SDA Student Competition 2025
Remote PPG-based Vital Signs Monitoring
microSD Express-Enabled Remote PPG for 1440p Uncompressed Capture on Jetson Orin Nano

Shira Barmats Shaked Levi
Advisor: Yair Moshe

Signal and Image Processing Laboratory (SIPL)
Andrew and Erna Viterbi Faculty of Electrical & Computer Engineering
Technion – Israel Institute of Technology, Haifa, Israel, https://sipl.ece.technion.ac.il/

September 2025

Abstract

We present a dual-mode system for contactless vital-sign monitoring from 1440p facial video using remote photoplethysmography (rPPG). Running on an NVIDIA Jetson Orin Nano Super Developer Kit, our prototype records the uncompressed stream to a microSD Express card while simultaneously executing a downscaled, low-latency pipeline for live feedback. microSD Express over PCIe/NVMe allows sustained 30 fps uncompressed 1440p capture, where legacy microSD solutions struggle to sustain comparable bandwidth, making high-fidelity offline analysis practical. On a 10-recording dataset against a Polar Verity Sense reference, our system achieved 8.3 bpm MAE in real-time and 6.6 bpm MAE offline, reflecting the accuracy gains unlocked by preserving uncompressed 1440p video for post-capture processing. By quantifying I/O throughput and end-to-end performance, and demonstrating a functioning system, we show that microSD Express is a key enabler for edge vision workloads in consumer telehealth and wellness applications.

1 Introduction

Photoplethysmography (PPG) is an optical technique that senses changes in blood volume in the skin's microvasculature, typically using a light source and photodetector at the skin surface. The resulting waveform has been used to derive vital signs such as heart rate and respiratory rate. Remote PPG (rPPG) extends PPG into the imaging domain: a camera measures tiny, periodic color fluctuations from exposed skin (e.g., the face) without physical contact. rPPG enables continuous, non-contact vital-sign monitoring suited to telehealth and wellness settings where comfort and hygiene matter. Most rPPG pipelines share four stages: (1) face/skin detection and region-of-interest selection. (2) spatial aggregation to suppress noise. (3) temporal color processing. (4) physiological inference to estimate vital signs.

Recent surveys highlight increasing rPPG robustness in unconstrained environments, expanding feasibility for remote consultations and consumer devices, while also underscoring challenges from motion, illumination, and compression [1], [2]. These challenges can be mitigated by capturing streams of high spatial and temporal resolution, and retaining them in an uncompressed format. Higher spatial and temporal detail enables more stable motion tracking and better region-of-interest selection. More skin pixels improve signal quality as the signal-to-noise ratio (SNR) scales roughly with the square root of the number of independent samples. In addition, uncompressed capture and storage preserves subtle inter-frame color variations that common codecs often attenuate and that are important for accurate rPPG waveform reconstruction.

AI Use Disclosure: We used ChatGPT to: (i) edit English for clarity and fluency; (ii) suggest minor code refactors and fixes; (iii) generate icons used in Fig. 2. All algorithms, experiments, results, and interpretations are original; all AI outputs were reviewed and edited by the authors.

2 Proposed Solution

While high-resolution cameras have become inexpensive, the bandwidth and computational resources required to process every pixel in real-time remain prohibitive. Even on a capable edge platform such as an NVIDIA Jetson Orin Nano Super, sustaining full-resolution rPPG incurs unacceptable latency. This motivates our solution: capture and store the uncompressed 1440p stream via microSD Express, chosen over SSD for its ultra-compact, field-swappable form factor and the ability to physically transfer recordings between devices, while running a downscaled rPPG path for immediate feedback. High-precision offline processing is then applied to the stored data to refine estimates and extract additional metrics.

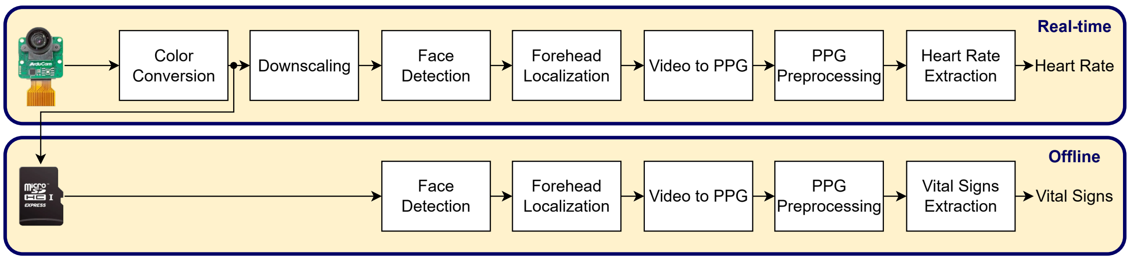


Figure 1: Overview of our prototype rPPG system.

