. 2019, pp. 383–388. DOI: 10.5220/0007389803830388.

[12] D.-Y. Kim, K. Lee, and C.-B. Sohn, "Assessment of roi selection for facial video-based rppg," *Sensors*, vol. 21, no. 23, 2021, ISSN: 1424-8220. [Online]. Available: https://www.mdpi.com/1424-8220/21/23/7923.

[13] S. Kwon, J. Kim, D. Lee, and K. Park, "Roi analysis for remote photoplethysmography on facial video," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 4938–4941. DOI: 10.1109/EMBC.2015.7319499.

[14] S. Kwon, J. Kim, D. Lee, and K. Park, "Roi analysis for remote photoplethysmography on facial video," English, in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2015*, ser. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Institute of Electrical and Electronics Engineers Inc., Nov. 2015, pp. 4938–4941. DOI: 10.1109/EMBC.2015.7319499.

[15] D.-Y. Kim, K. Lee, and C.-B. Sohn, "Assessment of roi selection for facial video-based rppg," *Sensors*, vol. 21, no. 23, 2021, ISSN: 1424-8220. DOI: 10.3390/s21237923. [Online]. Available: https://www.mdpi.com/1424-8220/21/23/7923.

[16] J. Hu, A. Bruno, D. Zagieboylo, *et al.*, "To centralize or not to centralize: A tale of swarm coordination," *CoRR*, vol. abs/1805.01786, 2018. arXiv: 1805.01786. [Online]. Available: http://arxiv.org/abs/1805.01786.

[17] K. Motoyama, H. Kawamura, M. Yamamoto, and A. Ohuchi, "Development of autonomous blimp robot with intelligent control," in *Entertainment Computing: Technologies and Application*, R. Nakatsu and J. Hoshino, Eds. Boston, MA: Springer US, 2003, pp. 191–198. DOI: 10.1007/978-0-387-35660-0_23. [Online]. Available: https://doi.org/10.1007/978-0-387-35660-0_23.

[18] S. Ferdous, A. Mohammadi, and S. Lakshmanan, "Developing a low-cost autonomous blimp with a reduced number of actuators," in *Unmanned Systems Technology XXI*, May 2019, p. 13. DOI: 10.1117/12.2519252.

[19] G. Gorjup and M. Liarokapis, "A low-cost, open-source, robotic airship for education and research," *IEEE Access*, vol. 8, pp. 70713–70721, 2020. DOI: 10.1109/ACCESS.2020.2986772.

[20] R. Al-Jarrah and H. Roth, "Design blimp robot based on embedded system and software architecture with high level communication and fuzzy logic," in *2013 9th International Symposium on Mechatronics and its Applications (ISMA)*, 2013, pp. 1–6. DOI: 10.1109/ISMA.2013.6547395.

[21] T. Takaya, H. Kawamura, Y. Minagawa, M. Yamamoto, and A. Ouchi, "Motion control in three dimensional round system of blimp robot," in *2006 SICE-ICASE International Joint Conference*, 2006, pp. 1291–1294. DOI: 10.1109/SICE.2006.315440.

[22] Y. T. Liu, E. Price, P. Goldschmid, M. J. Black, and A. Ahmad, "Autonomous blimp control using deep reinforcement learning," *CoRR*, vol. abs/2109.10719, 2021. arXiv: 2109.10719. [Online]. Available: https://arxiv.org/abs/2109.10719.

[23] A. Rottmann, C. Plagemann, P. Hilgers, and W. Burgard, "Autonomous blimp control using model-free reinforcement learning in a continuous state and action space," in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007, pp. 1895–1900. DOI: 10.1109/IROS.2007.4399531.

[24] Y. Liu, Z. Pan, D. Stirling, and F. Naghdy, "Q-learning for navigation control of autonomous blimp," Jan. 2009.

[25] N. Yao, E. Anaya, Q. Tao, S. Cho, H. Zheng, and F. Zhang, "Monocular vision-based human following on miniature robotic blimp," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 3244–3249. DOI: 10.1109/ICRA.2017.7989369.

[26] Y.-W. Huang, C.-L. Lu, K.-L. Chen, *et al.*, *Duckiefloat: A collision-tolerant resource-constrained blimp for long-term autonomy in subterranean environments*, 2019. DOI: 10.48550/ARXIV.1910.14275. [Online]. Available: https://arxiv.org/abs/1910.14275.

[27] T. Fukao, T. Oshibuchi, K. Osuka, T. Kohno, and Y. Tomoi, "Outdoor blimp robots for rescue surveillance systems," in *2008 SICE Annual Conference*, 2008, pp. 982–987. DOI: 10.1109/SICE.2008.4654798.

[28] T. Fukao, T. Kanzawa, and K. Osuka, "Tracking control of an aerial blimp robot based on image information," in *2007 IEEE International Conference on Control Applications*, 2007, pp. 874–879. DOI: 10.1109/CCA.2007.4389343.

