Projection-based Robust Interpretable Signal Mixing for Remote Heart Rate Estimation

Cecilia G. Morales[1*†] cgmorale@andrew.cmu.edu
Fanurs Chi-En Teh[1*†] fanurs@cmu.edu
Kai Li[1,2*] kaii.li@mail.utoronto.ca
Pushpak Agrawal[1,3] pushpaka@andrew.cmu.edu
Artur Dubrawski[1] awd@cs.cmu.edu

[1]Carnegie Mellon University, USA [2]University of Toronto, Canada [3]Vellore Institute of Technology

Abstract

Remote photoplethysmography (rPPG) enables contactless vital sign monitoring using standard RGB cameras. However, existing methods rely on fixed parameters optimized for particular lighting conditions and camera setups, limiting adaptability to diverse deployment environments. This paper introduces the Projection-based Robust Signal Mixing (PRISM) algorithm, a training-free method that jointly optimizes photometric detrending and color mixing through online parameter adaptation based on signal quality assessment. PRISM achieves state-of-the-art performance among unsupervised methods, with MAE of 0.77 bpm on PURE and 0.66 bpm on UBFC-rPPG, and accuracy of 97.3% and 97.5% respectively at a 5 bpm threshold. Statistical analysis confirms PRISM performs equivalently to leading supervised methods ($p > 0.2$), while maintaining real-time CPU performance without training. This validates that adaptive time series optimization significantly improves rPPG across diverse conditions.

1 Introduction

Remote photoplethysmography (rPPG) enables contactless cardiac monitoring by detecting pulse-induced color variations on skin using standard RGB cameras to extract physiological time series [1]. At its core, rPPG requires combining raw RGB channels into a single pulse signal, with the combination strategy directly affecting performance and generalizability across applications such as telemedicine consultations and spotting driver drowsiness [2]. While supervised deep models can adapt to diverse conditions, they require extensive domain-specific datasets and substantial computation [3]. Current unsupervised methods remain interpretable and computationally efficient but rely on fixed parameters, limiting their adaptability to real-world lighting variations [4].

The plane-orthogonal-to-skin (POS) algorithm exemplifies this trade-off: it removes low-frequency drift, suppresses specularities via a projection orthogonal to mean skin color, and mixes channels under an assumption of stable spectral composition with varying intensity [5]. Such assumptions hold reasonably well indoors under artificial lighting (fluorescent/incandescent), but natural illumination is nonstationary; the incident spectrum varies with solar angle, atmosphere, shadows, and scene context, so fixed projection axes can fail when spectra and luminance covary [6]. Invariance-based approaches such as LGI mitigate some nuisance variation but remain static once chosen [7].

We build on the POS algorithm's structure by retaining its interpretable projection design, but replace the fixed color mixing with two adaptive parameters, α and λ, which are optimized per window

[*]Co-first authors contributed equally to this work.
[†]Corresponding authors

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Learning from Time Series for Health.

to maximize heart rate signal quality. This principled objective yields a globally pulse-dominated yet temporally smooth signal that resists illumination shifts and motion, while remaining fully interpretable and CPU-efficient. It preserves POS's strengths while significantly boosting accuracy for autonomous or human-in-the-loop care.

This paper introduces PRISM, a training-free method that adaptively optimizes rPPG signal extraction through online parameter selection. We demonstrate state-of-the-art unsupervised performance in PURE and comparable to supervised methods in UBFC-rPPG while maintaining real-time efficiency.

2 Related Works

While supervised deep learning approaches have shown promising results in rPPG, they face significant challenges in generalization across diverse populations, lighting conditions, and camera systems due to their dependence on large labeled datasets [8]. Unsupervised methods address these limitations by relying on fundamental physiological and optical principles rather than learned patterns. Since PRISM is a training-free algorithm, we focus this review on unsupervised rPPG approaches that are computationally efficient enough for real-time processing [2]. These methods typically rely on color space transformations, blind source separation, and physiological priors to extract the pulse signal from raw video data [9].

De Haan and Jeanne proposed the CHROM method, which used a skin reflection model to compute orthogonal chrominance signals, achieving 92% agreement with contact PPG and improved motion robustness over blind source separation like ICA and PCA [10]. Wang et al. extended this with the Plane-Orthogonal-to-Skin (POS) method, leveraging optical and physiological skin modeling. POS outperformed CHROM, Green, PCA, and ICA in varying skin tones and activity levels in controlled settings [5].