As shown in Figure 1, the real-time and offline stages share key functional blocks: face detection, forehead localization, video-to-PPG conversion, and PPG pre-processing. In the real-time stage, a 45-seconds stream from a camera connected via the MIPI-CSI2 interface is captured at its native 2560×1440 resolution at 30 fps and written uncompressed to the microSD Express card for subsequent offline processing. Simultaneously, frames are downscaled spatially and temporally to 720p (1280×720) at 15 fps and processed at reduced resolution. Immediately after capture, the raw video stream is converted to a uniform color format. Because both color conversion and image downscaling operate over the full frame, they are the most compute-intensive stages and are executed on the Jetson GPU via GStreamer [3].

Face detection and forehead localization are performed with MediaPipe FaceMesh [4], which returns 3D facial landmarks for each frame. From selected landmarks we compute the interpupillary distance as a scale reference and estimate the forehead center. These anchors define a trapezoidal region of interest: its lower edge aligns with the brows, its upper edge extends above the forehead center, and its width and height are parameterized as fixed fractions of the interpupillary distance. To suppress frame-to-frame jitter, the polygon's vertices are temporally smoothed with an exponential moving average.

The cropped region-of-interest is temporally aggregated within a sliding window and is used to reconstruct a PPG waveform from the video. Over the past two decades, video-to-PPG methods have largely fallen into three families: blind-source-separation, model-based, and, more recently, learning-based approaches. In practice, blind-source-separation methods tend to deliver inferior results under motion, illumination changes, and variation in camera settings or skin tone. Learning-based methods can achieve strong performance, but they require large, diverse, well-labeled datasets with synchronized ground-truth PPG, often generalize poorly across cameras, frame rates, skin tones, and lighting, impose higher computational cost, and are less interpretable. Model-based methods offer competitive performance without these limitations, so we adopt this family. Among model-based techniques, two notable options are PBV (blood-volume pulse) and POS (plane orthogonal to skin) [5]. PBV [6] treats the blood-volume pulse as a subtle periodic modulation in skin reflectance and uses color-space projections to amplify the pulsatile component while suppressing motion and lighting artifacts. POS [7] projects the normalized RGB signal onto two chrominance directions orthogonal to the skin-tone vector, then combines them with adaptive weights to yield a robust PPG signal. Our system supports both POS and PBV, allowing the user to select either method.

The estimated PPG is then pre-processed to stabilize the waveform and improve downstream estimation. We remove the DC component and apply a band-pass filter, after which vital signs are estimated. This is done by identifying peaks in the PPG waveform and applying statistical analyses of them with the HeartPy toolkit [8]. In the real-time path, we report heart rate only. In the offline path, the higher-quality PPG signal and lack of latency constraints yield a more accurate heart-rate estimate and enable additional metrics, including heart-rate variability, respiration rate, and a stress/fatigue index. **A short video demonstrating our system is available at this link**

3 Performance Evaluation

We quantitatively evaluated both the accuracy of the estimated vital signs and the runtime system performance in the real-time and offline pipelines. To assess heart-rate accuracy, we compared our estimates against a Polar Verity Sense reference across five participants spanning different age groups and genders. The Polar Verity Sense is a commercially available arm-worn optical heart-rate monitor widely used in sports and validation studies. Each participant was measured twice, once at rest and once after brief physical activity to elevate heart rate, yielding ten recordings. Relative to the reference, our system with the POS video-to-PPG method [7] achieved a mean absolute error (MAE) of 8.3 bpm in real-time and 6.6 bpm offline. The improved offline accuracy is expected because offline processing yields a cleaner PPG waveform. For the remaining vital signs, no reference device was available. We therefore verified that estimates fell within physiologically plausible ranges and exhibited the expected increases from rest to post-exercise.

For performance evaluation, we profiled the major functional blocks of the system. Results are summarized in Table 1.

Table 1: Per-frame latency of major pipeline blocks in our system.

Functional block	**Real-time [ms]**	**Offline [ms]**
microSD Express card write	33	-
microSD Express card read	-	34
Downscaling	34	-
Face detection & forehead localization	22	43
Video-to-PPG (POS)	14	26
Preprocessing & vital-sign estimation	10	10
GUI & other	6	6
End-to-end processing	52	85

The last row of Table 1 reports the end-to-end per-frame latency. For 1440p offline processing, the latency is 85 ms ($\approx$ 11.7 fps), which is insufficient for 30 fps real-time operation. For real-time processing at 720p resolution the end-to-end latency is 52 ms ($\approx$ 19.2 fps) which is enough for real-time processing at 15 fps. The first row of Table 1 reports the per-frame card write time. In our real-time pipeline, writing a frame to the card takes about 33 ms (i.e., $\approx$ 30.3 fps), sufficient for sustained real-time. At 1440p with 1.5 bytes/pixel (YUV 4:2:0, 8-bpp), the sustained write bandwidth is

$$2560 \times 1440 \times 1.5 \times 30 \approx 1.67 \times 10^8 \text{ B/s} (\approx 167 \text{ MB/s}).$$

This level of performance is achievable with microSD Express. Legacy microSD cards cannot sustain such bandwidth and therefore do not support real-time operation with 1440p video, making microSD Express a key practical enabler for our application.