[29] I. Abdelkader, Y. El-Sonbaty, and M. El-Habrouk, *OpenMV: A Python powered, extensible machine vision camera*, 2017. arXiv: 1711.10464.

[30] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, *Yolov4: Optimal speed and accuracy of object detection*, 2020. arXiv: 2004.10934.

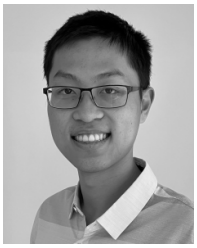
[31] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016. DOI: 10.1109/LSP.2016.2603342.

[32] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography," *Pattern Recogn. Lett.*, vol. 124, no. C, pp. 82–90, Jun. 2019, ISSN: 0167-8655. DOI: 10.1016/j.patrec.2017.10.017. [Online]. Available: https://doi.org/10.1016/j.patrec.2017.10.017.

[33] T. Lin, M. Maire, S. J. Belongie, *et al.*, "Microsoft COCO: common objects in context," *CoRR*, vol. abs/1405.0312, 2014. arXiv: 1405.0312. [Online]. Available: http://arxiv.org/abs/1405.0312.

Hen-Wei Huang received his BS and MS in mechanical engineering from National Taiwan University, Taiwan, in 2011 and 2012, respectively. He received his PhD in robotics technology from ETH Zürich in 2018.

He is an Instructor in the Department of Medicine, Brigham and Women's Hospital, Harvard Medical School and a Visiting Scientist at the Koch Institute for Integrative Cancer Research at Massachusetts Institute of Technology. His research interests include in vivo wireless sensor network, personalized medicine, MEMS, robotics, and ingestible/implantable electronics.



Jack Chen received his BASc in Engineering Science from University of Toronto in 2022.

He is currently a graduate student in the Department of Mechanical Engineering at Massachusetts Institute of Technology. He was previously a research assistant with the Division of Gastroenterology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.

Philip Rupp received his B.S. and M.S degree in information technology and electrical engineering from ETH Zürich, Switzerland. He was a Research Assistant with the Division of Gastroenterology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA. He is working as a project engineer at Weider Wärmepumpen, an Austrian heat pump manufacturer. His research interests include HVAC technologies, machine learning, computer vision, brain–computer interfaces and nanophotonic devices.

Claas Ehmke received his B.Sc. in Information Technology and Electrical Engineering from Technical University Munich, Germany, in 2017 and his M.Sc. in Robotics, Systems, and Control from ETH Zürich in 2021. He is currently working towards a Ph.D. in Robotics at the Multi-Scale Robotics Lab from ETH Zürich, Switzerland.

He was previously a research assistant at Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, and worked on autonomous driving estimation algorithms at Singapore-MIT Alliance for Research and Technology Centre and AMZ racing.

Peter Chai recieved his MMS in artifical organs, biomaterials and cellular technology at the Alpert Medical School at Brown University. He ultimately obtained his MD from the Alpert Medical School and Brown Unviersity and completed residency training in emergency medicine at Brown. He additionally completed subspecialty training in medical toxicology at the University of Massachusetts Medical School.

He is an Associate Professor in the Department of Emergency Medicine, Brigham and Women's Hospital, Harvard Medical School and Affiliate Researcher at The Fenway Institute. His current research focuses on linked behavioral interventions that leverage ingestible, implantable and other computer vision based systems to detect individual changes in health.

Riya Dhar received her B.Tech in Electronics and Communication Engineering from Vellore Institute of Technology, in 2020. She was previously a research assistant at Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology. Currently she works as a Data Scientist in a Unicorn Silicon Valley startup.

Ian Ballinger received his B.S. in Electrical Engineering from the University of Vermont in 2020. He currently works for the Traverso Lab at Brigham and Women's Hospital as a Research Engineer II, and Massachusetts Institute of Technology as a Technical Associate.

Giovanni Traverso received his BA from Trinity College, University of Cambridge, UK in 1998, and his PhD from Johns Hopkins University in 2002. He subsequently completed medical school at the University of Cambridge, internal medicine residency at the Brigham and Women's Hospital, and his gastroenterology fellowship training at Massachusetts General Hospital, both at Harvard Medical School.

He is an Assistant Professor in the Department of Mechanical Engineering at the Massachusetts Institute of Technology, Cambridge, MA, and also a gastroenterologist in the Division of Gastroenterology, Brigham and Women's Hospital, Harvard Medical School. His current research program is focused on developing the next generation of drug delivery systems to enable safe and efficient delivery of therapeutics as well developing novel ingestible electronic devices for sensing a broad array of physiologic and pathophysiologic parameters. Additionally, Dr. Traverso continues his efforts towards the development of novel diagnostic tests that enable the early detection of cancer.