More recent methods include Orthogonal Matrix Image Transformation (OMIT) using QR decomposition and Local Group Invariance (LGI) with signal priors. LGI was designed to handle adaptive lighting via invariance priors [7], it still relies on predetermined algorithmic choices that may not generalize across all conditions. PRISM addresses these limitations through a different approach. Rather than using fixed parameters or invariance assumptions, PRISM maintains POS's interpretable green minus red/blue projection while enabling dynamic optimization rather than fixed parameter selection. Unlike existing approaches, PRISM provides a unified framework that continuously adapts both color mixing and temporal filtering parameters based on real-time signal quality, ensuring optimal performance across diverse lighting conditions, motion artifacts, and individual physiological variations without requiring parameter tuning.

3 Methods

PRISM adaptively selects projection and detrending parameters based on a signal quality objective. It extends the structure of the Plane-Orthogonal-to-Skin (POS) algorithm [5], while replacing its fixed projection and filtering parameters with ones optimized online for each video.

Time series of RGB values are extracted by averaging over a facial region of interest detected and tracked throughout the video using YOLOv5 [11]. Each raw channel $X(t) \in \{R(t), G(t), B(t)\}$ is normalized by a smooth spline baseline $\tilde{X}(t)$, yielding $\hat{X}(t) = X(t)/\tilde{X}(t)$, to suppress slow changes from illumination and sensor gain while preserving fine-grained pulse signals. A smoothness parameter $\lambda > 0$ controls the spline, expressed in units of s^3 throughout this text.

To extract the pulse signal, we form a linear combination of the normalized color channels. The POS method projects RGB signals into a chrominance plane orthogonal to the intensity direction and combines two empirically chosen components using adaptive weighting. PRISM preserves this structure but restricts to a single axis within that plane, forming a direct projection:

$$s(t) = \hat{G}(t) - \left(\alpha \hat{B}(t) + (1-\alpha)\hat{R}(t)\right) , \tag{1}$$

where $\alpha \in [0, 1]$ controls the red–blue balance. This construction remains interpretable and allows adaptive selection of the projection axis. The pulse signal $s(t)$ is then split into non-overlapping windows of $\delta t = 10$ seconds, and we extract a heart rate estimate h_i from each.

To select optimal parameters (λ^*, α^*), we define an objective that balances two properties of the signal $s(t)$: spectral concentration within the physiological band and temporal stability of the estimated heart rate. Specifically, we minimize:

$$(\lambda^*, \alpha^*) = \arg\min_{\lambda,\alpha} \left[k \cdot \mathrm{TV}(\lambda, \alpha) - C(\lambda, \alpha)\right] \ , \tag{2}$$

where $k = 1/3$ is set empirically, C measures the fraction of power concentrated in the plausible heart rate band (typically 0.5-4.0 Hz), and TV is the temporal variation that quantifies how much the heart rate estimates vary between consecutive windows, defined as $\mathrm{TV} = \frac{1}{t_n - t_1} \sum_{i=1}^{n-1} |h_{i+1} - h_i|$.

This encourages both sharp spectral peaks and smoothly evolving heart rate sequences. The scalar k balances the two terms. Empirically, this objective correlates strongly with lower heart rate error across datasets. The search runs over a small grid of (λ, α) values and requires no training data. Once selected, (λ^*, α^*) are held fixed and used to compute final heart-rate estimates. Additional implementation details are provided in Appendix B.

3.1 Datasets

To assess the performance of PRISM under realistic conditions, including motion and varied illumination, we utilize the two standard rPPG benchmark datasets.

PURE [12]: A widely used real-world dataset consisting of RGB videos collected using eco274CVGE camera at resolution (640×480, 30 Hz) from 10 subjects (8 male, 2 female) performing six head motion tasks ranging from steady sitting to talking, translation, and rotation. Synchronized ground truth PPG and SpO_2 signals were recorded using a finger-clip oximeter at 60 Hz .

UBFC-rPPG [13]: A commonly used dataset consisting of 42 RGB videos (640×480, 30 Hz) recorded indoors with a Logitech C920 HD Pro webcam under natural and artificial lighting. PPG signals were simultaneously captured using a finger-clip pulse oximeter.

3.2 Implementation Enhancements to rPPG-Toolbox

We used the rPPG-toolbox [14] to ensure reproducibility, but implemented several modifications to enhance heart rate estimation accuracy across all methods.

Frequency Resolution: We increased the FFT bin count from $N = 2^9 = 512$ bins (frequency resolution of approximately 3.5 bpm) to $N = 2^{14}$, achieving a substantially finer resolution of approximately 0.1 bpm.

Temporal Windowing: We replaced the default 60-second analysis windows with 10-second non-overlapping windows. Shorter windows better track natural heart rate variability and provide more temporally precise measurements, particularly during motion scenarios.

Face Detection: Our ablation studies revealed that YOLOv5 [11], a learning-based face detector already available in the toolbox, significantly outperforms the default HaarCascade method. Superior detection accuracy ensures consistent face region-of-interest (ROI) extraction, reducing signal contamination from background pixels.