Appendix A microSD Express Card Performance Evaluation

We use a 256 GB SanDisk microSD Express card. To evaluate the card's real-world throughput over PCIe/NVMe, we used Ubuntu's `fio` (Flexible I/O Tester), which generates synthetic I/O workloads and reports bandwidth and latency. We ran `fio` with buffered I/O for 45 seconds to emulate one rPPG capture session of our application, yielding a sustained write rate of 186 MB/s. This rate is within $\approx 11\%$ of the sustained write rate of 210 MB/s reported in the card specification [9]. From the measured write bandwidth, the practical upper bound for writing uncompressed 1440p video at 1.5 bytes/pixel (YUV 4:2:0, 8-bpp) can be computed:

$$\text{fps}_{\text{max}} = \frac{186 \times 10^6}{2560 \times 1440 \times 1.5} \approx 33.5 \text{ fps}$$

Consistent with this bound, our application reliably writes 1440p video at 30 fps, closely approaching the computed upper bound.

Appendix B System Configuration

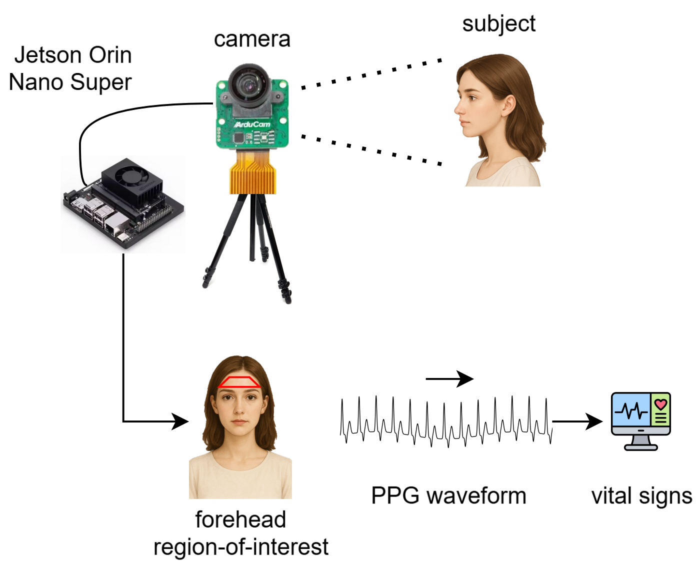
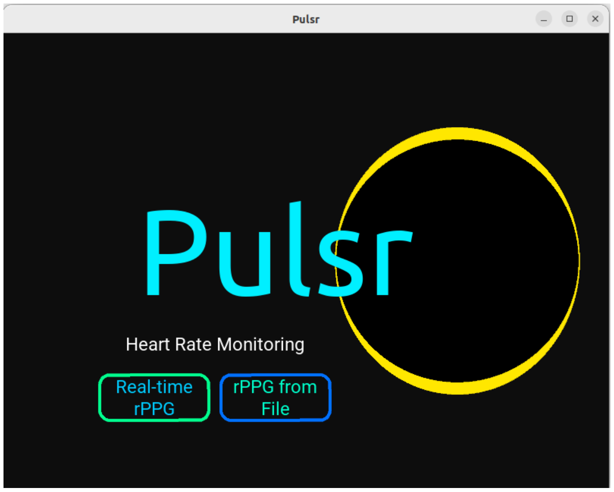
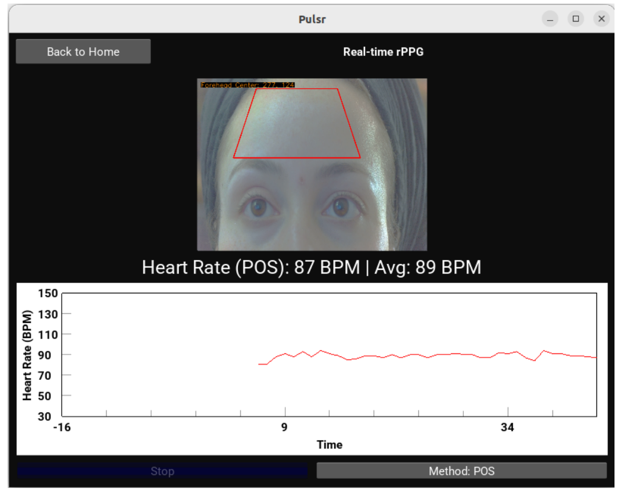


Figure 2: System configuration. Video of a seated subject is captured by a fixed Arducam IMX477 Camera Module and processed on the Jetson Orin Nano Super Developer Kit. A forehead region of interest is localized. From this region of interest, a PPG waveform is reconstructed to estimate vital signs, which are then presented to the user.

