A Deployable Sensor Platform for Contactless Health Observations in Uncontrolled Environments

Anthony J. Regli* and Brian Serrano.†
University of Maryland: College Park, College Park, Maryland, 20740

This work aims to investigate and improve the capabilities of contactless vital sign monitoring systems (CVSMs) in environments with increased variance in lighting and subject pose using a deployable hardware platform designed for use in the DARPA Triage Challenge (DTC). Current research into CVSMs primarily focuses on their use in controlled environments, like sitting at a desk or driving a vehicle. The deployable platform, henceforth referred to as the Sensor Package for Updated Diagnosis (SPUD), is designed to operate in a disaster-response triage scenario as represented by the DTC, and consists of a Raspberry Pi along with a near-infrared time-of-flight (TOF) camera used to obtain intensity and depth images. Heart rate (HR) and respiratory rate (RR) are estimated from the datastream via a ROS2 pipeline that can then be integrated into a larger network of autonomous systems for potential use as part of the DTC. We analyze how the system performs under a variety of subject positions. We struggled to replicate the success of HR CVSM techniques, but research is continuing to improve the accuracy of this system. For an ideal subject seated at a desk, we find an RR estimation to be within 10% of ground truth in 85% of tests. In other testing configurations, RR estimate accuracy stays within 10% of ground truth in a minimum of 55% of tests.

I. Introduction

Mass casualty incidents (MCI) pose a significant challenge for disaster response professionals. Providing care for a large number of potential casualties necessitates the ability to rapidly assess and evaluate injury and potential risk. While initial assessment may be quick for many casualties, those who are assessed and deemed healthy enough to not need active medical intervention are at risk of their condition worsening due to an unnoticed injury, especially without active monitoring. In this situation, medical technicians require active monitoring of vital signs to remain aware of situational developments and to allow technicians to respond more rapidly to those with developing injuries.

A wide variety of medical technology exists to serve this purpose, but the majority of it is designed for use in hospitals and clinical settings where time is often a less-important factor, and where patients are typically static. In the case of an MCI, medics may not have time to equip casualties with monitors like respiratory belts and pulse oximeters. This paper presents a potential alternative method to continuously monitor casualty vitals in an MCI in the form of a deployable hardware platform known as the Sensor Package for Updated Diagnosis (SPUD), developed for use by the UMD Collective Dynamics and Controls Laboratory (CDCL) in the DARPA Triage Challenge. SPUD performs contactless vital sign monitoring (CVSM) through the use of an infrared/time of flight (TOF) camera.

As demonstrated by [1, 2], CVSM via RGB, infrared, and depth video is capable of accurately measuring both heart rate (HR) and respiratory rate (RR) of a subject in a controlled environment. Much of the available research focuses on subjects seated in controlled locations like at a desk or while driving a car, which is significantly different than the circumstances that would be expected in an MCI. By leveraging existing signal processing techniques in a new domain, we create a deployable hardware platform capable of estimating and communicating vital signs.

We will first cover relevant background to the signal processing techniques found in this work, then describe the specific software and hardware architecture used for SPUD. We will describe experimental setup and testing used to validate the platform. Lastly, we will analyze the results and discuss possible vectors for continued research and improvement of the platform.

*Undergraduate Researcher, University of Maryland, 756 E Passyunk Ave, Philadelphia, PA, AIAA University Student Member 1925070
†Undergraduate Researcher, University of Maryland, 14205 Chelmsford Rd, Rockville, MD, AIAA University Student Member 1928021.

II. Background

While there are CVSM techniques that use RGB cameras, SPUD uses a Time-of-Flight (TOF) camera for CVSM, which is capable of measuring both infrared intensity and image depth. As show in [2, 3], depth images enable accurate estimation of subject RR, and can also be used to improve accuracy of HR estimation. This camera also includes active illumination to allow control over subject lighting.