Dynamic Face Tracking: We enabled dynamic face tracking to continuously adapt to subject movement throughout the video duration, replacing the toolbox's default static tracking assumption. Continuous adaptation to head motion maintains ROI alignment, which is critical for motion-heavy scenarios like PURE.

ROI Configuration: We configured the bounding box parameter by setting `largebox = False`, thereby minimizing background interference and ensuring more accurate extraction of the facial region of interest.

These modifications improve the performance of all evaluated methods, as demonstrated in Appendix D.

4 Results and Discussion

We evaluate performance using five metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), standard deviation (SD), Pearson correlation (R), and accuracy within a 5 bpm threshold,

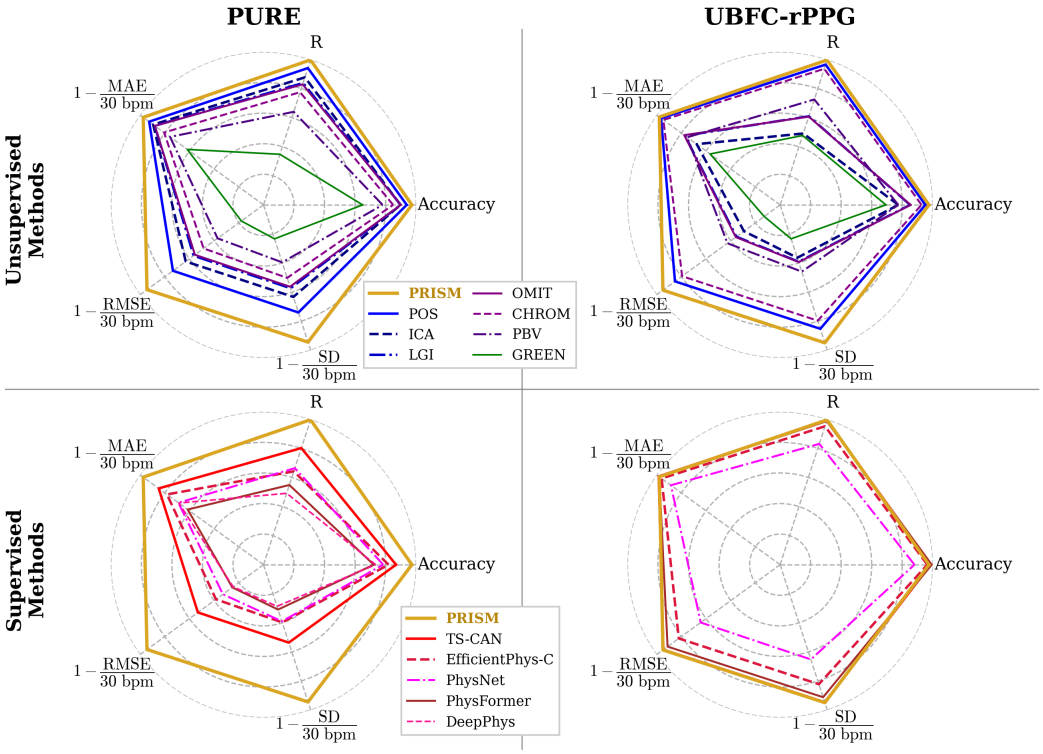

Figure 1: Radar plots comparing performance across five evaluation metrics on the PURE and UBFC-rPPG datasets. PRISM outperforms other methods across most of the metrics. All metrics are normalized such that larger area indicates better performance.

where predictions within ± 5 bpm of the ground truth are considered correct. This method of computing the accuracy provides an intuitive performance metric that aligns with clinically acceptable error ranges for heart rate monitoring and is more readily interpretable by medical practitioners than traditional correlation-based metrics. They are computed on non-overlapping 10-second windows to reflect real-time tracking performance and account for natural heart rate variability, results are shown in Table 1. Figure 1 summarizes performance across all methods using radar plots, with each axis normalized such that larger area indicates better performance.

PRISM demonstrates superior performance among other unsupervised methods across both datasets. On PURE, PRISM achieves the lowest MAE, 0.77 bpm, and RMSE, 1.69 bpm, significantly outperforming the previous best unsupervised method, POS, which achieved 2.23 bpm MAE and 8.00 RMSE. This corresponds to a 65% reduction in MAE and 79% reduction in RMSE. On UBFC-rPPG, PRISM achieves 0.66 bpm MAE and 1.54 bpm RMSE, again outperforming POS's 1.08 bpm MAE while improving RMSE by a substantial 65%.

