Impact of Image Processing Pipeline on Video-based Heart Rate Estimation

Ayane Tohma[a], Kosuke Kurihara[a*], Yoshihiro Maeda[b], Daisuke Sugimura[c], and Takayuki Hamamoto[a]

[a]Tokyo University of Science, Tokyo, Japan
[b]Shibaura Institute of Technology, Tokyo, Japan
[c]Tokyo Metropolitan University, Tokyo, Japan

ABSTRACT

Although previous video-based heart rate (HR) estimation methods operate on color images processed through an image processing pipeline, such pipelines may distort blood volume pulse (BVP) components. In this paper, we analyze the impact of the image processing pipeline on video-based HR estimation. In addition, we introduce a BVP signal extraction method that utilizes raw data from the RGB image sensor. Based on medical knowledge, we can effectively extract the BVP signal from raw data, allowing improvement of HR estimation performance.

Keywords: RGB camera, image processing pipeline, video-based heart rate estimation

1. INTRODUCTION

Heart rate (HR) is essential to assess the physical and mental condition of humans. Conventional HR measurement relies on contact-based sensors, such as pulse oximeters. Although widely adopted in clinical and research settings, such sensors may cause physical discomfort and impose constraints during prolonged use.

Non-contact video-based HR estimation has received considerable research attention.[1,2] This approach is based on the fact that subtle skin color changes, caused by blood volume pulse (BVP) containing cardiac pulsation, can be captured by cameras in a non-contact manner.[3,4]

Several video-based HR estimation methods have been proposed by introducing color-difference representations,[5] spatio-temporal analysis,[4,6] and supervised learning frameworks.[7,8] While these methods have achieved performance improvements, they typically operate on RGB videos where an image processing pipeline is applied to the raw output of digital cameras, commonly referred to as RAW images.

The image processing pipeline consists of demosaicing (i.e., reconstructing a full-color image from a RAW image captured through a color filter array),[9] white balancing,[10] and gamma correction,[10] which are designed to produce visually pleasing images aligned with human perception. Nevertheless, these operations may negatively impact the accuracy of BVP signal extraction. In particular, the amplitude of BVP-induced color variations is less than 2 bits of the analog-to-digital converter of the camera.[11,12] Therefore, when an image processing pipeline is applied to RAW images, these subtle components may be attenuated or lost.

Because image processing pipelines may alter or suppress BVP-relevant signals, it is important to consider the direct use of RAW images before any processing is applied. This approach may retain subtle temporal variations that are essential for accurate HR estimation. However, RAW images also contain substantial noise and irrelevant fluctuations, complicating the extraction of reliable physiological signals.[13] Thus, an appropriate signal processing strategy is needed to effectively utilize a RAW image sequence for video-based HR estimation.

In this paper, we analyze the impact of the image processing pipeline on the performance of video-based HR estimation. By comparing the HR estimation performance using RAW images directly with that using the RGB

Further author information: (Send correspondence to Kosuke Kurihara)

Kosuke Kurihara: E-mail: kuriharak@rs.tus.ac.jp. This research was supported by the JSPS Grants-in-Aid for Scientific Research (JP24K02964, JP24K23903, JP25K21231).

This is the author's version. The official version is available at https://doi.org/10.1117/12.3102486

video processed through an image processing pipeline, we examine whether the pipeline affects the performance of video-based HR estimation. Furthermore, we introduce a BVP signal extraction method for video-based HR estimation that can effectively process RAW images based on temporal characteristics of the BVP signal. By performing a quality assessment in the frequency domain for each RAW pixel, we can effectively use the RAW image sequence for BVP measurement, enabling performance improvement of video-based HR estimation.

2. IMAGE PROCESSING PIPELINE

We briefly describe the image processing pipeline typically applied to the RAW image of RGB cameras, which comprises four main steps: color filter array sampling, demosaicing, white balancing, and gamma correction.

