Enhancing Heart Rate Estimation with POS-SSA Remote Photoplethysmography

Ruixuan Wang [1,*], Wei Quan [1], Bogdan Matuszewski [1], Nick A. Heywood [2], Christopher Gaffney [3], and Katie Hoad [3]

[1] School of Engineering and Computing, University of Lancashire, Preston, United Kingdom
[2] East Lancashire Hospitals, NHS Trust, Blackburn, United Kingdom
[3] Lancaster Medical School, Lancaster University, Lancaster, United Kingdom
Email: RWang28@lancashire.ac.uk (R.W.); WQuan@lancashire.ac.uk (W.Q.); BMatuszewski1@lancashire.ac.uk (B.M.); Nick.Heywood@elht.nhs.uk (N.A.H.); C.Gaffney@lancaster.ac.uk (C.G.); K.Hoad@lancaster.ac.uk (K.H.)
*Corresponding author

***Abstract*—Remote Photoplethysmography (rPPG) enables non-contact heart rate monitoring from facial videos but it is highly susceptible to motion artefacts, illumination changes, and sensor noise. A framework combining the Plane Orthogonal-to-Skin (POS) method and Singular Spectrum Analysis (SSA) was proposed in this work to address these challenges by first projecting normalized Red Green Blue (RGB) signals onto a skin-tone–orthogonal subspace to suppress illumination and motion distortions and then decomposing the resulting signal into components that isolate physiologically meaningful oscillations. Evaluation on the PFF and UBFC-Phys dataset demonstrates that this approach consistently outperforms conventional single-channel, statistical, and chrominance-based methods by achieving a mean absolute error of 4.99 beats per minute (bpm) and correlation of 0.76 on PFF, and a mean absolute error of 4.11 bpm with correlation of 0.86 on UBFC-Phys. Furthermore, the comparison with results reported in the existing literature indicates that the proposed framework achieves competitive accuracy relative to popular learning-based rPPG approaches. These findings indicate that integrating chrominance projection with adaptive temporal decomposition significantly improves robustness and accuracy for contact-free heart rate estimation.**

***Keywords*—facial video-based remote Photoplethysmography (rPPG), Plane Orthogonal-to-Skin (POS), Singular Spectrum Analysis (SSA)**

I. Introduction

Heart Rate (HR) is a fundamental physiological parameter that provides critical insight into the functional state of the cardiovascular system, Autonomic Nervous System (ANS) balance, and overall physical health. It reflects the dynamic interaction between sympathetic and parasympathetic regulation [1] and serves as a direct indicator of cardiac activity. Deviations from normal heart rate patterns, including tachycardia, bradycardia, and abnormal variability, are closely associated with a wide range of clinical conditions such as cardiovascular disease, metabolic syndromes, sleep disorders, and neurological dysfunctions [2–4]. Accurate and continuous monitoring of HR enables early detection of clinical abnormalities, supports risk stratification, and guides therapeutic decision-making. In clinical practice, Heart Rate Variability (HRV) analysis is widely employed as a non-invasive biomarker for evaluating autonomic regulation [5] and predicting outcomes in cardiac and stress-related disorders [6]. Beyond traditional clinical settings, HR analysis also plays a central role in personalised health management, including fitness assessment, emotion recognition, fatigue detection, and stress monitoring, aligning with the emerging paradigm of preventive and precision medicine [7].

HR measurement techniques are broadly divided into contact-based approaches and non-contact-based approaches. Contact-based methods mainly consist of Electrocardiography (ECG), Photoplethysmography (PPG), and arterial pressure monitoring. They rely on direct interaction with the skin to capture physiological signals. ECG remains the gold standard for clinical cardiac assessment, providing highly accurate detection of electrical activity and heartbeat intervals [8]. Similarly, PPG is implemented in devices such as smartwatches, fitness trackers, and pulse oximeters, which measures blood volume changes using optical sensors to estimate heart rate and related parameters [9]. While these methods deliver precise and reliable results, they have inherent drawbacks such as motion artifacts, skin irritation, and discomfort during prolonged use. Their dependence on physical contact makes them less suitable for continuous or long-term monitoring in sensitive individuals or in environments where minimal intrusion is required [10].

In contrast, non-contact-based methods have emerged as a promising alternative that enables unobtrusive heart rate monitoring without physical sensors. Remote Photoplethysmography (rPPG) is the most widely used method, which captures subtle colour fluctuations in the skin caused by periodic blood volume changes using conventional RGB or Near-Infrared (NIR) cameras [11]. By applying advanced computer vision, signal processing,

Manuscript received November 1, 2025; revised December 4, 2025; accepted January 27, 2026; published April 24, 2026.

and deep learning algorithms, rPPG extracts cardiac-related signals from facial or exposed skin regions with rich vascularization. Recent advancements in spatial-temporal modelling, transformer networks, and self-supervised learning have enhanced its robustness under varying illumination, motion, and skin tone conditions. Beyond heart rate estimation, rPPG has been extended to stress and emotion analysis, fatigue detection, telemedicine, and remote health monitoring. Its comfort, scalability, and potential for real-time implementation on mobile and edge devices make it a highly attractive solution for continuous, contact-free physiological monitoring in both healthcare and everyday applications [12].

Over the last decade, deep learning–based rPPG approaches have demonstrated notable improvements in robustness under motion and illumination variations. Methods such as PhysNet [13], EfficientPhys [14], and transformer-based temporal models [15] leverage spatial–temporal representations to directly learn physiological patterns from facial video sequences. Recent development of TranSpike [16] further improve robustness by introducing pixel-wise frequency reconstruction to preserve pulsatile extrema and by modelling spike interactions across facial regions, thereby enhancing rPPG signal fidelity under challenging illumination and motion conditions. While these data-driven models often achieve strong performance in unconstrained environments, they typically require large-scale annotated datasets and incur higher computational costs and limited interpretability. These can restrict their practical deployment in resource-constrained or clinically sensitive settings.

In this context, signal-based frameworks remain attractive for applications where transparency, computational efficiency, and ease of deployment are critical. This highlights the need for approaches that improve robustness without relying on extensive training data or complex model architectures. The proposed POS–SSA framework combines chrominance-based projection with adaptive temporal decomposition to enhance physiological signal extraction while maintaining interpretability and low data dependency. It offers a practical and robust alternative for contact-free heart rate monitoring.

II. Literature Review

Early research on rPPG primarily focused on demonstrating its feasibility under controlled laboratory conditions. Researchers explored the relationships between light reflection, skin tone, and blood perfusion to identify suitable colour channels and spatial regions for signal extraction. Despite encouraging initial results, rPPG signals were found to be highly sensitive to external influences such as lighting changes, head motion, and camera sensor noise, which introduced instability and reduced measurement accuracy [17]. Consequently, improving signal quality and robustness became a central focus of subsequent studies. Over time, the field evolved from basic colour-channel analysis to more advanced algorithms that combined statistical, physiological, and computational principles to separate pulsatile information from non-physiological sources of variation.

Over the years, researchers have proposed a variety of methods to extract rPPG signals with increasing levels of sophistication. Early approaches primarily focused on analysing pixel intensity variations in visible-light channels to recover pulsatile information from facial videos. For instance, early studies demonstrated that the green channel carries the strongest pulsatile component due to its high sensitivity to blood volume changes, while the combination of red and green signals could further enhance the periodic component of the waveform [18]. As the field progressed, statistical signal processing methods such as Principal Component Analysis (PCA) were introduced to exploit inter-channel correlations and isolate physiological components from background noise [19]. PCA-based methods effectively reduced illumination bias and motion-related artefacts by projecting multi-channel signals onto orthogonal bases, allowing more reliable estimation of the underlying cardiac rhythm.