As part of our group's work on the DTC, several CVSM software packages have been developed within CDCL. For the work described in this paper, we specifically selected a software package based on the technique described in [3]. The general signal processing pipeline uses a synthesis of depth and intensity information to compensate for subject movement during HR estimation. More specifically, we sample a 50 second stream of video of a stationary subject. For each frame, we select 3 ROIs: one on the top of the subject's cheek (ROI 1), and 2 overlaying the subject's lungs (ROIs 2 and 3). We take the mean value of the intensity and depth image over ROI 1, along with the mean depth value over ROIs 2 and 3, giving us time-series datastreams for cheek image intensity, cheek depth, and depth signals for both lungs.

Fig. 1 Labelled ROIs from Camera Feed. ROI 1 denotes the subject's cheek, and ROIs 2 and 3 cover the subject's lungs

RR estimation uses a detrended signal from each lung ROI, attained by subtracting the sample mean from each dataset. After taking the Fourier transform of both signals, we select the frequency corresponding to the peak transform magnitude between 0.08 Hz and 0.667 Hz, corresponding to 4.8 and 40.02 breaths per minute respectively. This leaves us with 2 RR estimates, one for each lung signal.

If either estimate is zero (possible if a lung is obstructed for part of or all of the sample video), then whichever estimate is nonzero is returned. If both are nonzero and the difference between them is small (<0.1 BPM), then the left lung estimate is returned. If both estimates are nonzero and there is a significant difference between them, then the function returns the value corresponding to whichever lung ROI had the greater area over the course of the recording, as higher area likely implies more of the lung is visible.

HR estimation uses a combination of intensity and depth information to compensate for subject motion during the footage. As described in [3], when using an IR camera with active illumination, subject motion towards and away from the camera causes a noticeable change in intensity value that is independent of changes due to the subject's HR. This underlying motion signal will also be visible in the average cheek depth signal, and thus the depth signal can be used to compensate for subject motion. This package recalculates a compensated intensity signal. Given an initial intensity and depth signal I_{raw} and D_{raw}, we use the following equation from [3]. to calculate a compensated intensity signal that reduces the affect of subject motion on changes in intensity.

$$I_{comp} = \frac{I_{raw}}{a * (D_{raw})^{-b}}; \text{where } a, b = \underset{a,b}{\operatorname{argmin}}\, Correlation(\frac{I_{raw}}{a * (D_{raw})^{-b}}, D_{raw}) \tag{1}$$

Fig. 2 Intensity image from TOF Camera Feed

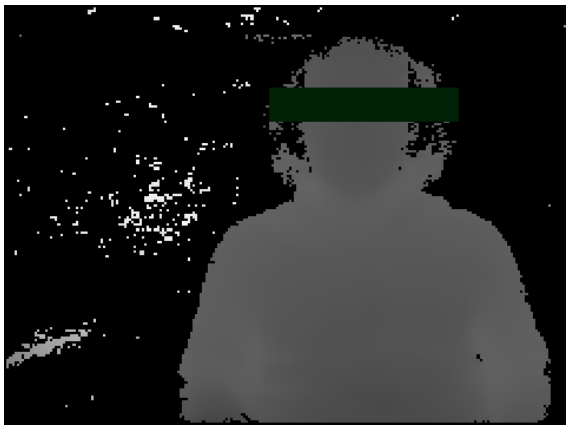


Fig. 3 Depth image from TOF camera feed

After intensity compensation, the signal is detrended by subtracting the mean signal value, and then a Fourier transform gives us the frequency spectrum. We take the frequency corresponding to the maximum Fourier magnitude between 0.8 Hz and 3 Hz (42 and 180 BPM) as our estimated HR from the video signal.

III. Methods

A. Hardware

SPUD is designed as a leave-behind sensor package to be deployed from robotic platforms. The short cylindrical form factor improves the probability of landing in the upright orientation for autonomous deployment. It also has a self-contained power source and compute capabilities to satisfy the leave-behind design constraint.

The sensor package contains a Raspberry Pi 5 and a Geekworm X1203 uninterruptible power supply for computing and power/battery management. A PMD Flexx2 TOF camera is utilized for HR and RR estimation. A reSpeaker USB 4-Mic Array XVF3000 v2.0 and two Raspberry Pi AI cameras are included for additional multimodal sensing. A 10,000 mAh LiPo battery provides uninterrupted power without access to external power.

Fig. 4 SPUD platform views from multiple perspectives.