2.1 Sampling With Color Filter Array

Let $\boldsymbol{C}^{(d)} \in \mathbb{Z}^{H\times W}$ denote a RAW image for the $d \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}\}$-th color channel, where H and W represent the vertical and horizontal sizes, respectively. According to the Bayer color filter array,[14] the pixel value at the position $\boldsymbol{p}$ in $\boldsymbol{C}^{(d)}$, denoted by $\boldsymbol{C}^{(d)}_{\boldsymbol{p}}$, is given by

$$\boldsymbol{C}^{(d)}_{\boldsymbol{p}} = \begin{cases} d_{\boldsymbol{p}} & \text{if } \ \boldsymbol{p} \in \Omega^{(d)} \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

where $d_{\boldsymbol{p}}$ denotes the sampled value of the d-th color pixel at position $\boldsymbol{p}$. In addition, $\Omega^{(d)}$ denotes the set of pixel positions corresponding to the d-th color channel in the Bayer color filter array.

2.2 Demosaicing

To interpolate the missing color components, demosaicing is performed on $\boldsymbol{C}^{(d)}$. A well-known demosaicing method is based on bicubic interpolation.[10] The d-th color channel of the demosaiced image $\boldsymbol{I}^{(d)} \in \mathbb{R}^{H\times W}$ is obtained as $\boldsymbol{I}^{(d)} = \boldsymbol{K}^{(d)} * \boldsymbol{C}^{(d)}$, where $*$ denotes the discrete convolution operator. In addition, $\boldsymbol{K}^{(d)} \in \mathbb{R}^{\kappa\times\kappa}$ denotes the convolution kernel matrix for the d-th color channel based on bicubic interpolation.

2.3 White Balance

White balance is effective in correcting color casts caused by ambient illumination. A common algorithm is based on the gray world assumption, assuming that the average color of a scene is achromatic; i.e., the spatial average of R, G, and B channels has the same value.[10] The gain for the d-th channel, denoted by $g^{(d)}$, is calculated as

$$g^{(d)} = \frac{\max\limits_{d\in\{\mathrm{R},\mathrm{G},\mathrm{B}\}} \sum_{\boldsymbol{p}\in\Omega} \boldsymbol{I}^{(d)}_{\boldsymbol{p}}}{\sum_{\boldsymbol{p}\in\Omega} \boldsymbol{I}^{(d)}_{\boldsymbol{p}}}, \tag{2}$$

where $\boldsymbol{I}^{(d)}_{\boldsymbol{p}}$ is the pixel value of channel d at position $\boldsymbol{p}$. The color-corrected image is obtained as $\tilde{\boldsymbol{I}}^{(d)} = g^{(d)}\boldsymbol{I}^{(d)}$.

2.4 Gamma Correction

Gamma correction is applied at the final stage to adjust image brightness and contrast, accounting for the nonlinear response of display devices and characteristics of human visual perception. The d-th channel of the resulting RGB image $\boldsymbol{Y}^{(d)}$ is calculated as $\boldsymbol{Y}^{(d)} = (\tilde{\boldsymbol{I}}^{(d)})^{\frac{1}{\gamma}}$, where γ denotes the gamma correction factor.