Building upon these foundations, more sophisticated models were developed to address illumination and motion interferences more explicitly. The Chrominance-Based rPPG (CHROM) algorithm introduced a colour-space transformation that computes a weighted combination of chrominance signals, thereby suppressing intensity-related fluctuations and improving the signal-to-noise ratio under varying lighting conditions [20]. Similarly, the Plane Orthogonal to Skin (POS) algorithm projects normalised RGB signals onto a plane orthogonal to the skin-tone vector, reducing the influence of both global illumination variations and minor motion disturbances [21]. These advancements improved the robustness and reliability of rPPG signal extraction compared to earlier methods, forming the basis for many subsequent developments in the field.

Despite these notable improvements, conventional algorithms such as CHROM and POS still exhibit performance degradation in real-world scenarios, particularly under conditions of substantial motion, varying ambient illumination, camera noise, or diverse subject appearances [22]. Such interferences can distort the extracted waveform and lead to inaccurate heart-rate estimation, limiting their applicability in unconstrained environments. To overcome these challenges, recent studies have explored deep learning techniques that leverage convolutional and recurrent architectures to model complex spatial–temporal relationships within video data. These models have demonstrated improved robustness and generalization across conditions by learning motion and illumination-invariant representations directly from data. However, their reliance on large annotated datasets and the inherent lack of interpretability remain practical barriers to widespread adoption in medical and consumer-grade systems.

To address these limitations, the present study proposes a hybrid rPPG extraction framework that integrates chrominance-based projection with data-adaptive decomposition. In the first stage, the POS algorithm is employed to project normalised RGB signals onto a

subspace orthogonal to the skin-tone vector, which effectively mitigates common-mode distortions arising from illumination variation, sensor bias, and motion-induced leakage. This process yields a stable, illumination-invariant rPPG signal that preserves the essential pulsatile dynamics. In the second stage, Singular Spectrum Analysis (SSA) [23] is applied to decompose the POS-derived signal into a set of orthogonal components using Hankel embedding and singular value decomposition. Components exhibiting quasi-periodic behaviour within the physiological heart-rate frequency band are retained, while noise-dominated components are suppressed. The reconstructed waveform is therefore denoised, physiologically meaningful, and suitable for robust heart-rate estimation in practical, unconstrained environments.

III. POS-SSA Methodology

The processing pipeline of the proposed POS–SSA framework is shown in Fig. 1, which comprises four steps. First, the full face is detected from each video frame and used as the Region of Interest (ROI), providing a localised skin area for subsequent signal acquisition. Next, the averaged RGB signals are transformed by the POS algorithm, which projects them onto a chrominance subspace orthogonal to the skin tone vector and yields composite cardio signal more robust to illumination changes. This signal is then refined through the SSA technique, where decomposition into elementary components separates pulsatile information from noise and motion artefacts. Finally, the components with quasi-periodic behaviour and spectral energy in the physiological heart-rate band are analysed in the frequency domain via Fourier transform in order to estimate heart rate. Collectively, these stages form a compact yet robust framework for enhancing signal quality and improving estimation accuracy under varied conditions.

A. Facial Region Detection and Normalisation

Facial detection and normalisation ensured that subsequent rPPG analysis was based on a spatially consistent and geometrically standardised Region of Interest (ROI). Faces were localized in each video frame using the Viola–Jones cascade classifier [24], which employs Haar-like features and AdaBoost-trained classifiers for robust detection under varying lighting and pose conditions. Within the detected ROI, ten salient feature points were extracted using the minimum eigenvalue method [25] and tracked across frames with a bidirectional error–minimizing algorithm to maintain spatial and temporal consistency. The similarity transformation was estimated from the tracked features, and it was then applied to correct translation, in-plane rotation, and scale variations. Finally, the ROI was resampled to a fixed resolution and orientation, yielding a normalised and temporally stable facial segment suitable for reliable rPPG signal extraction.

The full-face region was selected as the region of interest rather than smaller subregions (e.g., cheeks or forehead) to maximise signal stability across subjects and recording conditions. Prior studies have demonstrated that using a larger facial ROI reduces sensitivity to local motion, partial occlusion, and regional illumination non-uniformity [26]. Averaging over the full face also improves the signal-to-noise ratio by aggregating pulsatile information from multiple vascularised areas [27], which is particularly beneficial for subsequent SSA-based decomposition.

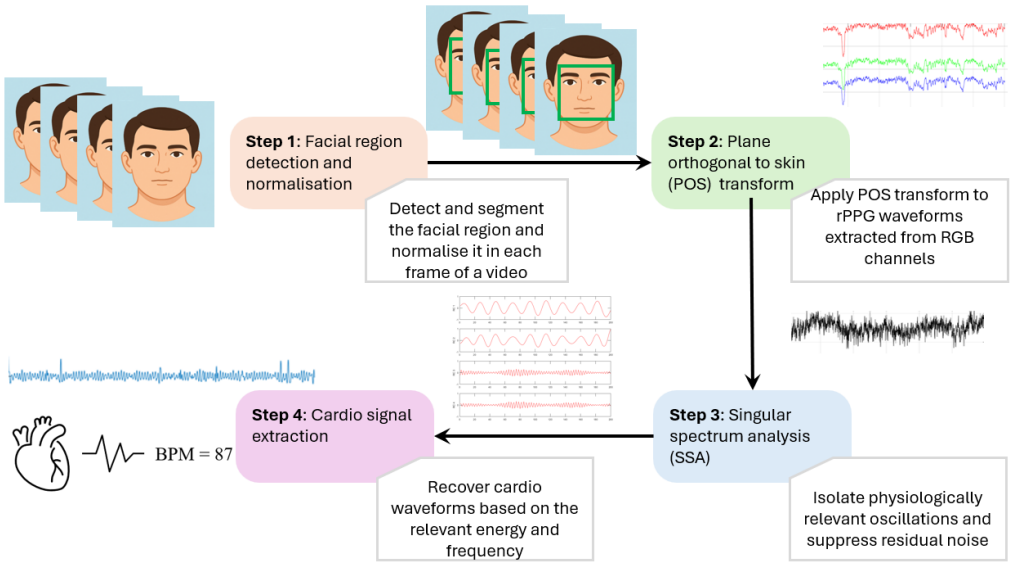

Fig. 1. Pipeline of the proposed POS-SSA framework.

B. POS Transform

The POS method was first proposed by Wang *et al*. [21], which is a widely used method for extracting remote Photoplethysmography (rPPG) signals. It is designed to enhance the physiological component of skin-tone variations in video while suppressing noise from motion and illumination changes, thereby enabling more reliable heart rate estimation in real-world conditions. The method first preprocesses the colour signals to remove slow

illumination drifts and then projects them onto an orthogonal subspace in the RGB space to emphasise the pulsatile component. Finally, it recovers clean physiological signals via dynamic normalisation and combination of the two orthogonal projection signals.

Given a video sequence, the facial region is first detected and localised. For each frame, the mean pixel intensities of the red, green, and blue channels are extracted and denoted as $R(t), G(t), B(t)$, respectively. To reduce the effects of illumination fluctuations and global intensity changes, these raw colour signals are detrended by removing their slowly varying components. This is achieved by applying a moving average filter for each channel. The resulting detrended signals are therefore defined as deviations from their global mean values, where $\bar{R}(t), \bar{G}(t)$ and $\bar{B}(t)$ represent the temporal mean of the red, green, and blue channels, respectively.

Based on the distribution characteristics of skin colour in the RGB colour space, the POS algorithm constructs two signal components, defined as follows:

$$X(t) = R'(t) - G'(t) \tag{1}$$

$$Y(t) = R'(t) + G'(t) - 2B'(t) \tag{2}$$

$X(t)$ and $Y(t)$ emphasize pulsatile information while reducing illumination and motion-induced noise. To avoid domination of one component due to amplitude imbalance, their standard deviations σ_X and σ_Y over a temporal window are computed and used for normalisation:

$$\sigma_X = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(X(t) - \bar{X})^2} \tag{3}$$

$$\sigma_Y = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(Y(t) - \bar{Y})^2} \tag{4}$$

$\bar{X}$ and $\bar{Y}$ are the mean values of $X(t)$ and $Y(t)$ over the window 1 to N.