The accuracy results further demonstrate PRISM's effectiveness, achieving 97.3% accuracy on PURE and 97.5% accuracy on UBFC-rPPG, outperforming all other unsupervised methods. Notably, PRISM's performance approaches that of state-of-the-art supervised methods while maintaining the computational efficiency and deployment advantages of unsupervised approaches. Namely, on UBFC-rPPG, PRISM matches the accuracy of 97.5% of DeepPhys and approaches TSCAN's 0.51 bpm MAE performance.

Statistical analysis using paired t-tests confirms PRISM's performance on UBFC-rPPG is statistically equivalent to leading supervised methods, TS-CAN ($p = 0.208$), PhysFormer ($p = 0.880$), and DeepPhys ($p = 0.595$). Thus, PRISM achieves statistically comparable performance to deep learning approaches without requiring training data.

Both PURE and UBFC-rPPG datasets contain natural varying lighting conditions, demonstrating PRISM's adaptability compared to methods designed for artificial illumination or via invariance

priors. Additionally, Appendix C presents PRISM's per-condition results on PURE, showing robust performance across different motion scenarios.

These results validate PRISM's adaptive parameter selection strategy, demonstrating that online optimization of color mixing and temporal filtering parameters can significantly improve rPPG extraction quality across diverse experimental conditions and datasets.

Table 1: Comprehensive performance comparison across all metrics on the PURE and UBFC-rPPG benchmark datasets. PRISM's performance is highlighted. MAE, RMSE, and SD are in bpm. R denotes Pearson correlation. Acc. is the percentage of estimates within a ± 5 bpm threshold.

		Test Set									
		PURE [12]					UBFC-rPPG [13]				
	Method	MAE$^{\downarrow}$	RMSE$^{\downarrow}$	SD$^{\downarrow}$	R$^{\uparrow}$	Acc.$^{\uparrow}$	MAE$^{\downarrow}$	RMSE$^{\downarrow}$	SD$^{\downarrow}$	R$^{\uparrow}$	Acc.$^{\uparrow}$
Unsupervised	GREEN [15]	11.48	24.62	22.99	0.348	64.9%	13.01	26.13	23.00	0.477	69.3%
	ICA [16]	3.20	11.11	11.01	0.877	89.7%	9.73	21.21	19.09	0.491	77.4%
	CHROM [10]	5.84	15.45	14.94	0.775	84.9%	1.65	6.20	6.11	0.936	92.6%
	PBV [17]	7.30	18.84	18.14	0.640	78.0%	7.44	17.20	16.23	0.725	77.4%
	OMIT [18]	3.62	13.33	13.12	0.829	89.9%	6.72	19.24	18.19	0.607	85.5%
	POS [5]	2.23	8.00	7.81	0.941	93.8%	1.08	4.46	4.44	0.965	95.1%
	LGI [7]	3.58	13.02	12.81	0.838	89.9%	6.88	19.37	18.28	0.609	85.2%
	PRISM (Ours)	**0.77**	**1.69**	**1.69**	**0.997**	**97.2%**	0.66	**1.54**	**1.54**	**0.996**	97.5%
Supervised	DeepPhys [19]	9.42	22.62	21.41	0.491	73.5%	0.58	1.70	1.69	0.995	97.5%
	PhysNet [20]	9.33	20.13	18.31	0.664	78.4%	3.48	10.71	10.46	0.831	87.9%
	TS-CAN [21]	4.54	14.06	13.93	0.802	86.8%	**0.51**	1.64	1.64	0.995	98.6%
	PhysFormer [22]	11.58	22.43	20.74	0.548	72.2%	0.64	2.66	2.66	0.987	**99.3%**
	EfficientPhys-C [23]	6.55	18.45	18.10	0.642	81.9%	1.22	5.39	5.35	0.952	96.4%

5 Conclusions

PRISM demonstrates that unsupervised methods can match or exceed supervised performance through adaptive optimization. While these results are promising, the current evaluation focuses primarily on lighter skin tones present in the tested datasets, and more comprehensive testing across diverse skin tones is needed to fully validate the method's generalizability. Additionally, ongoing work is testing PRISM in more challenging outdoor and robot-mounted scenarios where environmental conditions are completely uncontrolled, which will provide further insights into the method's robustness and outdoor deployment potential. We also plan to integrate PRISM in the rPPG-Toolbox repository for easier access and reproducibility.

Acknowledgments and Disclosure of Funding

This work has been partially supported by the Defense Advance Research Projects Agency (award HR00112420329) and National Science Foundation (awards 2427948 and 2406231).

References

[1] Daniel McDuff. Camera Measurement of Physiological Vital Signs. *ACM Comput. Surv.*, 55(9), January 2023. ISSN 0360-0300. doi: 10.1145/3558518. URL `https://doi.org/10.1145/3558518`.