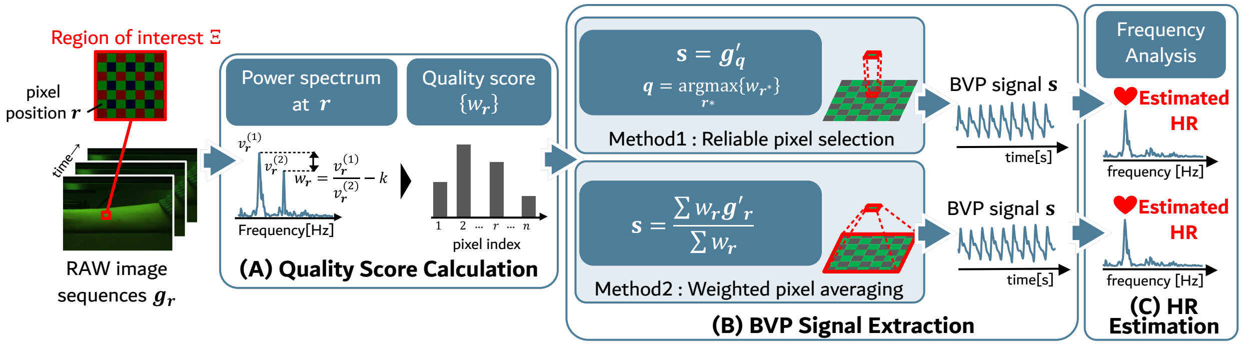

Region of interest Ξ
pixel position $\boldsymbol{r}$
time
RAW image sequences $\boldsymbol{g_r}$
Power spectrum at $\boldsymbol{r}$
Quality score $\{w_r\}$
$v_r^{(1)}$
$v_r^{(2)}$
$w_r = \frac{v_r^{(1)}}{v_r^{(2)}} - k$
Frequency[Hz]
1 2 … r … n
pixel index
(A) Quality Score Calculation
$\mathbf{s} = \boldsymbol{g}'_q$
$\boldsymbol{q} = \mathrm{argmax}_{\boldsymbol{r}*}\{w_{\boldsymbol{r}^*}\}$
Method1 : Reliable pixel selection
$\mathbf{s} = \frac{\sum w_r \boldsymbol{g}'_r}{\sum w_r}$
Method2 : Weighted pixel averaging
BVP signal s
time[s]
BVP signal s
time[s]
(B) BVP Signal Extraction
Frequency Analysis
Estimated HR
frequency [Hz]
Estimated HR
frequency [Hz]
(C) HR Estimation

Figure 1. Overview of our method: (A) Based on the quasi-periodic characteristics of the BVP, we calculate the quality score of the time-series signal for each green RAW pixel. (B) Using the quality score, we extract BVP signals from the RAW image sequence. (C) By searching the peak of the power spectrum of the extracted BVP signals, we estimate HR.

3. RAW-BASED HR ESTIMATION

3.1 Overview

We propose an HR estimation method that processes RAW images directly based on the physiological characteristics of the BVP signal. Figure 1 shows an overview of our method. According to past studies, due to the light absorbance characteristics of blood and skin, green components have the highest sensitivity to the BVP component among the red, green, and blue components.[15, 16] Based on this knowledge, we utilize the green component. To extract the BVP signal, we calculate the quality score for each green RAW pixel based on the temporal characteristics of the BVP signal. It is well established that BVP signals exhibit quasi-periodic characteristics driven by the cardiac cycle.[17] Based on this characteristic, we calculate the quality score by performing a frequency analysis. Using the quality scores, we can effectively extract the BVP signal, enabling us to improve the HR estimation performance. In the following section, we describe the details of our method.

3.2 Quality Score Calculation

Let the set of pixel positions belonging to the region of interest (ROI) be Ξ. For each green pixel within ROI, i.e., $\boldsymbol{r} \in (\Omega^{(\mathrm{G})} \cap \Xi)$, we define the time-series signal of the RAW image sequence as $\boldsymbol{g_r} \in \mathbb{R}^T$, where T denotes the length of the sequence.

We first assess the quality of $\boldsymbol{g_r}$. Previous studies have shown that the BVP signal is dominated by a single frequency component corresponding to the quasi-periodic cardiac cycle, which lies within the established physiological frequency range.[18] Based on this property, quality can be evaluated by measuring the spectral sharpness within the expected BVP frequency band.[12]

To this end, we first apply a band-pass filter with the passband f_{BVP}, which corresponds to the typical BVP frequency range. The resulting signal $\boldsymbol{g}'_{\boldsymbol{r}}$ is obtained as $\boldsymbol{g}'_{\boldsymbol{r}} = \mathrm{BPF}(\boldsymbol{g_r})$, where $\mathrm{BPF}(\cdot)$ represents the band-pass filter operator. We then compute the power spectrum of $\boldsymbol{g}'_{\boldsymbol{r}}$ as $\boldsymbol{v_r} = |\,\mathcal{F}(\boldsymbol{g}'_{\boldsymbol{r}})\,|^2$, where $\boldsymbol{v_r} \in \mathbb{R}^F$ denotes the power spectrum consisting of frequency components $f = (1, \ldots, F)$, and $\mathcal{F}(\cdot)$ represents the discrete Fourier transform.

