The 12th International Conference on E-Health and Bioengineering - EHB 2024
Grigore T. Popa University of Medicine and Pharmacy Iasi, November 14-15, 2024, Iasi, ROMANIA

Preliminary Analysis on the Usage of Hyperspectral Reconstruction for Imaging Photoplethysmography and Heart Rate Detection

Tobias Vogelsang[1], Alexander Woyczyk[1], Filip Packań[1,2], Jonas Kolonko[1], Alwin Hoffmann[3], Björn Schuller[2], Shahin Amiriparian[2], Sebastian Zaunseder[1]

Affiliation 1: Chair of Diagnostic Sensing, University of Augsburg, Augsburg, Germany, tobias.vogelsang@uni-a.de, jonas.kolonko@uni-a.de, alexander.woyczyk@uni-a.de, sebastian.zaunseder@uni-a.de

Affiliation 2: Chair of Health Informatics, Technical University of Munich, Munich, Germany, filip.packan@tum.de, schuller@tum.de, shahin.amiriparian@tum.de,

Affiliation 3: Xitaso GmbH, Augsburg, Germany, alwin.hoffmann@xitaso.com

***Abstract*— Hyperspectral imaging (HSI) of the skin enables far-reaching diagnostic statements concerning anatomical and physiological aspects. However, hyperspectral cameras are expensive and have limitations with regard to spatial and temporal resolution. Hyperspectral reconstruction provides a means to transfer RGB data to a hyperspectral representation, potentially overcoming limitations of equipment for HSI. This contribution investigates whether a state-of-the-art deep learning (DL) technique is usable to transform RGB videos to a hyperspectral representation and if such representation can be used to extract the blood volume pulse (BVP) and heart rate (HR). Our results indicate that the chosen DL technique performs well on the reconstruction task using the Hyper-Skin database. At the same time, the physiological information is preserved. E.g. with respect to HR extraction in own experimental data, using the original green channel yields a correlation coefficient of 0.81 to a reference HR. When using a synthesized green channel from the DL reconstruction, the correlation even rises to 0.93. Using a regression-based approach for hyperspectral reconstruction, we achieved a correlation of 0.92. Our findings indicate the potential of using hyperspectral reconstruction to yield physiological information from videos. Future works will focus on dedicated methods to process the reconstructed hyperspectral data to exploit the full potential of the pursued approach.**

Keywords*— *imaging photoplethysmography, hyperspectral reconstruction, heart rate, blood volume puls

I. Introduction

Over the last years, imaging photoplethysmography (iPPG) has been attracting immense interest. iPPG assesses the cutaneous perfusion by exploiting subtle color variations from videos to yield various physiological information. The technology is extremely attractive, as conventional RGB cameras can be used, and measurements do not require any contact. iPPG can capture multiple parameters such as heart rate (HR), heart rate variability (HRV), oxygen saturation, blood pressure, venous pulsation, strength of cutaneous perfusion and spatiotemporal dynamics of cutaneous perfusion [1–3]. Such variety opens up multiple clinical and non-clinical applications [4].

Conventional iPPG processing procedures create a blood volume pulse (BVP) signal, which is processed afterwards [3]. BVP signal creation invokes three steps: (1) segmentation and tracking of a region of interest (ROI), (2) formation of a raw signal from the ROI, and (3) signal filtering. The subsequent processing of the BVP signal depends on the target, e.g., to yield HR, most often time-frequency-analysis is used. The formation of the raw signal (step 2) turned out to be of particular relevance for iPPG. The usage of a single color channel (owing to the best signal quality typically the green color channel) or linear combinations of RGB are common [5]. Recently, the use of machine learning (ML) has become very popular in iPPG [6, 7]. Proposed ML approaches cover different strategies: image sequence to target parameter (end-to-end strategy), image sequence to BVP, and raw signal to BVP and raw signal to target parameter. Accordingly, the usage of ML can imply modifications to the outlined processing procedure, e.g., make ROI segmentation obsolete.