[2] Ru Jing Lee, Saaveethya Sivakumar, and King Hann Lim. Review on remote heart rate measurements using photoplethysmography. *Multimedia Tools and Applications*, 83(15):44699–44728, 2024.

[3] Aoxin Ni, Arian Azarang, and Nasser Kehtarnavaz. A review of deep learning-based contactless heart rate measurement methods. *Sensors*, 21(11):3719, 2021.

[4] Daniele Di Lernia, Gianluca Finotti, Manos Tsakiris, Giuseppe Riva, and Marnix Naber. Remote photoplethysmography (rPPG) in the wild: Remote heart rate imaging via online webcams. *Behavior Research Methods*, 56(7):6904–6914, October 2024. doi: 10.3758/s13428-024-02398-0. Epub 2024 Apr 17.

[5] Wenjin Wang, Albertus C. den Brinker, Sander Stuijk, and Gerard de Haan. Algorithmic Principles of Remote PPG. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2017. doi: 10.1109/TBME.2016.2609282.

[6] Yu-Chen Lin and Yuan-Hsiang Lin. A study of color illumination effect on the SNR of rPPG signals. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4301–4304, 2017. doi: 10.1109/EMBC.2017.8037807.

[7] Christian S Pilz, Sebastian Zaunseder, Jarek Krajewski, and Vladimir Blazek. Local group invariance for heart rate estimation from face videos in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 1254–1262, 2018.

[8] Chun-Hong Cheng, Kwan-Long Wong, Jing-Wei Chin, Tsz-Tai Chan, and Richard HY So. Deep learning methods for remote heart rate measurement: A review and future research agenda. *sensors*, 21(18):6296, 2021.

[9] Uday Debnath and Sungho Kim. A comprehensive review of heart rate measurement using remote photoplethysmography and deep learning. *BioMedical Engineering OnLine*, 24(1):73, 2025.

[10] Gerard de Haan and Vincent Jeanne. Robust Pulse Rate From Chrominance-Based rPPG. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013. doi: 10.1109/TBME.2013.2266196.

[11] Ultralytics. YOLOv5 by Ultralytics, 2025. URL `https://github.com/ultralytics/yolov5`.

[12] R. Stricker, S. Müller, and H.-M. Gross. Non-contact Video-based Pulse Rate Measurement on a Mobile Service Robot. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 1056–1062, Edinburgh, Scotland, UK, 2014. IEEE.

[13] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, 2017.

[14] Xin Liu, Girish Narayanswamy, Akshay Paruchuri, Xiaoyu Zhang, Jiankai Tang, Yuzhe Zhang, Roni Sengupta, Shwetak Patel, Yuntao Wang, and Daniel McDuff. rPPG-Toolbox: Deep Remote PPG Toolbox. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 68485–68510. Curran Associates, Inc., 2023. URL `https://proceedings.neurips.cc/paper_files/paper/2023/file/d7d0d548a6317407e02230f15ce75817-Paper-Datasets_and_Benchmarks.pdf`.

[15] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Opt. Express*, 16(26):21434–21445, Dec 2008. doi: 10.1364/OE.16.021434. URL `https://opg.optica.org/oe/abstract.cfm?URI=oe-16-26-21434`.

[16] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard. Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam. *IEEE Transactions on Biomedical Engineering*, 58(1): 7–11, 2011. doi: 10.1109/TBME.2010.2086456.

[17] G de Haan and A van Leest. Improved motion robustness of remote-PPG by using the blood volume pulse signature. *Physiological Measurement*, 35(9):1913, aug 2014. doi: 10.1088/0967-3334/35/9/1913. URL `https://dx.doi.org/10.1088/0967-3334/35/9/1913`.

[18] Constantino Álvarez Casado and Miguel Bordallo López. Face2PPG: An Unsupervised Pipeline for Blood Volume Pulse Extraction From Faces. *IEEE Journal of Biomedical and Health Informatics*, 27(11): 5530–5541, 2023. doi: 10.1109/JBHI.2023.3307942.

[19] Weixuan Chen and Daniel McDuff. DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[20] Zitong Yu, Xiaobai Li, and Guoying Zhao. Remote Photoplethysmograph Signal Measurement from Facial Videos Using Spatio-Temporal Networks. In *Proc. BMVC*, 2019.

[21] Xin Liu, Josh Fromm, Shwetak Patel, and Daniel McDuff. Multi-Task Temporal Shift Attention Networks for On-Device Contactless Vitals Measurement. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 19400–19411. Curran Associates, Inc., 2020. URL `https://proceedings.neurips.cc/paper_files/paper/2020/file/e1228be46de6a0234ac22ded31417bc7-Paper.pdf`.

[22] Zitong Yu, Yuming Shen, Jingang Shi, Hengshuang Zhao, Philip H.S. Torr, and Guoying Zhao. PhysFormer: Facial Video-Based Physiological Measurement With Temporal Difference Transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4186–4196, June 2022.