The second component is normalised by its standard deviation and subtracted from the first component to form a synthesised POS signal $S(t)$, which can be expressed as:

$$S(t) = X(t) - \frac{\sigma_X}{\sigma_Y}Y(t) \tag{5}$$

Eqs. (1) and (2) define two orthogonal chrominance signals derived from the normalised RGB channels, designed to suppress common-mode intensity variations. Eqs. (3) and (4) perform adaptive normalisation using the standard deviation within a temporal window to prevent dominance by any single channel. Eq. (5) combines the two normalised components into a single POS signal that emphasises pulsatile variations while attenuating motion and illumination artefacts.

Fig. 2 illustrates the raw RGB channel signals and the synthesised POS signal obtained after applying the plane orthogonal-to-skin projection. As shown in the figure, the raw colour signals (red, green, and blue) contain both the desired pulsatile component and substantial noise arising from illumination fluctuations and minor head movements. After the POS transformation, these non-physiological variations are reduced, yielding a more stable and periodic waveform that aligns closely with the underlying cardiac rhythm. The improvement demonstrates the effectiveness of the POS method in enhancing signal quality by suppressing motion-induced and lighting-related distortions while preserving the essential heart rate information. Consequently, the synthesized POS signal serves as a cleaner and more robust input for subsequent SSA-based decomposition, which decomposes $S(t)$ into oscillatory components and selectively reconstructs the cardiac-related component.

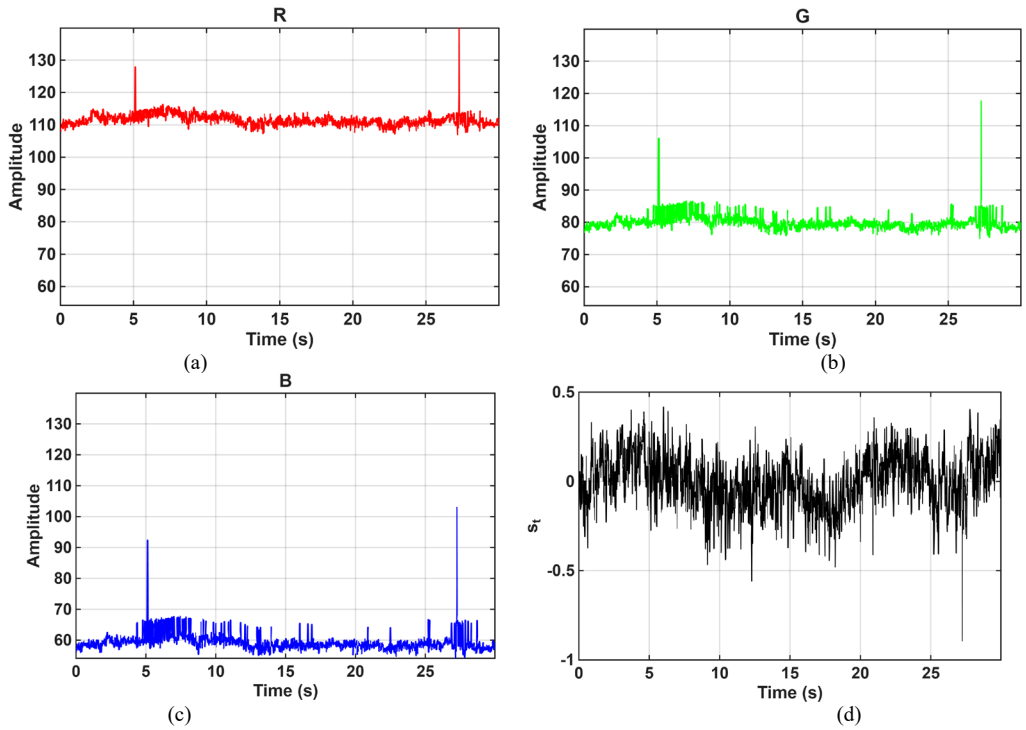

Fig. 2. Examples of colour rPPG signals extracted from the facial regions and the corresponding POS-derived rPPG signal: (a) red-channel rPPG; (b) green-channel rPPG; (c) blue-channel rPPG; (d) POS-derived rPPG.

C. Singular Spectrum Analysis

After synthesising the POS signal, SSA is applied to further isolate physiologically relevant oscillations and suppress residual noise. SSA is a non-parametric decomposition method commonly used for denoising, trend removal, and periodic component extraction in time series [23]. Since remote rPPG signals are often affected by illumination changes and measurement artefacts, SSA provides an effective temporal filtering strategy that enhances heart rate components, even under non-ideal acquisition conditions.

Let the original time series be defined as $S(t)$. For subsequent analysis, this signal is uniformly sampled at the video frame rate f_s, yielding a discrete sequence

$$S = \{s_1, s_2, \dots, s_N\},\ s_n = S(n\Delta t),\ \Delta t = \frac{1}{f_s} \tag{6}$$

For a chosen window length L such that $1 < L < N$, the series is embedded into a sequence of lagged vectors to form a trajectory matrix X:

$$X = \begin{bmatrix} s_1 & s_2 & \cdots & s_K \\ s_2 & s_3 & \cdots & s_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_L & s_{L+1} & \cdots & s_N \end{bmatrix} \in R^{L\times K} \tag{7}$$

where $K = N - L + 1$. This Hankel matrix captures the temporal dynamics of the series. By decomposing the trajectory matrix via Singular Value Decomposition (SVD), the original signal is separated into a set of elementary components. These components can be grouped into interpretable categories:

- Trend components: capturing slowly varying illumination or baseline drift.
- Oscillatory components: representing periodic physiological rhythms, including the heart rate–related signal.
- Noise components: which mainly correspond to high-frequency fluctuations caused by sensor noise or subtle motion artefacts.

By reconstructing the signal using only the oscillatory components within the cardiac frequency band, SSA effectively attenuates noise and irrelevant variations while enhancing the pulsatile component. Each video segment is 30 seconds in duration and captured at a frame rate of 50 fps, yielding a total of 1500 frames per segment ($N = 1500$). In principle, the choice of the window length L is a temporal parameter that depends on the desired resolution and stability of the decomposition. In practice, however, the feasibility of selecting large L values is also constrained by implementation factors, such as the spatial resolution of the input video frames and the computational capacity of the hardware [28]. High-resolution facial recordings (e.g., 1080 p or higher) increase the data volume per frame, which in turn amplifies the computational burden during trajectory matrix construction and decomposition. Likewise, devices with limited CPU/GPU resources or restricted memory bandwidth may encounter significant slowdowns when large values of L are used, particularly in scenarios involving multiple segments.

In this study, the SSA window length L was set to 20 after empirical evaluation and practical considerations. Preliminary experiments were conducted using different window lengths to assess their impact on signal decomposition quality and computational efficiency. It was observed that smaller values of L were insufficient to capture the temporal structure of cardiac oscillations, whereas larger values led to substantially increased computational cost due to the growth of the trajectory matrix without providing additional performance gains. Given a sampling rate of 50 fps, $L = 20$ provides a balance between temporal resolution, numerical stability, and computational feasibility. This choice ensures effective separation of oscillatory cardiac components while remaining suitable for near real-time implementation on standard hardware.