B. Software

We used Raspberry Pi OS as our operating system because Raspberry Pi AI Cameras are currently only supported by Raspberry Pi OS. Our software platform runs in a Docker container running ROS2 Humble that opens at system log in. Several ROS2 nodes are used for vitals estimation; the first node runs in C++ and uses the Flexx2 C++ API to take images from the camera and send them via ROS2 topic to 2 processing nodes at a rate of 30 fps. The second processing

node runs in Python and uses a YOLOv8 keypoint detection algorithm to find subject keypoints based on the intensity image from the Flexx2 camera. Keypoints along with depth and intensity images are then sent to the vitals estimation node for recording and processing using the technique described in the methods section of this paper.

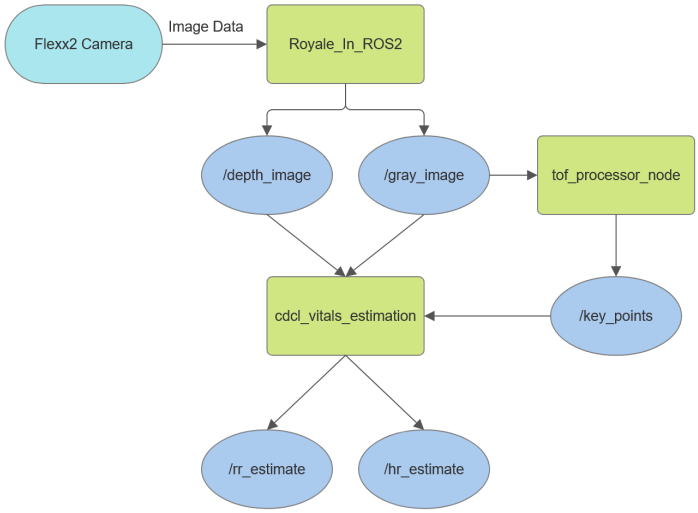


Fig. 5 Block Diagram of ROS2 Architecture.

C. Testing Methodology

SPUD was tested on 3 test subjects in varied configurations of both position and clothing. Each test run consisted of five 50-second measurement windows taken in succession. Each subject was tested in 4 configurations. Configuration 1 had subjects seated at a desk facing the camera at a distance of 1.5m while wearing thin clothes on their upper body, such as a t-shirt or a flannel button-down. Configuration 2 had subjects seated in the same position and distance, this time wearing a heavy winter jacket. Configuration 3 had subjects seated on the floor, to replicate a more likely position of a subject in an MCI, at a distance of 1.5m from the SPUD camera, again wearing standard clothing. Lastly, configuration 4 had subjects seated on the ground now wearing a thick jacket. We took ground truth HR measurements using a CVS Health™ C20 Pulse Oximeter. Ground truth RR measurements were taken with a Vernier Go Direct® Respiration Belt, which measures axial force on a belt placed around the subject's upper torso. To find respiratory rate from force measurements, we performed a Fourier transform on each 50 second window corresponding to the sample windows recorded by SPUD.

IV. Results

Across all testing configurations, SPUD was capable of reliably estimating RR for the majority of testing windows. We were unable to obtain an accurate HR measurement using the techniques described above, however potential causes of this inaccuracy and avenues for continued research will be discussed later in this paper. As seen in Fig. 7, in 71.9% of cases, our respiratory rate estimation was within a 10% error range of the true calculated value. In contrast, HR estimates in Fig. 6 were within this bound only 12.5% of the time. Changes in testing configuration also led to a decrease in RR estimation accuracy, though in all configurations the majority of RR estimates fell within a 10% error bound of ground truth. HR estimates were not significantly altered by changes in testing configuration; all configurations showed difficulty in obtaining an accurate HR estimate.