[23] Xin Liu, Brian Hill, Ziheng Jiang, Shwetak Patel, and Daniel McDuff. EfficientPhys: Enabling Simple, Fast and Accurate Camera-Based Cardiac Measurement. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 5008–5017, January 2023.

[24] Carl De Boor and Carl De Boor. *A practical guide to splines*, volume 27. springer New York, 1978.

[25] George S Kimeldorf and Grace Wahba. A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics*, 41(2):495–502, 1970.

A Ablation

To understand PRISM's performance and validate our design choices, we performed an extensive ablation study, presented in Table 2. We evaluated several variants of our method:

- **Full PRISM:** The complete model with joint optimization of both the color mixing weight α and the smoothness parameter λ.
- **Best Fixed α / λ:** The model with one parameter fixed, while the other is optimized. This isolates the contribution of adapting each parameter individually.
- **Concentration-only:** The model optimizes the objective function without the temporal variation penalty ($k = 0$), relying solely on spectral concentration.
- **TV-only:** Model optimizes the objective function without concentration value.

The results demonstrate that each component of PRISM is critical. The full model consistently achieves the lowest MAE, indicating the most reliable performance. While fixing α or λ can sometimes give a lower RMSE by avoiding occasional large errors coming from a subject in the dataset, the higher MAE shows a decrease in accuracy. A crucial observation, the optimal fixed parameters differ significantly between datasets (best fixed α is 0.6 for PURE but 0.8 for UBFC-rPPG) this verifies the need for an adaptive approach to generalize across different conditions.

Removing the temporal variation penalty shoots up the error rates, confirming the need for penalty in our objective function. Removing the concentration had the most significant impact, confirming its critical role in guiding the optimization toward physiologically relevant frequencies. We hypothesize this is because the concentration term C acts as a frequency-domain constraint that distinguishes genuine pulse signals from noise and artifacts. Without this physiological prior, the temporal variation penalty alone cannot differentiate between smooth but non-physiological signals, such as low-frequency illumination changes, and actual cardiac activity, leading to parameter selections that produce temporally consistent but physiologically meaningless results.

Table 2: Comprehensive ablation study of PRISM's components on the PURE and UBFC-rPPG datasets. We compare the full model against variants with fixed parameters or a modified objective function. The 'Best Fixed' settings show the result for the single best-performing fixed parameter value from our experiments. 'Concentration-only' refers to optimizing without the temporal variation penalty ($k = 0$). 'TV-Only' uses optimization function without C.

Dataset	Configuration	MAE$^{\downarrow}$	RMSE$^{\downarrow}$	Acc.$^{\uparrow}$	Selected λ	Selected α
PURE	Full PRISM (Ours)	**0.77**	1.69	97.3%	Adaptive	Adaptive
	Best Fixed α ($\alpha = 0.6$)	0.82	**1.26**	**97.5%**	0.155	0.6
	Best Fixed λ ($\lambda = 0.5$)	0.99	1.74	96.4%	0.5	0.466
	Concentration-only ($k = 0$)	4.02	5.47	85.8%	0.107	0.171
	TV-only ($C = 0$)	14.30	16.55	64.2%	0.01	0.0
UBFC-rPPG	Full PRISM (Ours)	**0.66**	1.54	97.5%	Adaptive	Adaptive
	Best Fixed α ($\alpha = 0.8$)	0.70	**1.15**	**97.6%**	0.058	0.8
	Best Fixed λ ($\lambda = 0.05$)	1.04	1.94	96.6%	0.05	0.560
	Concentration-only ($k = 0$)	1.83	2.93	93.6%	0.052	0.479
	TV-only ($C = 0$)	4.74	8.33	81.9%	0.01	0.0

B PRISM implementation details

The PRISM algorithm is built as a drop-in replacement for traditional projection-based rPPG pipelines such as POS, but with online parameter adaptation for robustness under varying illumination and motion. It begins with preprocessing and normalization, followed by an online grid-search of the optimal color projection and heart rate estimation.

Each input video is first processed with face detection [11]. In each frame, a rectangle region of interest is defined over the face, and pixel intensities within this region are averaged for each RGB channel, producing the raw signals $R(t)$, $G(t)$, and $B(t)$. To suppress illumination drift while preserving pulsatile components, each channel is normalized by dividing it by a smooth estimate of its baseline. For each $X(t) \in \{R(t), G(t), B(t)\}$, we compute

$$\hat{X}(t) = \frac{X(t)}{\tilde{X}(t)}, \tag{3}$$

where $\tilde{X}(t)$ is a smoothing spline fit to $X(t)$ by minimizing the residual functional [24, 25],

$$\tilde{X} = \arg\min_{h \in \mathcal{H}^2} \left\{ \sum_{k=1}^{K} w_k \left(h(t_k) - X(t_k)\right)^2 + \lambda \int \left(h''(t)\right)^2 dt \right\} . \tag{4}$$

Here, $\mathcal{H}^2$ denotes the Sobolev space of twice-differentiable functions, w_k are frame weights (set uniformly), and $\lambda > 0$ controls the smoothness. Smaller values of λ allow the spline to follow fluctuations more closely, while larger values enforce a smoother baseline.