D. Cardio Signal Extraction

Each SSA-processed component represents a distinct temporal mode of variation within the original signal. To isolate physiologically relevant structures from noise or motion-induced artifacts, a two-step component selection strategy was employed:

- **Energy Analysis**: To quantify the contribution of each SSA component, the squared singular value of each component was divided by the sum of all squared singular values, which is defined as the normalised energy ratio. It reflects the proportion of total signal energy represented by each component, and components with higher ratios were retained as candidate signals.
- **Frequency-Domain Analysis**: The retained components were examined using Fast Fourier Transform (FFT). Only those with dominant peaks within the typical heart rate band (0.5–4 Hz) were considered valid physiological components

Fig. 3 illustrates the energy distribution across the first ten SSA components while Fig. 4 presents the corresponding time-domain waveforms of the six components with the highest energy contributions. In Fig. 3, it can be observed that the first few components capture the majority of the total signal energy, which indicates that they contain dominant physiological information associated with cardiac activity. The rapid decline in energy after the fourth component reflects the diminishing influence of meaningful oscillatory structures and the growing dominance of noise or residual illumination effects in the lower-ranked components. Complementing this, Fig. 4 provides a visual comparison of the temporal behaviour of these components. The higher-energy components (particularly components 2 to 4) exhibit clear quasi-periodic oscillations consistent with heart rate–related dynamics, whereas the lower-energy components appear more irregular with erratic fluctuations and less defined rhythmic patterns. This contrast highlights the effectiveness of SSA in decomposing the POS signal into interpretable components, where only a limited subset contributes physiologically relevant information. Based on both the quantitative energy distribution and qualitative waveform inspection, the first four components (each contributing at least 1% of the total energy) were retained

for subsequent spectral analysis and ensure that the reconstructed signal maintains strong physiological fidelity while minimizing the inclusion of noise or motion artefacts.

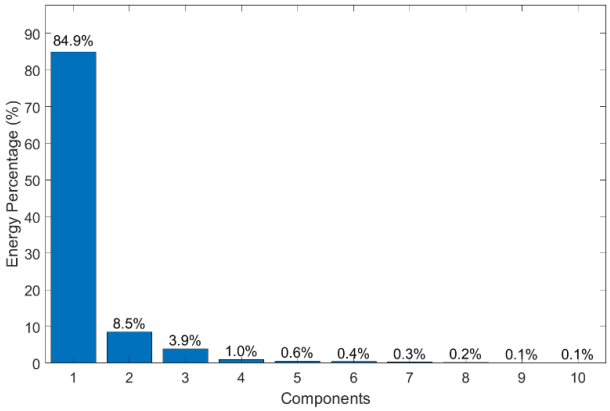

Fig. 3. Energy distribution of top 10 components.

FFT was then applied to each of the retained components. The spectral content of first 4 SSA calculated components RC1 to RC4 was inspected to identify dominant peaks corresponding to cardiac activity as shown in Fig 5. While the first component RC1 exhibited the highest energy, its peak frequency was near 0 Hz, which suggests it represented a baseline trend rather than physiological variation. Therefore, it was excluded. Among the remaining components, only the second component RC2 displayed a clear peak within the 0.5–4 Hz band, aligning with expected heart rate frequencies, and was thus identified and kept as the most physiologically relevant for reconstructing the rPPG signal. Through this integrated approach of POS-filtered signals followed by energy-based ranking and frequency-domain validation, physiologically meaningful information was extracted with enhanced robustness against noise and improving the fidelity of heart rate estimation from facial video data.

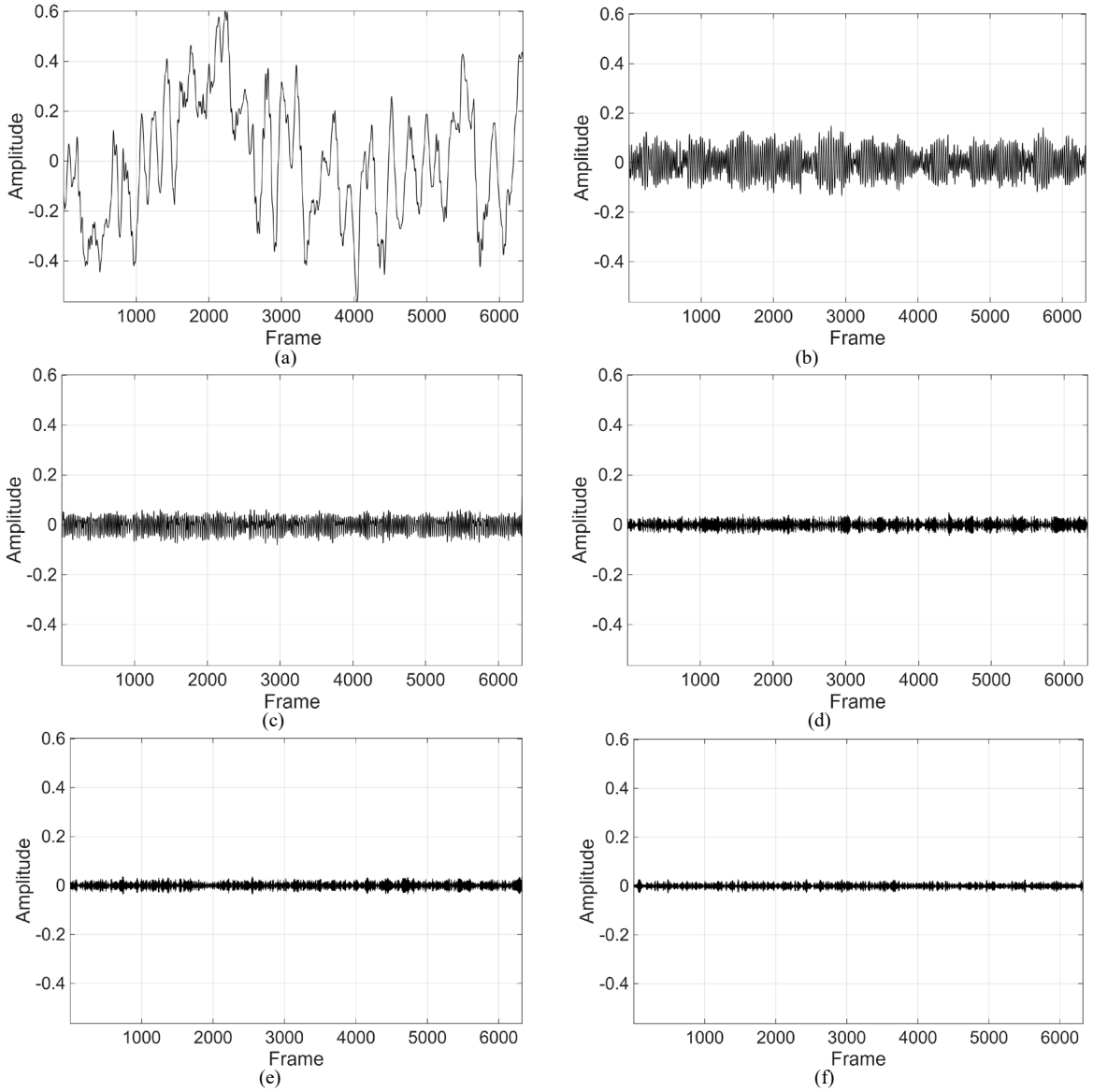

Fig. 4. Time-domain representation of the first six Reconstructed Components (RCs) obtained by applying SSA to the POS-derived rPPG signal: (a) RC1; (b) RC2; (c) RC3; (d) RC4; (e) RC5; (f) RC6.