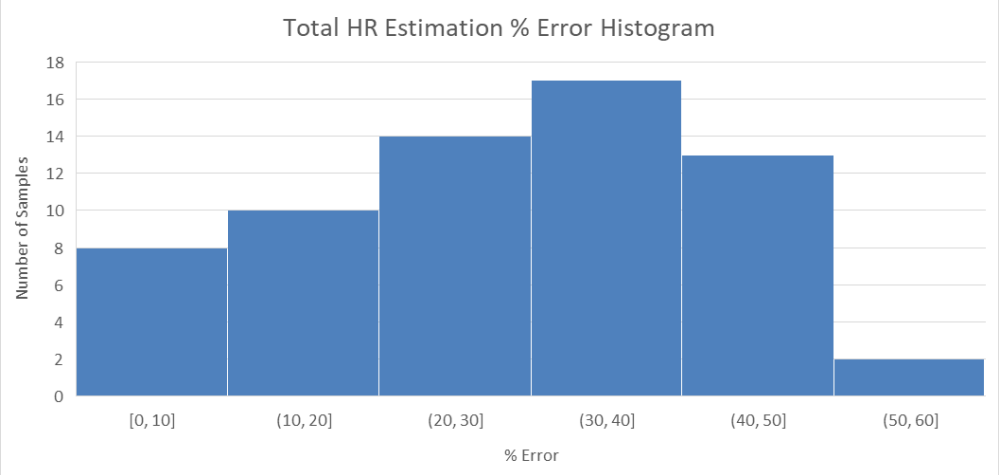


Fig. 6 HR Estimate percent error histogram across all testing data.

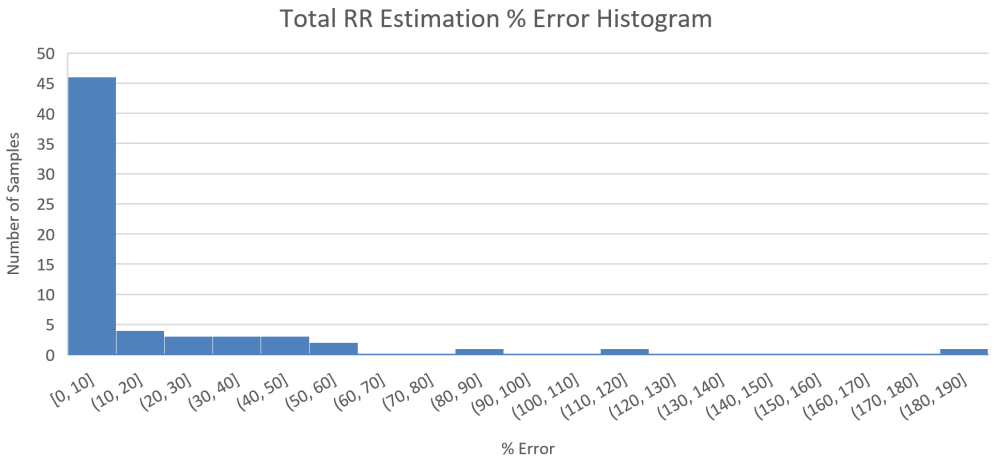


Fig. 7 RR Estimate percent error histogram across all testing data.

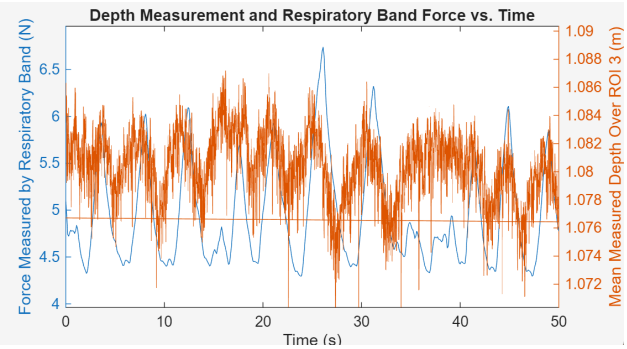


Fig. 8 Comparison of ROI 2 mean depth value with ground truth chest band force shows accurate measurement of underlying respiratory signal.

RR estimation showed differences in accuracy based on subject configuration. Subjects in configuration 1 showed the most accurate RR estimation, with 85% of estimates falling within a 10% error of ground truth (Fig. 9). HR

estimates in configuration 1 show no clear pattern. Subjects in configuration 3, seated on the floor with standard clothes, show a slight decrease in accuracy, with only 73% of estimates falling inside the 10% error bound of ground truth (Fig. 11). The next best accuracy was that of condition 2, with 66% of estimates falling within a 10% error bound of ground truth (Fig. 10). Lastly, condition 4 had 53% of RR estimates within 10% of ground truth (Fig. 12). The motion of subject's lungs is clearly visible within the measured depth signal when compared to the ground truth chest band force measurements, as shown in Fig. 8.