Once normalized, we construct a pulse signal $s(t)$ as a linear combination of the normalized color channels. In the Plane-Orthogonal-to-Skin (POS) method, the signal is given by

$$s_{\text{POS}}(t) = s_1(t) + \gamma \cdot s_2(t) \ , \tag{5}$$

where $s_1(t) = \hat{G}(t) - \hat{B}(t)$, $s_2(t) = \hat{G}(t) + \hat{B}(t) - 2\hat{R}(t)$, and γ is adaptively set as the ratio of standard deviations $\sigma(s_1)/\sigma(s_2)$. Expanding and regrouping terms, we obtain

$$s_{\text{POS}}(t) = (\gamma + 1)\hat{G}(t) + (\gamma - 1)\hat{B}(t) - 2\gamma\hat{R}(t) \ . \tag{6}$$

We can factor out $(1 + \gamma)$ and introduce $\alpha = (1 - \gamma)/(1 + \gamma)$ to obtain an equivalent form:

$$s_{\text{PRISM}}(t) = \hat{G}(t) - \Big(\alpha \cdot \hat{B}(t) + (1 - \alpha) \cdot \hat{R}(t)\Big) \ , \tag{7}$$

with $\alpha \in [-1, 1]$ to be selected adaptively to optimize for signal quality. The formulation retains interpretability while enabling data-driven projection that adapts to varying spectral and motion conditions.

To effectively extract the pulse signal across varying conditions, the parameters α and λ must be selected to balance noise suppression, spectral fidelity, and robustness to motion and illumination changes. Rather than relying on fixed values, PRISM adaptively tunes these parameters based on the observed signal characteristics. This is achieved by segmenting the input stream into consecutive non-overlapping windows of duration $\delta t = 10$ s, over which signal quality can be locally assessed.

Within each window, the normalized signal $s(t)$ is computed for candidate (λ, α) pairs. A fast Fourier transform (FFT) is performed to extract the dominant frequency to give an estimated heart rate h_i. These estimates form the basis for an objective function that encourages sharp spectral peaks in the physiological band and smooth heart rate evolution over time.

To select the optimal parameters (λ^*, α^*), PRISM defines an objective that balances two criteria: spectral concentration of the pulse signal and temporal stability of the estimated heart rates. The spectral concentration C is defined as the fraction of signal power falling within some physiological band $[f_{\text{min}}, f_{\text{max}}]$:

$$C = \frac{\int_{f_{\text{min}}}^{f_{\text{max}}} |S(f)|^2 \, df}{\int_0^{f_{\text{nyquist}}} |S(f)|^2 \, df} \ , \tag{8}$$

where $S(f)$ is the Fourier transform of $s(t)$ and $f_{\text{nyquist}} = \frac{1}{2} f_{\text{frame}}$ is the Nyquist frequency, equal to half the frame rate. Temporal variation is computed as

$$\text{TV} = \frac{1}{t_n - t_1} \sum_{i=1}^{n-1} |h_{i+1} - h_i| \ , \tag{9}$$

where t_1 and t_n are the midpoints of the first and last windows. The joint objective minimized by PRISM is

$$\mathcal{L}(\lambda, \alpha) = k \cdot \mathrm{TV}(\lambda, \alpha) - C(\lambda, \alpha) \ , \tag{10}$$

with $k = 1/3$, set empirically as a heuristic. This objective function encourages signals with high spectral concentration in the plausible physiological range and smooth trajectories of estimated heart rate over time. In practice, a discrete grid search is performed over $\alpha \in \{0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$ and $\lambda \in \{0.01, 0.05, 0.1, 0.5, 1.0\}$. The parameter pair minimizing $\mathcal{L}$ is selected and used to compute final estimates. Empirically, α values below 0.5 are seldom optimal and are therefore excluded from the search. This is intuitive as a large α corresponds to the primary projection given in POS (i.e., $s_1(t) = \hat{G}(t) - \hat{B}(t)$), meanwhile a small α corresponds closer to the secondary projection (i.e., $s_2(t) = \hat{G}(t) + \hat{B}(t) - 2\hat{R}(t)$).