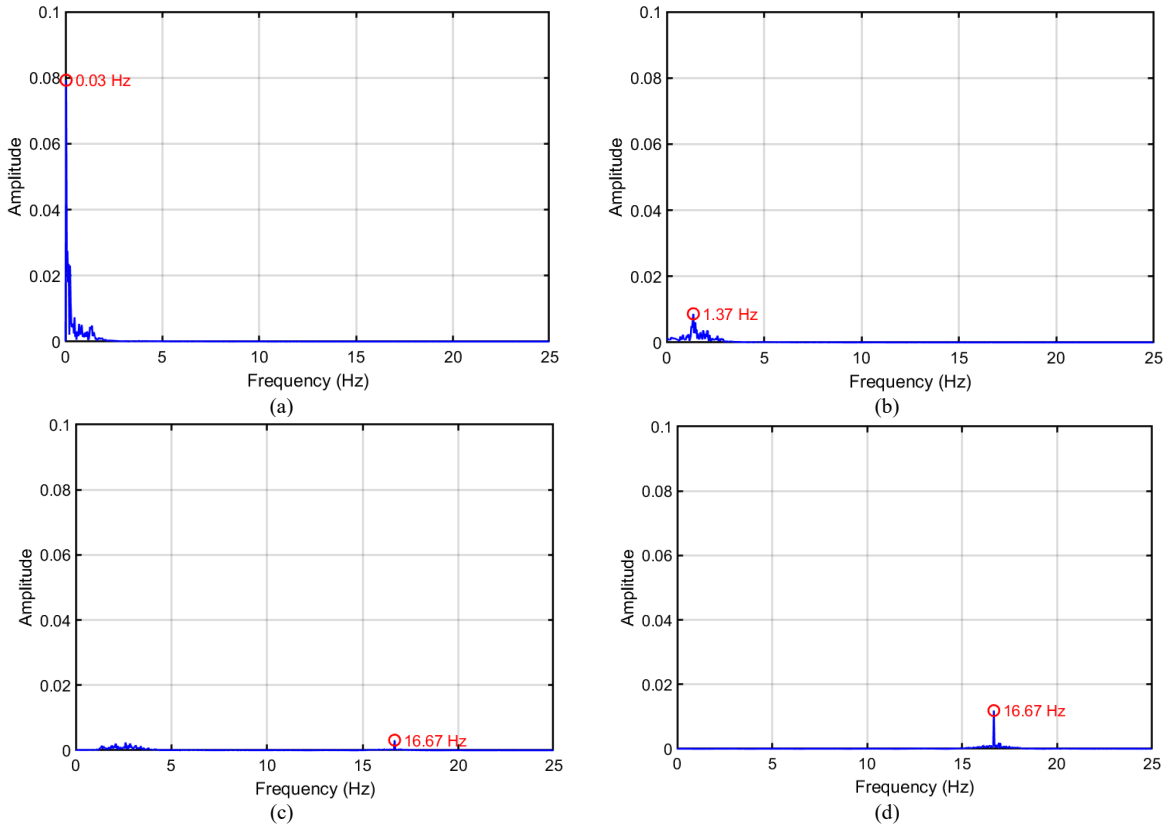

Fig. 5. Spectrum analysis for first 4 SSA components: (a) RC1; (b) RC2; (c) RC3; (d) RC4.

In summary, starting from the POS-processed facial signal, the application of SSA enabled a structured decomposition into interpretable components. The subsequent combination of energy-based ranking and frequency-domain analysis provided a principled approach to isolate physiologically meaningful information. This integrated strategy enhances robustness against noise and improves the fidelity of heart rate estimation from facial video data.

IV. Experiments

A. Experimental Datasets

To rigorously evaluate the effectiveness of the proposed POS-SSA framework for remote heart rate estimation, the experiments were conducted on the PFF dataset [29] and the UBFC-Phys dataset [30]. Both datasets provide complementary recording conditions and subject variability, which enable a comprehensive assessment.

The PFF dataset contains facial video recordings from 13 participants under varying illumination conditions with synchronised ground-truth heart rate data. The dataset was collected with informed consent under institutional ethical approval, as described in [31]. The access was granted by the authors for research purposes. Each recording lasts approximately three minutes, during which both physiological fluctuations in heart rate and natural variations in illumination and head motion occur. Whereas the UBFC-Phys consists of facial videos from 56 participants recorded under controlled indoor conditions. It is widely used as a benchmark for rPPG evaluation and provides higher inter-subject diversity in skin tone and facial appearance, allowing validation of cross-dataset robustness.

For consistency across both datasets, all recordings were partitioned into non-overlapping 30-second clips recorded at 50 fps. Each segment was processed independently using the proposed POS–SSA framework, and the estimated heart rates were compared against the reference values to enable a comprehensive quantitative assessment of performance.

B. Performance Analysis

In this study, all experiments were conducted on a desktop PC equipped with an Intel Core Ultra 5 245KF CPU operating at 4.20 GHz, 48 GB of RAM, and without GPU acceleration. The average processing time for a single 30-second segment was approximately 40 s. This indicates that the current MATLAB-based implementation is suitable for offline analysis but does not yet achieve real-time performance. Runtime profiling shows that the primary computational bottleneck arises from the SSA stage, particularly during the construction of the trajectory (Hankel) matrix and the subsequent singular value decomposition.

Fig. 6 presents a detailed comparison of the predicted and ground-truth heart rate signals for eight representatives, including four from PFF dataset and four from UBFC-Phys dataset. The figure illustrates the temporal tracking capability and overall accuracy of the proposed POS-SSA framework. The results clearly demonstrate that the predicted heart rate trajectories closely follow the reference measurements throughout the

recording duration. Across all examples, the estimated heart rates exhibit smooth transitions and synchronized temporal dynamics, which confirms the framework's ability to capture both the baseline cardiac rhythm and transient fluctuations induced by physiological or environmental factors. Importantly, deviations between the estimated and ground-truth signals remain consistently low, typically within ±6 bpm, which reflects the method's robustness in mitigating common challenges such as illumination variation, minor head motion, and camera noise.