Table 1 List of Subject Testing Conditions

Configuration Number	Seat Position	Clothing Worn	% of samples within 10% error of ground truth RR
1	Desk	Standard Clothing	85%
2	Desk	Winter Jacket	65%
3	Floor	Standard clothing	73%
4	Floor	Winter Jacket	53%

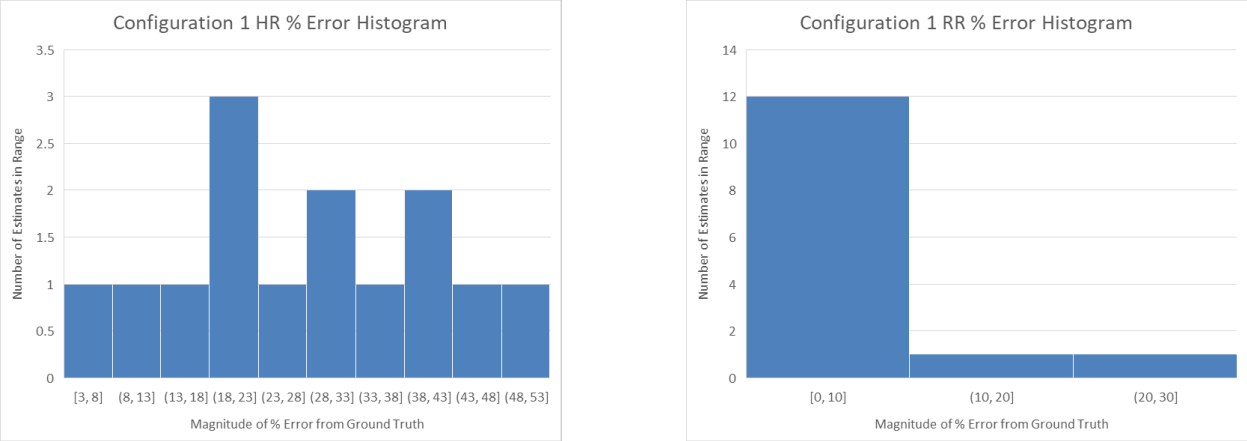


Fig. 9 Configuration 1 Error Histograms

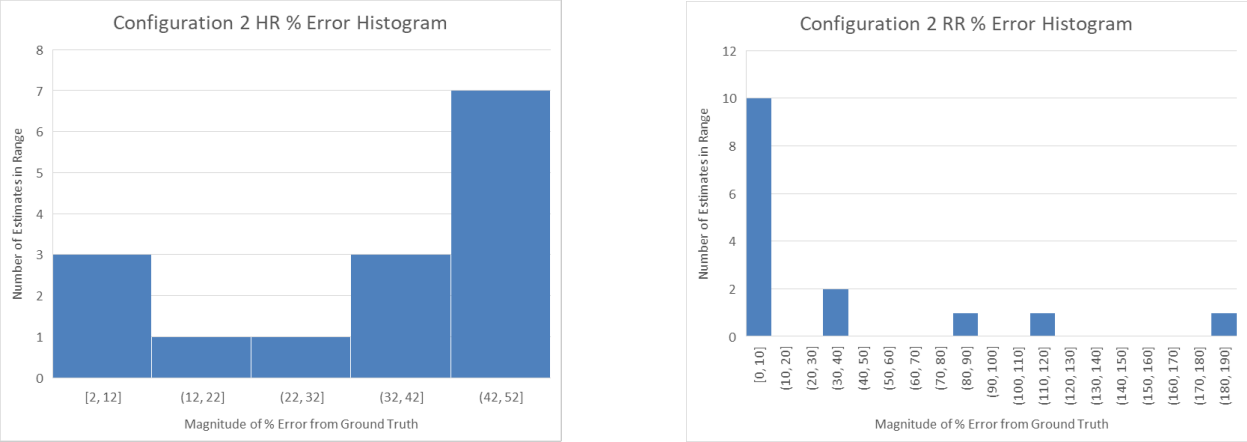
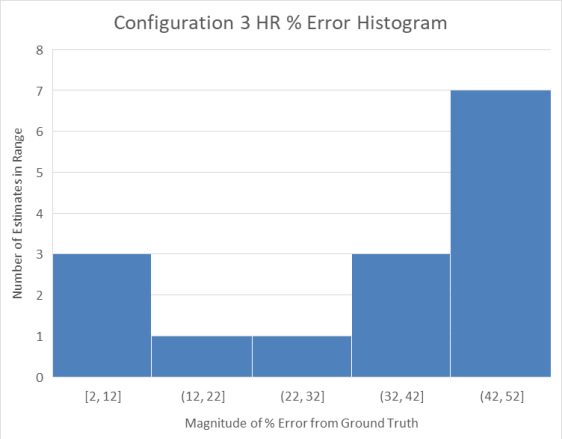