Occasionally, harmonic frequencies dominate the power spectra of rPPG signals, which can result in predictions that are twice the true heart rate frequency. To alleviate this, two bands are used for the spectral concentration C: a high band $[0.75\,\mathrm{Hz},\ 4.0\,\mathrm{Hz}]$ and a low band $[0.5\,\mathrm{Hz},\ 3.0\,\mathrm{Hz}]$. An rPPG candidate is derived from each band, $s_{\text{PRISM},H}(t)$ and $s_{\text{PRISM},L}(t)$, and the estimated windowed heart-rate sequences are given as: $\{f_H^{(k)}\}_{k=1}^K$ and $\{f_L^{(k)}\}_{k=1}^K$. Let

$$\bar{f}_H \triangleq \frac{1}{K}\sum_{k=1}^{K} f_H^{(k)}, \qquad \bar{f}_L \triangleq \frac{1}{K}\sum_{k=1}^{K} f_L^{(k)}. \tag{11}$$

Denote the high-band estimate as a harmonic of the low-band estimate if $\bar{f}_H \approx 2\,\bar{f}_L$. The final selection is

$$\widehat{s}_{\text{PRISM}}(t) = \begin{cases} s_{\text{PRISM},L}(t), & \text{if } \bar{f}_H \approx 2\,\bar{f}_L, \\ s_{\text{PRISM},H}(t), & \text{otherwise.} \end{cases} \tag{12}$$

Since parameter selection relies only on recent signal windows, PRISM operates in an online manner. After a brief initialization period of roughly one minute, the system converges to stable parameter values and can thereafter update them periodically as conditions change, or maintain them if the environment remains stable. Moreover, inference can begin immediately with predefined α and λ values, which, as shown in Appendix D, already provide competitive performance. Subsequent adaptation then refines these parameters to better match the observed data.

C Per-condition performance on PURE

Table 3: Per-condition performance of PRISM on the PURE dataset. MAE, RMSE, and SD are reported in beats per minute (bpm). R is the Pearson correlation coefficient. Acc. is the percentage of estimates within a ± 5 bpm threshold.

Condition	MAE$^{\downarrow}$	RMSE$^{\downarrow}$	SD$^{\downarrow}$	R$^{\uparrow}$	Acc.$^{\uparrow}$
Steady	0.48	0.72	0.72	0.999	100.00
Talking	1.48	3.00	2.93	0.993	88.89
Slow Translation	0.52	0.78	0.77	0.999	100.00
Fast Translation	1.26	2.50	2.50	0.993	93.75
Small Head Rotation	0.40	0.54	0.54	1.000	100.00
Medium Head Rotation	0.53	0.95	0.95	0.999	100.00

D Ablation: Toolbox Modifications vs. Algorithm Contribution

To isolate PRISM's algorithmic contribution from implementation improvements, we evaluated our method under two configurations:

Standard rPPG-Toolbox Configuration: Using default settings (512 FFT bins, 60-second windows, HaarCascade detection, static tracking), PRISM achieves:

- PURE: MAE = 2.66 bpm
- UBFC-rPPG: MAE = 3.10 bpm

Enhanced Configuration (reported in Table 1): With our modifications (subsection 3.2), PRISM achieves:

- PURE: MAE = 0.77 bpm (71% improvement)
- UBFC-rPPG: MAE = 0.66 bpm (79% improvement)

Importantly, our modifications improve the performance of unsupervised methods while showing no statistical difference on supervised approaches. Table 4 demonstrates the impact across representative methods. For unsupervised methods like POS, the enhanced configuration provides consistent improvements on both datasets (39% improvement on PURE, 75% on UBFC-rPPG). Interestingly, the supervised method TS-CAN shows degraded performance on PURE with the enhanced configuration (23% worse), while improving substantially on UBFC-rPPG (60% improvement). This suggests that supervised methods may have been optimized for the specific characteristics of the standard toolbox settings, whereas unsupervised methods benefit more uniformly from improved signal preprocessing. PRISM's adaptive optimization demonstrates robust performance under both configurations, achieving competitive results with standard settings and state-of-the-art performance with enhanced settings.

Table 4: Performance comparison of leading unsupervised and supervised methods under standard versus enhanced rPPG-Toolbox configurations. Unsupervised methods (US) benefit consistently from enhanced settings, while supervised methods (S) show dataset-dependent responses, suggesting possible overfitting to standard toolbox characteristics.

	Method	**PURE**		**UBFC-rPPG**	
		Standard MAE (bpm)	**Enhanced** MAE (bpm)	**Standard** MAE (bpm)	**Enhanced** MAE (bpm)
US	POS [5]	3.67	2.23	4.00	1.08
US	**PRISM**	**2.66**	**0.77**	**3.10**	**0.66**
S	TS-CAN [21]	3.69	4.54	1.29	0.51