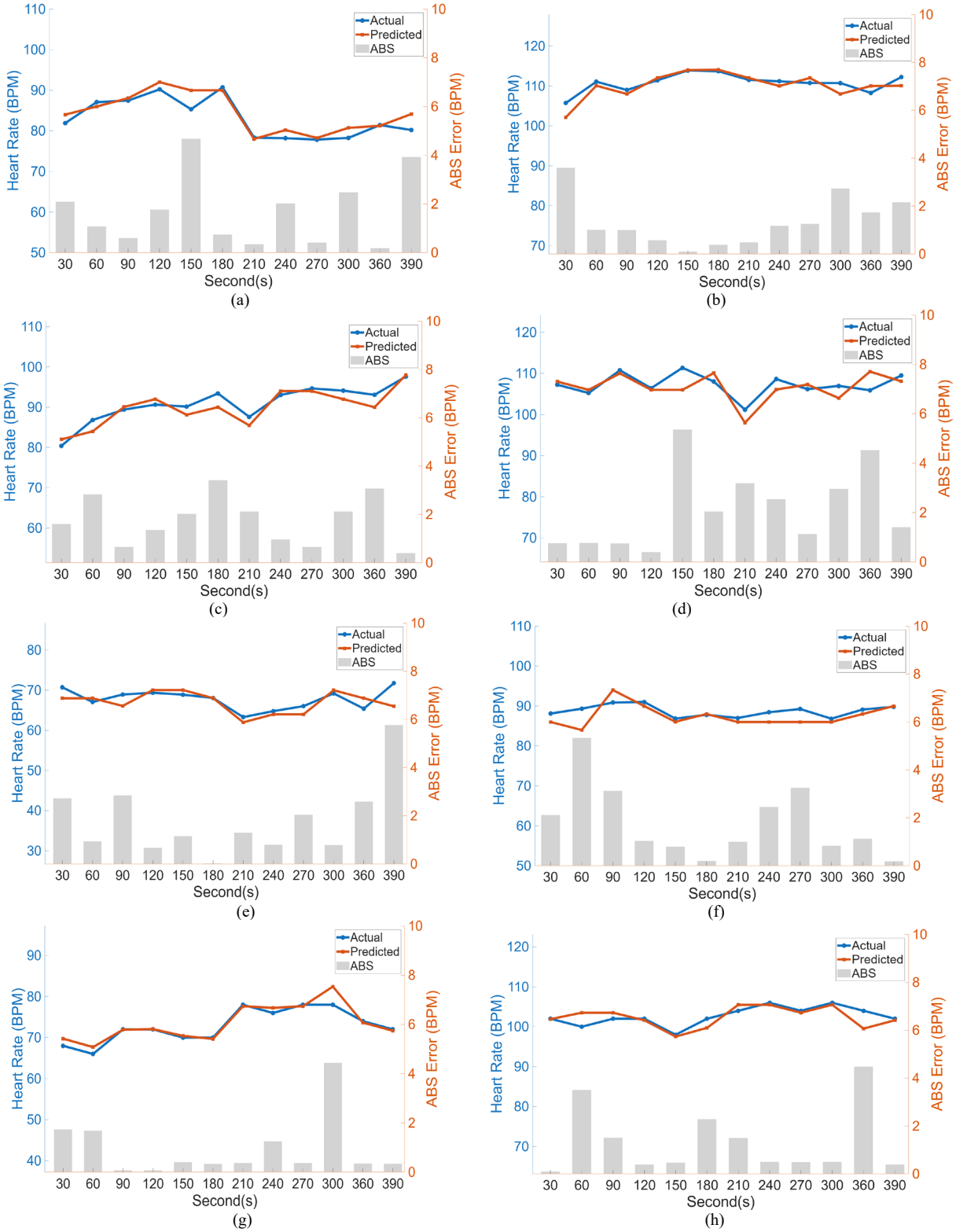

Fig. 6. Representative examples of heart rate estimation using the proposed framework on the PFF and UBFC-Phys datasets: (a) P1-PFF (b) P2-PFF (c) P3-PFF (d) P4-PFF; (e) P5-UBFC-Phys (f) P6-UFBC-Phys (g) P7-UFBC-Phys (h) P8-UFBC-Phys.

A closer inspection of the individual subplots in the figure reveals subtle distinctions in tracking behaviour among participants, which is attributable to differences in facial features, skin tone, and motion intensity. For example, participants P2 and P3 demonstrate nearly perfect alignment between predicted and ground-truth curves that indicates highly stable performance under uniform lighting conditions. In contrast, participants P1, P4 and P7 show slightly larger fluctuations during sections of rapid heart rate change, which suggests the presence of transient noise or reduced signal-to-noise ratio due to subtle motion or lighting transitions. Nevertheless, even in these cases, the POS-SSA output quickly realigns with the true signal, illustrating its capacity for self-correction through adaptive decomposition and frequency-domain validation. The narrow error margins and consistent waveform overlap across all examples further confirm the system's temporal coherence and noise resilience.

Moreover, the Absolute (ABS) error bars displayed alongside the trajectories provide quantitative evidence of the algorithm's precision. The limited amplitude of these error bars that were concentrated within a narrow range across the full duration of each recording emphasises the framework's stability and reliability for long-term monitoring. This temporal consistency is critical for practical applications such as fatigue assessment, stress detection, or telemedicine, where accurate trend tracking is more valuable than instantaneous measurements. Unlike conventional rPPG methods that often suffer from cumulative drift or delayed response under varying conditions, the integration of SSA into the POS pipeline ensures continuous refinement of the extracted signal by isolating physiologically meaningful oscillatory modes and suppressing residual noise.

C. Benchmarking

To contextualise the performance of the proposed POS-SSA framework, its results were compared against several well-established algorithms spanning the spectrum from simple single-channel extraction to statistically informed separation and chrominance-based enhancement. All comparative methods were implemented in this work and include the Green (G) channel only method [10], the Green plus Red (G+R) channels chrominance-based method [18], the blind source separation based on PCA [19], the POS based on RGB signals [21] and the CHROME [22].

Performance was quantitatively assessed on both the PFF and UBFC-Phys datasets using five standard metrics: Mean Absolute Error (MAE), root mean square error (RMSE), Mean Error (ME) to quantify bias, Standard Deviation of error (STD) to reflect variability, and the Pearson correlation coefficient (r) to measure agreement with ground-truth heart rate signals. The results are summarised in Table I.

TABLE I. COMPARISON OF RPPG METHODS

Methods	PFF Dataset						UBFC-Phys Dataset					
	G	G+R	PCA	POS	CHROM	POS+SSA	G	G+R	PCA	POS	CHROM	POS+SSA
MAE (bpm)	22.47	24.19	23.03	9.48	8.51	4.99	17.28	9.93	21.30	6.00	6.79	4.11
RMSE (bpm)	30.99	32.88	31.07	22.28	20.53	13.05	28.21	19.86	30.55	11.99	14.45	6.47
ME (bpm)	8.56	6.34	15.47	1.35	1.84	1.16	1.30	2.56	3.51	1.29	2.31	1.29
STD (bpm)	29.74	32.26	26.94	22.23	20.46	13.01	28.25	19.74	30.43	11.95	14.40	6.18
Pearson r	0.12	0.15	0.18	0.43	0.51	0.76	0.10	0.28	0.15	0.59	0.55	0.86

Across both datasets, the proposed POS–SSA framework consistently outperforms all comparative methods in terms of accuracy, bias reduction, and correlation. On the PFF dataset, the POS–SSA achieves the lowest MAE of 4.99 bpm and RMSE of 13.05 bpm. It substantially improves upon the POS (MAE 9.48 bpm, RMSE 22.28 bpm) and CHROM (MAE 8.51 bpm, RMSE 20.53 bpm) methods. The mean error is reduced to 1.16 bpm, which indicates minimal systematic bias, while the Pearson correlation coefficient reaches 0.76, which is the highest among all evaluated approaches. These results demonstrate that the integration of SSA significantly enhances the robustness of POS by suppressing residual noise and non-physiological variations.

Similar trends are observed on the UBFC-Phys dataset, where the POS–SSA achieves an MAE of 4.11 bpm and an RMSE of 6.30 bpm. These outperform POS and CHROM by a clear margin. Notably, the Pearson correlation coefficient increases to 0.86, which shows strong agreement with the ground-truth measurements and confirming the robustness of the proposed method across datasets with different subject populations and acquisition conditions.

The table further reveals broader performance patterns among the compared methods. The G and G+R baselines, which rely on direct FFT analysis of raw colour signals, perform poorly on both datasets, exhibiting large errors and weak correlations ($r < 0.3$). PCA provides moderate improvements by exploiting inter-channel correlations but remains sensitive to motion and illumination changes. Chrominance-based approaches such as POS and CHROM yield substantial gains, which highlights the importance of colour-space projection for noise suppression. However, the proposed POS–SSA framework consistently achieves the lowest error variability across both datasets with a standard deviation (STD) of 13.01 bpm on the PFF dataset and 6.18 bpm on the UBFC-Phys dataset. This demonstrates superior stability and cross-subject consistency.

Fig. 7 presents a scatter plot comparing predicted and actual heart rates across all methods for the PFF and UBFC-Phys datasets. Each point corresponds to one test clip, with the solid line indicating the ideal one-to-one relationship. For the PFF datasets shown in Fig. 7(a), the distribution illustrates that the G, G+R, and PCA methods produce wide scatter with many points deviating

substantially from the line. It reflects large estimation errors and weak correlations. The POS and CHROME approaches have reduced the spread but with noticeable deviations, particularly at higher heart rates. In contrast, the POS-SSA results (red points) cluster tightly around the ideal line and demonstrate both higher accuracy and greater consistency across the full physiological range. This visual evidence complements the numerical improvements reported in Table I in which the reduced scatter confirms the lower error variability, while the alignment with the one-to-one line illustrates the stronger correlation ($r = 0.76$).

A similar trend is observed for the UBFC-Phys dataset in Fig. 7(b). The baseline methods again display broad scatter and poor alignment with the ideal line, whereas POS and CHROM provide moderate improvements. The POS–SSA framework achieves the tightest clustering and strongest alignment with the diagonal, reflecting its superior accuracy and robustness under controlled indoor conditions with higher inter-subject variability. The close agreement observed across both datasets visually confirms the numerical improvements reported in Table I and demonstrates that the proposed POS–SSA framework delivers accurate, stable, and generalisable heart rate estimates across diverse recording conditions and subject populations.