From $\boldsymbol{v_r}$, we search for the first and second largest frequency components, denoted by $v_{\boldsymbol{r}}^{(1)}$ and $v_{\boldsymbol{r}}^{(2)}$, as

$$v_{\boldsymbol{r}}^{(1)} = \max_{f \in f_{\mathrm{BVP}}} \boldsymbol{v_r}[f] \,, \qquad v_{\boldsymbol{r}}^{(2)} = \max_{f \in \{f_{\mathrm{BVP}} \setminus f_{\boldsymbol{r}}^{(1)}\}} \boldsymbol{v_r}[f] \,, \tag{3}$$

where $f_{\boldsymbol{r}}^{(1)}$ represents the frequency index of $v_{\boldsymbol{r}}^{(1)}$.

We calculate the quality score at $\boldsymbol{r}$, denoted by $w_{\boldsymbol{r}}$ as

$$w_{\boldsymbol{r}} = \frac{v_{\boldsymbol{r}}^{(1)}}{v_{\boldsymbol{r}}^{(2)}} - k \;, \tag{4}$$

where $k = 1$ denotes the normalization constant that ensures that the minimum value of $w_{\boldsymbol{r}}$ is 0.

3.3 BVP Signal Extraction

We extract the BVP signal $\boldsymbol{s} \in \mathbb{R}^T$ from the green pixel time-series signal $\boldsymbol{g}'_{\boldsymbol{r}}$. To this end, two BVP signal extraction strategies on the basis of quality scores $w_{\boldsymbol{r}}$ are considered.

Reliable pixel selection: We select the pixel with the highest quality score as the suitable pixel for HR estimation. We extract the BVP signal $\boldsymbol{s}$ as

$$\boldsymbol{s} = \boldsymbol{g}'_{\boldsymbol{q}} \, , \quad \text{where} \;\; \boldsymbol{q} = \mathop{\arg\max}_{\boldsymbol{r}^* \in \{\Omega^{(\mathrm{G})} \cap \Xi\}} \; w_{\boldsymbol{r}^*} \; . \tag{5}$$

Weighted pixel averaging: According to past studies, multiple pixel averaging can reduce noise and improve HR estimation performance.[11,12] Thus, we extract $\boldsymbol{s}$ by calculating the weighted average of $\boldsymbol{g}_{\boldsymbol{r}}$ using $w_{\boldsymbol{r}}$ as

$$\boldsymbol{s} = \frac{\sum_{\boldsymbol{r} \in \{\Omega^{(\mathrm{G})} \cap \Xi\}} w_{\boldsymbol{r}} \, \boldsymbol{g}'_{\boldsymbol{r}}}{\sum_{\boldsymbol{r} \in \{\Omega^{(\mathrm{G})} \cap \Xi\}} w_{\boldsymbol{r}}} \; . \tag{6}$$

3.4 HR Estimation

To calculate HR, we first apply a band-pass filter with the frequency range f_{BVP} to the obtained BVP signal $\boldsymbol{s}$. The filtered BVP signal is denoted by $\boldsymbol{s}' = \mathrm{BPF}(\boldsymbol{s})$. We then compute the power spectrum of $\boldsymbol{s}'$ as $\boldsymbol{x} = |\mathcal{F}(\boldsymbol{s}')|^2$, where $\boldsymbol{x} \in \mathbb{R}^F$ denotes the obtained power spectrum consisting of frequency components $f = (1, \ldots, F)$.

