Feature-Disentangling RGB-NIR Fusion Network for Remote Driver Physiological Measurement

Tayssir Bouraffa* Ziyuan Wang*
Daniel Strüber
Chalmers University of Technology
Gothenburg, Sweden
tayssir.bouraffa@volvocars.com, wangzi@chalmers.se, danstru@chalmers.se

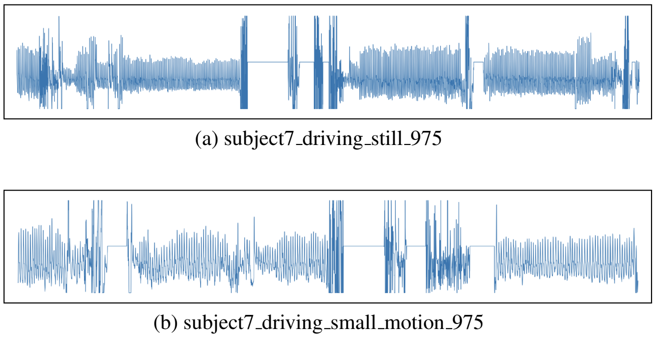
Appendix A: Excluded Cases

As illustrated in Table 1, specific instances from the MR-NIRP dataset were excluded from the study to prevent factors that could compromise the accuracy of the experiments.

Specifically, five sequences were too dark to detect the subject's face, and one case involved corrupted video frames. Additionally, in two instances the recorded ground-truth PPG signals contained extended large spans of zeroes, suggesting an issue in the sampling process, as outlined in Figure 1. Additionally, all samples from Subject 12 were omitted from the experiments due to excessive noise detected in the ground-truth signal during PSD analysis, suggesting a potential sampling error, as illustrated in Figure 2. Furthermore, the HR extracted from these PPG signals was consistently below than 50 Beats/min which is not typical for a healthy adult male like Subject 12. This issue was observed in the majority of this subject's recordings, rendering them unsuitable for accurate ground-truth vital sign extraction [1].

Table 1. List of excluded cases from MR-NIRP car dataset

Justification	Excluded Cases
5***Dark Frames**	subject5_garage_still_975 subject6_garage_still_975 subject6_garage_small_motion_975 subject6_garage_large_motion_975 subject2_driving_still_940
Corrupted Frames	subject2_garage_small_motion_940
3***PPG Sampling Error**	subject7_driving_small_motion_975 subject7_driving_still_975 subject12*


(a) subject7_driving_still_975

(b) subject7_driving_small_motion_975

Figure 1. Ground-truth PPG signals showing extensive spans of zeros, indicating sampling errors.

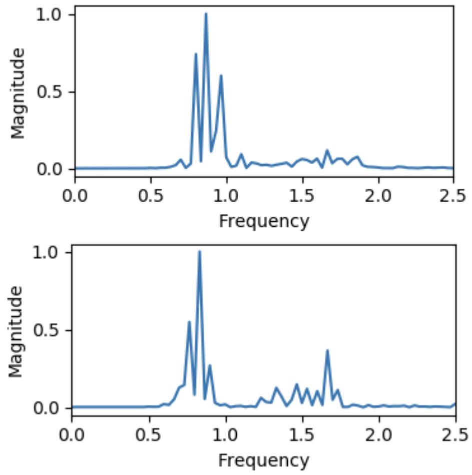


Figure 2. Example PSD plots from Subject 12. The samples were taken from the recordings "subject12_garage_small_motion_975" and "subject12_driving_still_940" respectively.

Appendix B: Evaluation Metrics

To evaluate the effectiveness of rPPG algorithms and accurately predicting HR under dynamic vehicular conditions,

*Equal contribution.

we selected six metrics that address different aspects of algorithm performance, including error rates and signal quality, and processing speed.

Mean Absolute Error (MAE): Measures the average of the absolute differences between the ground-truth (R_{GT}) and the predicted signal rate (R_{Pred}) for HR over all observation windows (T).

$$\textbf{MAE} = \frac{1}{T}\sum_{i=1}^{T} |R_{GT} - R_{Pred}|$$

Root Mean Square Error (RMSE): Assesses the magnitude of the prediction error in comparison to the ground-truth (R_{GT}) and the predicted signal rates (R_{Pred}) over all observation windows (T).

$$\textbf{RMSE} = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(R_{GT} - R_{Pred})^2}$$

Mean Absolute Percentage Error (MAPE): Computes the average of the absolute percentage differences between the ground-truth (R_{GT}) and the predicted signal rates (R_{Pred}), represented as a percentage of the ground-truth signal rate across all observation windows (T).

$$\textbf{MAPE} = \frac{100}{T}\sum_{i=1}^{T}\left|\frac{R_{GT} - R_{Pred}}{R_{GT}}\right|$$

Pearson Correlation Coefficient (ρ): A statistical metric used to assess the linear correlation between the ground-truth (R_{GT}) and the predicted signal rates (R_{Pred}) across all windows (T).

$$\rho = \frac{\sum_{i=1}^{T}(R_{GT} - \overline{R_{GT}})(R_{Pred} - \overline{R_{Pred}})}{\sqrt{\sum_{i=1}^{T}(R_{GT} - \overline{R_{GT}})^2 \sum_{i=1}^{T}(R_{Pred} - \overline{R_{Pred}})^2}}$$

Signal-to-Noise Ratio (SNR): Calculated as the ratio between the area under the curve of the power spectrum near the first and second harmonic of the ground-truth signal rate frequency and the area under the curve for the remainder of the power spectrum.

$$\textbf{SNR} = \frac{1}{T}\sum_{i=1}^{T}\left|10\log_{10}\left(\frac{\sum_{f=lf}^{hf}\left(\hat{S}(f)\cdot U_t(f)\right)^2}{\sum_{f=lf}^{hf}\left(\hat{S}(f)\cdot(1-U_t(f))\right)^2}\right)\right|$$

$\hat{S}$ denotes the power spectrum of the predicted signal S, f refers to the frequency, and $U_t(f)$ is a binary template that is set to 1 around the first and second harmonics of the ground-truth signal and 0 elsewhere. In this approach, only the power spectrum within the frequency ranges of $(lf - hf)$ of 0.75 - 2.5 Hz for HR.

Frames Per Second (FPS): Measures the processing speed of the model during evaluation, representing how many frames the algorithm can process per second. This metric is crucial for assessing real-time performance, especially for applications in dynamic vehicular environments where timely HR estimation is essential for effective driver monitoring.

References

[1] Y. Hafeez and G. Shamai A., "Sinus bradycardia," *StatPearls [Internet]. Treasure Island, FL: StatPearls Publishing. https://www.ncbi.nlm. nih.gov/books/NBK493201/*, 2023. 1