Fig. 10 Configuration 2 Error Histograms



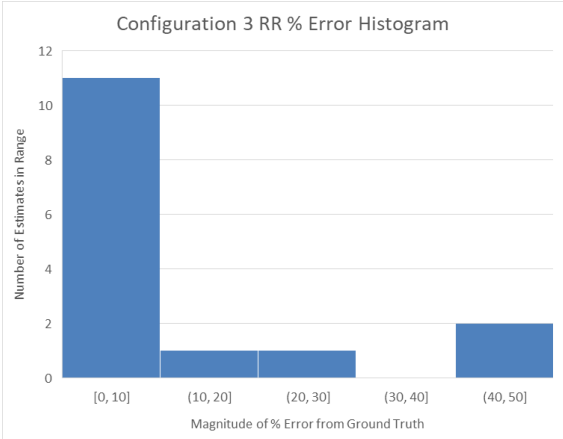
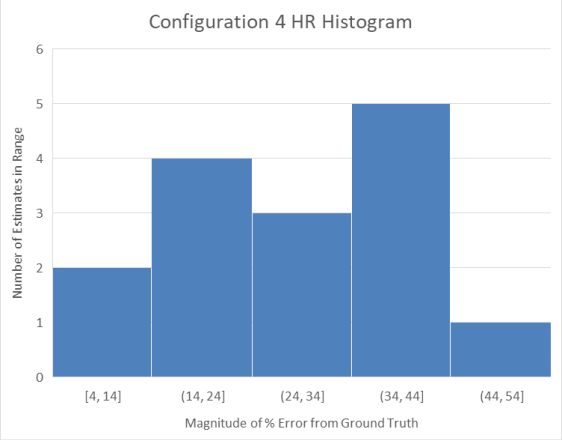