Hyperspectral imaging (HSI) also can yield diagnostic statements remotely [8, 9]. By exploiting spectroscopic characteristics of cutaneous chromophores, HSI enables far-reaching statements on metabolism, perfusion and anatomical aspects. Hyperspectral recordings, however, typically suffer from reduced spatial and/or temporal resolution and equipment for HSI is expensive. Hyperspectral reconstruction has the potential to overcome such limitations by computing hyperspectral data from RGB data as obtained by regular cameras. Hyperspectral reconstruction is an area of active research [10, 11]. Recently, deep neural networks have been demonstrated to yield most accurate results for hyperspectral

reconstruction considering common error measures as mean squared error (MSE) or spectral angle mapper (SAM). However, using hyperspectral reconstruction in the context of iPPG is not common and, to the best of our knowledge, there are no works to investigate to which extend deep neural networks maintain the information on BVP.

This contribution presents a preliminary evaluation on the effect of hyperspectral reconstruction by DL for iPPG. The contribution covers training of a state of the art network for hyperspectral reconstruction using the Hyper-Skin database [12], formation of an iPPG signal in the green channel (using either the native green channel or reconstructed green channels) using own experimental data [13], and finally an assessment regarding the hyperspectral reconstruction and the ability of the used network to retain information on the BVP and capture HR.

II. Used Data

A. Hyper-Skin data set

1) Overview

To train the methods for hyperspectral reconstruction we use the Hyper-Skin data set [12]. The Hyper-Skin database was developed with the objective of optimizing research in the field of hyperspectral reconstruction from RGB images with a medical context. Moreover, as stated by the authors, the database is designed to facilitate the examination of skin characteristics, including melanin and hemoglobin concentration. In order to allow for comparisons with the NTIRE challenges, the database was designed to match the database of the NTIRE challenges [14, 15].

2) Experimental Procedure

The faces of healthy subjects were captured from the front, left, and right with both a neutral and a smiling facial expression, resulting in six data cubes per subject. To ensure the accuracy and consistency of the data, the subjects were instructed to stand still with the help of a chin rest.

3) Technical Setup

The recordings were conducted under controlled illumination conditions, utilizing two halogen lamps, and employed spectral scanning with the FX-10 from Specim (Oulu, Finland). The camera is capable of capturing both the visible spectrum (400 nm–700 nm) and the near-infrared spectrum (700 nm–1000 nm), with a spatial resolution of 1024×1024. For each defined spectrum, 224 bands were captured, resulting in a maximum of 448 bands. The authors [12] downsampled the number of bands to ensure comparability with the NTIRE challenges database. The Matlab function "HSI2RGB" was employed to extract RGB reference images from all hyperspectral images. For this study, only the VIS was utilized.

4) Participants

A total of 51 healthy participants volunteered for the experiment. The majority of them were in their early 20s or 30s, and there were slightly more male than female participants.

B. Imaging PPG Data

1) Overview

To assess the ability to pertain the BVP and extract the HR, we use data from own multimodal experiments. Our experiments invoke healthy volunteers of Caucasian origin who underwent different stimuli, namely paced deep breathing (PDB), multiple orthostatic maneuvers, and cold pressure test (CPT). Throughout the experiment, we recorded three videos and multiple biosignals as reference. The study was approved by the Ethics Committee at TU Dresden (EK 311082018). All subjects gave written consent. A more detailed explanation on the study can be found in [13]. Below, we provide details on those aspects that are relevant to this contribution.

2) Experimental Procedure

Fig. 2 depicts the experimental protocol. The experiments had a duration of approximately 49 minutes and were carried out on a tilt table. During the experiment, the tilt-table was alternated between supine and upright position every seven minutes, defining seven phases. Between orthostatic maneuvers participants had resting epochs and executed CPT or PDB. Each participant executed at least one CPT (denoted as CPT1), which was randomly assigned to phase 1 or phase 3. A random subset of participants executed another CPT (CPT2) in phase 5. During CPT, participants immersed their left hand into cold water at a temperature of approximately 4° C for 60 s (termination was possible at any time if participants felt uncomfortable). CPT is known to induce a sympathetic activation owing to the painful stimuli. As a result, blood pressure typically increases rapidly and stays increased throughout the time of immersion. HR shows a more undefined behavior. Typically, there is a fast increase upon immersion. Afterwards, the behavior varies between subjects. In some cases, increased heart rate levels are preserved while in other cases heart rate recedes, sometimes falling below the initial level. Within this contribution, we use data excerpts of 90 s starting 30 s before CPT1 and ending 60 s after CPT1.