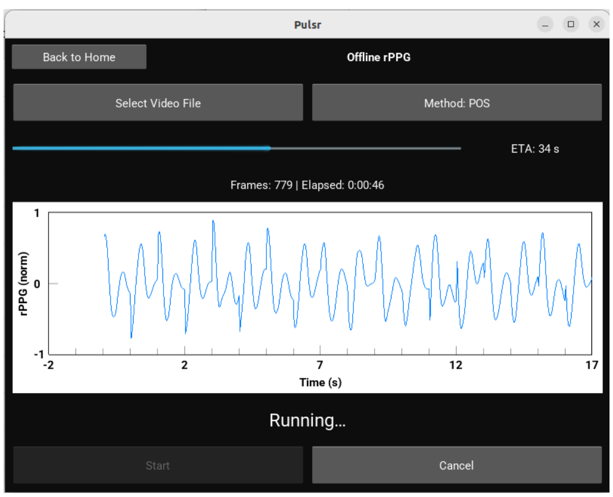
Appendix C GUI Screenshots



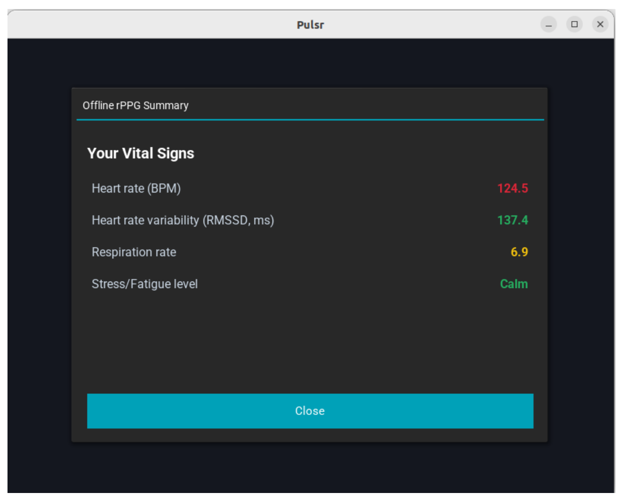
(a) Home screen with mode selection (real-time or offline).



(b) Real-time mode interface. Heart rate estimates are displayed in real-time. The forehead ROI is marked by a red trapezoid.



(c) Offline mode interface showing the rPPG waveform during processing and a progress bar.



(d) Vital-sign estimates produced after offline processing. Normal values are green while elevated values are orange or red.

Figure 3: Screenshots of the application GUI.

References

[1] H. Xiao, T. Liu, Y. Sun, Y. Li, S. Zhao, and A. Avolio, "Remote photoplethysmography for heart rate measurement: A review," *Biomedical Signal Processing and Control*, vol. 88, p. 105 608, 2024.

[2] P. Pirzada, A. Wilde, and D. Harris-Birtill, "Remote photoplethysmography for heart rate and blood oxygenation measurement: A review," *IEEE Sensors Journal*, vol. 24, no. 15, pp. 23 436–23 453, 2024.

[3] *GStreamer, Open-source multimedia framework*, version 1.26.5, GStreamer Project. [Online]. Available: `https://gstreamer.freedesktop.org/`.

[4] C. Lugaresi, J. Tang, H. Nash, *et al.*, "MediaPipe: A framework for perceiving and processing reality," in *Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR)*, vol. 2019, 2019.

[5] X. Liu, G. Narayanswamy, A. Paruchuri, *et al.*, "rPPG-Toolbox: Deep remote PPG toolbox," *Advances in Neural Information Processing Systems*, vol. 36, pp. 68 485–68 510, 2023.

[6] G. De Haan and A. Van Leest, "Improved motion robustness of remote-PPG by using the blood volume pulse signature," *Physiological measurement*, vol. 35, no. 9, p. 1913, 2014.

[7] 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, 2016.

[8] P. Van Gent, H. Farah, N. Van Nes, and B. Van Arem, "HeartPy: A novel heart rate algorithm for the analysis of noisy signals," *Transportation research part F: traffic psychology and behaviour*, vol. 66, pp. 368–378, 2019.

[9] "SanDisk microSD Express microSD Card — Data Sheet," SanDisk Corporation. (2024), [Online]. Available: `https://documents.westerndigital.com/content/dam/doc-library/en_us/assets/public/sandisk/product/memory-cards/microsd-express/data-sheet-microsd-express.pdf`.