Since the dominant frequency component in $\boldsymbol{x}$ is expected to be HR,[19] we identify the frequency index with the highest power spectrum as

$$f^* = \mathop{\arg\max}_{f \in f_{\mathrm{BVP}}} \boldsymbol{x}[f] \; . \tag{7}$$

We finally obtain the estimated HR in beats per minute [bpm] by $\mathrm{HR} = 60 \times f^*$.

4. EXPERIMENTS

To demonstrate the effectiveness of our method, we conducted experiments using real RAW image sequences that were collected under controlled conditions.

4.1 Dataset

We recorded RAW image sequences capturing the right arm for 135 seconds using an RGB camera (JAI AD-130GE). The RGB camera has a 1/3-inch CCD sensor with a Bayer filter color array and captures 12-bit RAW images with a resolution of 1296 × 966 pixels and a frame rate of 30 fps. Five healthy subjects participated and were instructed to sit still and place their right arms on a table. To obtain a ground-truth HR, we used a contact-type pulse oximetry sensor (CONTEC CMS50D+), attached to the finger of each subject.

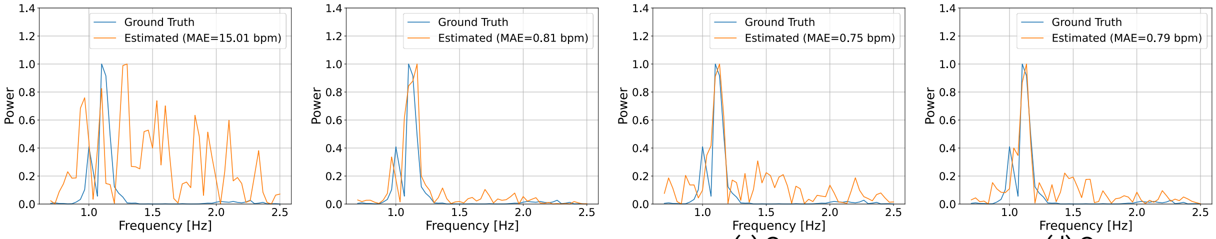
4.2 Comparison Methods

As a comparison method, we employed a video-based HR estimation method using RGB image sequence that was processed through an image processing pipeline, referred to as *Pipeline*. The image processing pipeline consisted of bicubic demosaicing (Sect. 2.2), gray-world white balance correction (Sect. 2.3), gamma encoding (Sect. 2.4), and final 8-bit quantization to generate RGB image sequence. In addition, we adopted a method that directly utilizes RAW image sequence without quality score, referred to as *Raw*. To ensure a fair comparison, these comparison methods employ identical BVP signal extraction procedures: extracting the green channel time-series signal from the same ROI Ξ.

The BVP frequency range f_{BVP} was set as 0.7 - 2.5 Hz based on medical knowledge on BVP.[18] The length of the input sequence T for our method and the comparison methods was set at 900. For *Pipeline*, the kernel size κ in the demosaicing process was set to 7. The gamma factor was set to $\gamma = 2.2$. The size of the ROI Ξ for our method and *Raw* was set to 7×7 pixels, while that for *Pipeline* was set to 1×1 pixels, considering the convolutional kernel size $\kappa = 7$ to ensure a fair comparison.

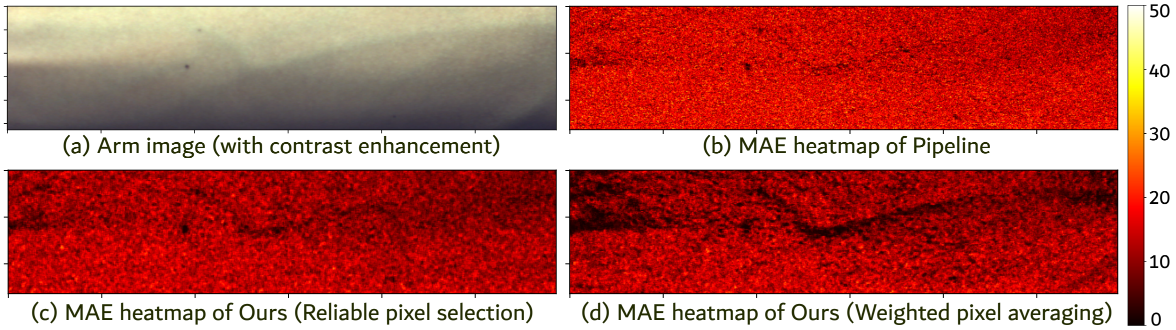
Table 1. Quantitative results using MAE [bpm] and SR [%].