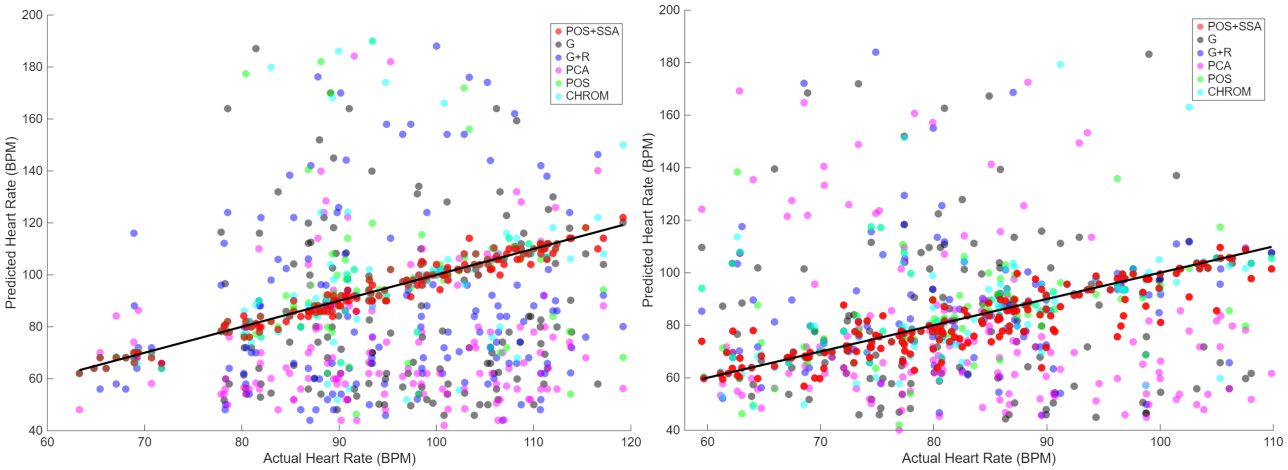

Fig. 7. Scatter plot of predicted against actual heart rates: (a) PFF dataset; (b) UBFC-Phys dataset.

V. Discussion

A. Robustness of Proposed POS-SSA Framework

rPPG remains inherently challenging due to its sensitivity to motion artefacts, illumination variability, partial occlusions, and camera noise. The proposed POS–SSA framework mitigates these issues by combining colour-space projection with adaptive temporal decomposition. The POS stage attenuates illumination and minor motion effects by projecting normalised RGB signals onto a plane orthogonal to the skin-tone vector, yielding a more stable composite signal than raw colour channels. SSA then exploits the temporal structure of this signal by decomposing it into orthogonal components with distinct dynamical characteristics. By retaining only quasi-periodic components whose spectral energy lies within the physiological heart-rate band, SSA suppresses baseline drift, residual motion artefacts, and stochastic noise that persist after chrominance projection. This two-stage strategy improves robustness over using POS alone, as confirmed by the consistent reductions in error metrics and increases in correlation observed across both the PFF and UBFC-Phys datasets. By combining the POS colour-space projection with SSA for adaptive signal decomposition, the method could effectively mitigate common distortions caused by motion artifacts, illumination fluctuations, and camera noise. Experimental results on both datasets demonstrated that the proposed framework outperforms conventional single-channel, statistical, and chrominance-based approaches.

B. Learning-Based rPPG Methods

Recent deep learning–based rPPG methods have demonstrated strong performance by learning complex spatial–temporal representations directly from facial videos. These methods include PhysNet [13], EfficientPhys [14], PhysFormer [32], DeepPhys [33] and TS-CAN [34] that can adaptively weight facial regions, capture non-linear dependencies, and suppress local artefacts through data-driven optimisation. In particular, learning-based approaches are well suited to handling heterogeneous motion patterns and spatially localised disturbances.

In contrast, the proposed POS–SSA framework follows a signal-driven, region-wise strategy that prioritises robustness, transparency, and low data dependency. While it does not learn spatial attention, it avoids reliance on large annotated datasets and reduces the risk of domain overfitting. The POS–SSA offers lower computational complexity and predictable behaviour, which make it attractive for applications where training data are limited or explainability is required. Liu *et al.* [35] developed a comprehensive toolbox, rPPG-Toolbox, which integrates a wide range of widely used rPPG models and provides

unified support for public benchmark datasets, data augmentation strategies, and standardised performance evaluation. Table II summarises the reported performance of these methods testing on the UBFC-Phys dataset in terms of MAE and Pearson correlation coefficient as documented in their study. It provides valuable contextual insight by showing that the proposed POS-SSA framework achieves error levels, which are broadly comparable to those state-of-the-art learning-based approaches despite its signal-driven nature and lack of data-dependent training.

TABLE II. EVALUATION OF LEARNING-BASED RPPG METHODS ON THE UBFC-PHYS DATASET REPORTED BY LIU *ET AL*. [35]

Methods	TS-CAN	PhysNet	PhysFormer	DeepPhys	EfficientPhys
MAE (bpm)	5.13	5.79	6.63	6.62	4.93
Pearson *r*	0.76	0.70	0.69	0.66	0.79

However, in highly unconstrained scenarios involving severe motion, large occlusions, or complex spatial interference, state-of-the-art learning-based methods are expected to achieve superior performance due to their richer spatial modelling capacity. Recent work on rPPG under real-world and extreme lighting conditions employs end-to-end transformer-based architectures with explicit interference disentanglement, background reference modelling, and long-term spatiotemporal context learning [36]. These methods are particularly effective in outdoor environments with drastic illumination changes, periodic external interference, and complex motion patterns, such as driving scenarios. By jointly modelling foreground facial regions and background interference at a fine spatial resolution, these approaches can actively disentangle lighting-induced artefacts that overwhelm subtle biosignals, which provides a capability beyond the scope of the proposed POS–SSA framework.

C. Region-Wise and Pixel-Wise rPPG

Region-wise rPPG approaches, including the proposed POS–SSA framework, aggregate colour information over facial regions to improve signal stability and signal-to-noise ratio. This strategy is robust to sensor noise, minor motion, and local illumination non-uniformity, and offers low computational complexity and high interpretability. However, spatial averaging can attenuate locally informative pulsatile cues when physiological signals are unevenly distributed across the face due to local motion, occlusion, or heterogeneous illumination. Pixel-wise rPPG approaches, by contrast, preserve temporal dynamics at the individual pixel level and are therefore more sensitive to spatially localised pulsatile variations. Methods such as TranSpike [16] exploit pixel-wise frequency reconstruction to retain pulse extrema and spike-like structures, offering improved robustness under extreme or spatially heterogeneous conditions. The trade-off is increased sensitivity to noise, higher computational cost, and reliance on complex learning-based models, which may limit interpretability and deployment in resource-constrained settings. Overall, region-wise and pixel-wise methods represent complementary strategies: the former prioritise stability and efficiency, while the latter emphasise spatial sensitivity under highly challenging conditions.

VI. CONCLUSION

This study proposed a hybrid POS-SSA framework for improving the accuracy and robustness of rPPG-based heart rate estimation from facial videos. By combining the POS colour-space projection with SSA for adaptive signal decomposition, the method effectively mitigates common distortions caused by motion artifacts, illumination fluctuations, and camera noise. Experimental results on the PFF and UBFC-Phys datasets demonstrated that the proposed framework outperforms traditional single-channel, statistical, and chrominance-based approaches. Moreover, when compared with recent learning-based rPPG approaches, the POS–SSA achieves compitive performance while offering several practical advantages. Unlike learning-based approaches that rely on large and annotated datasets and complex spatial–temporal representations, the proposed framework remains data-efficient, computationally predictable, and inherently interpretable. However, the approach still exhibits certain limitations. The computational cost associated with SSA decomposition may constrain real-time deployment on low-power devices, and performance may degrade under extreme motion or occlusion conditions. Furthermore, as the current validation was performed primarily under controlled illumination with limited motion variation, broader evaluation in more dynamic, real-world settings is warranted to confirm its generalisability.