Fig. 11 Configuration 3 Error Histograms



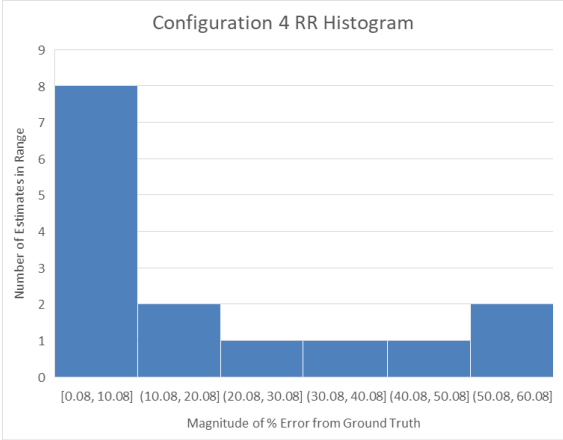


Fig. 12 Configuration 4 Error Histograms

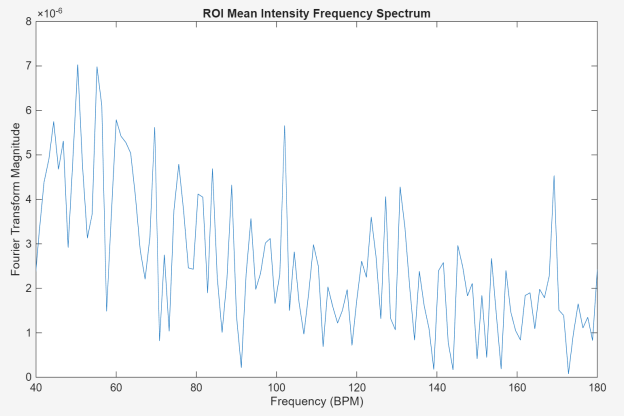


Fig. 13 Intensity Spectrum from Subject Cheek ROI

V. Discussion

In all testing configurations, SPUD provides an accurate means of estimating and monitoring subject RR. RR estimate accuracy appears to drop when the subject is wearing a jacket, indicating that thicker layers of fabric obstruct the depth signal being used to estimate RR. Additionally, the significant accuracy decrease between a subject wearing a jacket in a chair to a subject wearing a jacket on the ground may be due to the geometry of stiff fabric when worn seated on the floor. One subject commented on how the stiff fabric in their jacket made it puff out in front of their chest when seated on the ground. This puffing could create a gap between the subject's chest and the fabric of the jacket, making it difficult for the RR movement to make it through the jacket in a way that is visible to SPUD.

The inability to detect an accurate HR may be due to any number of factors. The vast majority of HR signals showed significant noise in the range of 40-60BPM. The source of this interference is yet to be determined, though several initial theories have been explored. The noise does not seem to be related to any individual subject's attributes, as all 3 subjects showed similar levels of noise. While intensity and depth image data is sampled at a rate of 30fps, due to computing power limitations, the keypoint detection algorithm runs at a rate of just under 10fps. The low frame rate of keypoints could mean that there are insufficient data to accurately obtain the HR signal. Research into increasing keypoint processing speed is ongoing.

Additionally, we are working on using independent component analysis (ICA) to further isolate variation in intensity signal that is due to sources other than the underlying HR signal. We plan to explore the use ICA of keypoint positions and intensity values of several pixels within the ROI to obtain a better underlying intensity signal. The other major goal of additional research is testing the system on a wider variety of subjects under a wider variety of conditions. The end goal is for SPUD to be deployed outdoors, thus additional verification is needed to ensure that the outdoor environment does not introduce unexpected inaccuracy in measurements.

VI. Conclusion

This work presents a novel hardware platform for CVSM in uncontrolled and dynamic environments like those that would be found in an MCI response scenario. We presented accurate estimation of RR using platform sensors, along with demonstrating the hardware required to make such a software platform deployable. We conducted tests to build on prior HR estimation research. While we were unable to replicate the success of prior HR estimation research, we plan to continue work using the possible avenues for research described in the discussion section to improve HR estimation.

Informed Consent

All participants in this study gave informed consent.

Acknowledgments

Special thanks to Dr. Derek Paley, Srijal Shekhar Poojari, Kleio Baxevani, and Zach Bortoff for their help, guidance, and mentorship throughout this project, and to all of our testing participants.

References

[1] Pourbemany, J., Essa, A., and Zhu, Y., "Real-Time Video-based Heart and Respiration Rate Monitoring," *NAECON 2021 - IEEE National Aerospace and Electronics Conference*, 2021, pp. 332–336. https://doi.org/10.1109/NAECON49338.2021.9696378.

[2] Guo, K., Zhai, T., Pashollari, E., Varlamos, C. J., Ahmed, A., and Islam, M. N., "Contactless Vital Sign Monitoring System for Heart and Respiratory Rate Measurements with Motion Compensation Using a Near-Infrared Time-of-Flight Camera," *Applied Sciences*, Vol. 11, No. 22, 2021. https://doi.org/10.3390/app112210913, URL https://www.mdpi.com/2076-3417/11/22/10913.

[3] Guo, K., Zhai, T., Purushothama, M. H., Dobre, A., Meah, S., Pashollari, E., Vaish, A., DeWilde, C., and Islam, M. N., "Contactless Vital Sign Monitoring System for In-Vehicle Driver Monitoring Using a Near-Infrared Time-of-Flight Camera," *Applied Sciences*, Vol. 12, No. 9, 2022. https://doi.org/10.3390/app12094416, URL https://www.mdpi.com/2076-3417/12/9/4416.