	Pipeline	Raw	Ours (Reliable pixel selection)	Ours (Weighted pixel averaging)
MAE ↓	14.57	12.40	**11.37**	11.56
SR ↑	24.53	34.99	31.53	**35.28**



(a) Pipeline (b) RAW (c) Ours (Reliable pixel selection) (d) Ours (Weighted pixel averaging)

Figure 2. Comparison results for power spectrum of estimated BVP signals for Subject #4.



(a) Arm image (with contrast enhancement)

(b) MAE heatmap of Pipeline

(c) MAE heatmap of Ours (Reliable pixel selection)

(d) MAE heatmap of Ours (Weighted pixel averaging)

Figure 3. Arm image and MAE heatmap of subject #4.

4.3 Evaluation Metrics

We evaluated HR estimation performance using two metrics: mean absolute error (MAE) and success rate (SR). MAE is defined as the average of absolute errors between the estimated HR and the ground-truth HR. SR is defined as the percentage of cases in which the HR estimation error was within ±5 bpm.[20, 21] The better HR estimation performance is represented by lower MAE and higher SR values. We comprehensively evaluated the HR estimation performance using these metrics by sliding the ROI across the entire arm region in 1-pixel increments and sliding the time window in 1-second increments.

4.4 Results

Table 1 shows the comparison results using MAE and SR. Compared to *Raw*, *Pipeline* exhibits lower accuracy, indicating that the image processing pipeline negatively affects HR estimation performance. Furthermore, our method exhibits performance superior to that of *Raw*, indicating the effectiveness of quality scores based on the quasi-periodic characteristics of the BVP signal.

Figure 2 shows an example of the power spectrum of the extracted BVP signals for each method of subject #4. It can be seen that the BVP signals extracted by our methods are consistent with the ground-truth signal. We visualized the MAE at each ROI by sliding the ROI across the entire arm region. Figure 3 (a) shows the entire arm of subject #4 with contrast enhancement. Figures 3 (b), (c), and (d) show MAE heatmaps of subject #4 of *Pipeline*, ours with reliable pixel selection, and ours with weighted pixel averaging, respectively. Note that the heatmaps in Figures 3 (b), (c), and (d) are spatially aligned with the arm image in Figure 3 (a). It can be seen that our methods exhibit darker colors, indicating better performance than *Pipeline*.

5. CONCLUSION

We analyzed the impact of the image processing pipeline on video-based HR estimation. Based on medical knowledge of the BVP signal, we also proposed an HR estimation method that can effectively extract BVP signals from RAW image sequence. The experimental results revealed that the image processing pipeline negatively affects HR estimation performance. In addition, we demonstrated that the introduction of quality scores can enhance the performance of HR estimation. Our results suggest that the inherent performance limitations of video-based HR estimation methods can be overcome by rethinking the processing of RAW image sequences.

REFERENCES

[1] Chen, X., Cheng, J., Song, R., Liu, Y., Ward, R., and Wang, Z. J., "Video-based heart rate measurement: Recent advances and future prospects," *IEEE Transactions on Instrumentation and Measurement* **68**(10), 3600–3615 (2018).

[2] Kurihara, K., Sugimura, D., and Hamamoto, T., "Adaptive fusion of RGB/NIR signals based on face/background cross-spectral analysis for heart rate estimation," in *Proc. IEEE International Conference on Image Processing (ICIP)*, 4534–4538 (2019).