Future work will focus on extending the POS-SSA framework to address these limitations and broaden its applicability. One promising direction is the incorporation of adaptive windowing or real-time incremental SSA to reduce computational latency while preserving signal quality. Additionally, integrating learning–based feature extraction with traditional signal processing could enhance robustness against large head movements, occlusions, and varying illumination. Future studies may also explore multimodal sensing fusion, combining rPPG with thermal or depth imaging to further strengthen resilience under complex environmental conditions. Beyond heart rate estimation, the framework can be extended for heart rate variability (HRV) analysis, stress detection, and emotion recognition, expanding its utility in healthcare, human–computer interaction, and remote monitoring scenarios. The long-term vision is to develop a fully real-time, camera-based vital sign monitoring system capable of delivering clinically meaningful measurements in both controlled and unconstrained environments, thereby contributing to the advancement of contact-free physiological sensing and intelligent health monitoring technologies.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Ruixuan Wang, Wei Quan, Bogdan Matuszewski, Nick A. Heywood, Christopher Gaffney, and Katie Hoad conducted the research; Ruixuan Wang and Wei Quan analysed the data; Ruixuan Wang and Wei Quan wrote the paper. All authors had approved the final version.

FUNDING

This work was supported by the University of Lancashire Doctoral Training Centre for Industry Collaboration.

ACKNOWLEDGMENT

The authors gratefully acknowledge Prof. Gee-Sern Hsu, Dr. Arul Murugan Ambikapathi, and Dr. Ming-Shiang Chen for providing access to the PFF dataset, which made the performance evaluation and comparative analysis possible.

REFERENCES

[1] M. Malik, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 17, no. 3, pp. 354–381, 1996.
[2] P. D. Stein and R. E. Kleiger, "Insights from the study of heart rate variability," *Annu. Rev. Med.*, vol. 50, pp. 249–261, 1999.
[3] G. E. Billman, "The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance," *Front. Physiol.*, vol. 4, 26, 2013.
[4] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Front. Public Health*, vol. 5, 258, 2017.
[5] J. F. Thayer, F. Ahs, M. Fredrikson, J. J. Sollers, and T. D. Wager, "A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart–brain interactions," *Neurosci. Biobehav. Rev.*, vol. 36, no. 2, pp. 747–756, 2012.
[6] U. Rajendra Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: A review," *Med. Biol. Eng. Comput.*, vol. 44, no. 12, pp. 1031–1051, 2006.
[7] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, 2005.
[8] M. Nemec, P. Mlejnek, P. Javorka, and L. Javorka, "Electrocardiography—Golden standard or obsolete method? A review," *Frontiers in Physiology*, vol. 13, 867033, pp. 1–12, 2022.
[9] Y. Xin, C. Zhang, H. Wang, and J. Li, "Photoplethysmography in wearable devices: A comprehensive review of signal processing and applications," *Electronics*, vol. 12, no. 13, art. 2923, pp. 1–21, 2023
[10] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, no. 3, pp. R1–R39, 2007.
[11] Y. Sun and N. Thakor, "Photoplethysmography revisited: From contact to noncontact, from point to imaging," *IEEE Trans. Biomed. Eng*., vol. 63, no. 3, pp. 463–477, 2016.
[12] Y. Bousefsaf, C. Maaoui, and A. Pruski, "Remote detection of mental workload changes using cardiac parameters assessed with a low-cost webcam," *Comput. Biol. Med*., vol. 53, pp. 154–163, 2014.
[13] Z. Yu, X. Li, and G. Zhao, "Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks," in *Proc. British Machine Vision Conf. (BMVC)*, 2019.
[14] X. Liu, B. Hill, Z. Jiang, S. Patel, and D. McDuff, "EfficientPhys: Enabling simple, fast and accurate camera-based cardiac measurement," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, 2023, pp. 4300–4309.
[15] M. Savic and G. Zhao, "PhySU-Net: Long temporal context transformer for remote photoplethysmography with self-supervised pre-training", in *Proc. Int. Conf. Pattern Recognition (ICPR)*, 2024.
[16] H. Shao, L. Luo, J. Qian, C. Hu, S. Chen, and J. Yang, "TranSpike: Pixel-wise frequency reconstruction and spike interaction for remote photoplethysmography," *Pattern Recognition*, vol. 171, 112329, 2026.
[17] T. Balakrishnan, S. Durand, and G. de Haan, "Skin reflection modeling for motion-robust remote photoplethysmography," *Biomed. Opt. Express*, vol. 11, no. 11, pp. 6243–6262, 2020.
[18] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Express*, vol. 16, no. 26, pp. 21434–21445, 2008.
[19] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Opt. Express*, vol. 18, no. 10, pp. 10762–10774, 2010.
[20] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng*., vol. 60, no. 10, pp. 2878–2886, 2013.
[21] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, 2017.
[22] D. McDuff, S. Gontarek, and R. Picard, "Remote measurement of cognitive stress via heart rate variability," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2014, pp. 2957–2960.
[23] C. Duan, X. Liang, and F. Dai, "Optimization of video heart rate detection based on improved SSA algorithm," *Sensors*, vol. 25, 501, 2025.
[24] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2001, pp. 511–518.
[25] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 1994, pp. 593–600.
[26] D.-Y. Kim, K. Lee, and C.-B. Sohn, "Assessment of ROI Selection for Facial Video-Based rPPG," *Sensors*, vol. 21, no. 23, 2021.
[27] K. L. Wong *et al*., "Optimising rPPG Signal Extraction by Exploiting Facial Surface Orientation," in *Proc. CVPR Workshops*, 2022.
[28] R. Wang, H. G. Ma, G. Q. Liu, and D. G. Zuo, "Selection of window length for singular spectrum analysis," *J. Franklin Inst*., vol. 352, no. 4, pp. 1541–1560, 2015.
[29] G.-S. Hsu, A. Ambikapathi, and M.-S. Chen, "Deep learning with time–frequency representation for pulse estimation from facial videos," *Neurocomputing*, vol. 417, pp. 155–166, 2020.
[30] R. Meziati Sabour, Y. Benezeth, P. De Oliveira, J. Chappé,and F. Yang, "UBFC-Phys: A multimodal database for psychophysiological studies of social stress," *IEEE Transactions on Affective Computing*, 2021.
[31] AvLab-CV. (2022). Pulse-From-Face-Database. GitHub repository. [Online]. Available: https://github.com/AvLab-CV/Pulse-From-Face-Database
[32] Z. Yu, Y. Shen, J. Shi, H. Zhao, P. H. S. Torr, and G. Zhao, "PhysFormer: Facial video-based physiological measurement with temporal difference transformer," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
[33] W. Chen and D. McDuff. "DeepPhys: Video-based physiological measurement using convolutional attention networks," in *Proc. the European Conference on Computer Vision (ECCV)*, 2018.
[34] X. Liu, B. Yang, Y. Li, H. Zhang, D. McDuff. "TS-CAN: Temporally shifted convolutional attention network for video-based physiological measurement," in *Proc. the European Conference on Computer Vision (ECCV)*, 2020.
[35] X. Liu, G. Narayanswamy, A. Paruchuri, X. Zhang, J. Tang, Y. Zhang, Y. Wang, S. Sengupta, S. Patel, and D. McDuff. "rPPG-Toolbox: Deep remote photoplethysmography toolbox," in *Proc. the 37th ACM International Conference on Neural Information Processing System (NIPS 2023)*, 2023.
[36] X. Liu, B. Yang, Y. Li, and D. McDuff. "Remote photoplethysmography in real-world and extreme lighting scenarios," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.