[3] Verkruysse, W., Svaasand, L. O., and Nelson, J. S., "Remote plethysmographic imaging using ambient light.," *Optics express* **16**(26), 21434–21445 (2008).

[4] Kurihara, K., Maeda, Y., Sugimura, D., and Hamamoto, T., "Spatio-temporal structure extraction of blood volume pulse using dynamic mode decomposition for heart rate estimation," *IEEE Access* **11**, 59081–59096 (2023).

[5] Wang, W., den Brinker, A. C., Stuijk, S., and de Haan, G., "Algorithmic principles of remote ppg," *IEEE Transactions on Biomedical Engineering* **64**(7), 1479–1491 (2017).

[6] Kurihara, K., Maeda, Y., Sugimura, D., and Hamamoto, T., "Blood volume pulse signal extraction based on spatio-temporal low-rank approximation for heart rate estimation," in *Proc. IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 1–5 (2022).

[7] Sun, Z. and Li, X., "Contrast-phys+: Unsupervised and weakly-supervised video-based remote physiological measurement via spatiotemporal contrast," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **46**(8), 5835–5851 (2024).

[8] Liu, X., Zhang, Y., Yu, Z., Lu, H., Yue, H., and Yang, J., "rppg-mae: Self-supervised pretraining with masked autoencoders for remote physiological measurements," *IEEE Transactions on Multimedia* **26**, 7278–7293 (2024).

[9] Gunturk, B., Glotzbach, J., Altunbasak, Y., Schafer, R., and Mersereau, R., "Demosaicking: color filter array interpolation," *IEEE Signal Processing Magazine* **22**(1), 44–54 (2005).

[10] Lukac, R., *Single-Sensor Imaging: Methods and Applications for Digital Cameras*, CRC Press, Inc., USA (2008).

[11] Kumar, M., Veeraraghavan, A., and Sabharwal, A., "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomedical optics express* **6**(5), 1565–1588 (2015).

[12] Kurihara, K., Sugimura, D., and Hamamoto, T., "Non-contact heart rate estimation via adaptive rgb/nir signal fusion," *IEEE Transactions on Image Processing* **30**, 6528–6543 (2021).

[13] Foi, A., Trimeche, M., Katkovnik, V., and Egiazarian, K., "Practical poissonian-gaussian noise modeling and fitting for single-image raw-data," *IEEE Transactions on Image Processing* **17**(10), 1737–1754 (2008).

[14] Lukac, R. and Plataniotis, K., "Color filter arrays: design and performance analysis," *IEEE Transactions on Consumer Electronics* **51**(4), 1260–1267 (2005).

[15] Blackford, E. B., Estepp, J. R., and McDuff, D., "Remote spectral measurements of the blood volume pulse with applications for imaging photoplethysmography," in *Proc. Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics*, **10501**, 105010Z (2018).

[16] Kurihara, K., Maeda, Y., Sugimura, D., and Hamamoto, T., "Physiological modeling with multispectral imaging for heart rate estimation," in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2957–2963 (2024).

[17] Elgendi, M., "On the analysis of fingertip photoplethysmogram signals," *Current Cardiology Reviews* **8**(1), 14–25 (2012).

[18] Palatini, P., "Need for a revision of the normal limits of resting heart rate," *Hypertension* **33**(2), 622–625 (1999).

[19] Poh, M., McDuff, D., and Picard, R. W., "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Transactions on Biomedical Engineering* **58**(1), 7–11 (2011).

[20] Kurihara, K., Maeda, Y., Sugimura, D., and Hamamoto, T., "Unified physiological and illumination modeling for heart rate estimation using dynamic mode decomposition and rgb/nir sensor," *IEICE Transactions on Information and Systems* , Early Access (2026).

[21] Lam, A. and Kuno, Y., "Robust heart rate measurement from video using select random patches," in *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 3640–3648 (2015).