3) Technical Setup

We recorded RGB videos and reference signals continuously throughout the experiment. Videos were recorded by three UI-3060CP-C-HQ Rev.2 RGB cameras (IDS Imaging Development Systems GmbH; Obersulm, Germany). The cameras were mounted on the tilt table and had fixed orientation with respect to the subject during the experiment. Within this work, only videos from camera 2 are considered. This camera was placed at a distance of approximately 40 cm to subject´s head and covered to a subject's head and a small portion of the shoulders (see Figure 2). Videos were captured at a color depth of 8 bit, a frame rate of 25 Hz, and a spatial resolution of 1280×960 pixel. We stored all videos in a format with lossless compression. We used artificial illumination by two spotlights Walimex pro LED Sirius 160 Daylight 65 W (color temperature 5600 K, color rendering index ≥ 90) (WALSER GmbH & Co. KG; Gersthofen, Germany). Biosignals were acquired by two

Biopac MP36 (Biopac; Goleta, United States of America) and HR. HR was derived from a single lead electrocardiogram (Einthoven II). Camera frame times are synchronized by feeding each camera's trigger signal into the Biopac units along with the biosignals.

4) Participants

Overall, 61 recordings were carried out using the setting described above. Within this preliminary study we used data from 20 subjects. The mean age and standard deviation of the subjects was 30.95±12.25.

III. Methods

A. iPPG Processing

1) Overview

The employed processing adheres to the general processing scheme to yield the BVP, and from it the HR, consisting of (1) segmentation and tracking of a ROI, (2) formation of a raw signal from the ROI, and (3) signal filtering. Using such filtered signal, we finally derive the HR in the frequency domain. Below, we provide details on the single steps.

2) Segmentation and Tracking

In order to create the iPPG signal quality, only pixels within a ROI, i.e., pixels showing skin, were used. The ROI was formed by a specialized segmentation algorithm introduced by Woyczyk et al. [16]. The algorithm uses a modified level set approach and subject dependent skin color models to derive a suitable ROI.

a) Skin Model Initialization

The segmentation algorithm uses Gaussian Mixture Models (GMMs) to model the color distributions of foreground (skin) and background (non skin). To initialize both models, a face detection algorithm searches for a face in the first frame. The employed face detector is distributed with OpenCV and uses a cascaded classifier with haar features [17]. The detector provides a bounding box indicating the position and size of the detected face. Using Jones and Rehg's static skin classifier, definitive background pixels are eliminated from the bounding box. The remaining pixels are then used to train the GMM of the foreground. As further refinement step, kernels showing a low overlap with the static skin model are excluded from the GMM, as this indicates a possibility of unwanted color information in the model. The background GMM on the other hand is trained on all pixels outside the bounding box. As the number of pixels is large enough, no further adjustments of the background kernels are made.

b) Level Set Segmentation

The actual segmentation (see Fig. 2) and tracking is carried out by a level set formulation which is loosely based on the work of Chan and Vese [18]. A level set segmentation formulates a function (ϕ function) over the two-dimensional image space, giving every pixel an energy value. The segmentation is then concluded by setting a threshold (usually a zero threshold) and assigning each pixel either to fore or background depending on its energy value, i.e., above threshold is foreground and below threshold is background. The crucial part of a level set segmentation is to define the energy term, which defines the ϕ function. In our solution, the ϕ function of an image is approximated by minimizing the total energy of a contour and the area enclosed. The total energy is derived from the length of the contour, its enclosed area, and lastly the distance of each pixel's intensity to the GMM of its assigned class. The final approximation of the ϕ function is calculated through an iterative algorithm.

3) Formation of Raw Signal

Within this contribution, we consider only the green channel (the original one and two reconstructed versions; see below for details on the reconstruction). The raw BVP signal for the native RGB data is calculated by averaging the green color channel intensities within the ROI. For the reconstructed data, a synthetic green channel is calculated by a weighted average (Gaussian, centered at 550 nm, standard deviation 20 nm) over the reconstruced HSI data cube. The reconstructed green channel is afterwards processed in the same way as the native green channel, i.e., averaging the intensities within the ROI to yield a time varying signal.

4) Signal Filtering

As we are only interested in the pulsatile behavior of BVP and HR, we apply a narrow band bandpass filter. The used filter is a 5th order Butterworth filter with lower and upper cutoff frequencies of 2/3 Hz and 4 Hz. The filter is applied in forward and backward direction to avoid phase shifts. An example of the PPG and iPPG can be seen in figure 2.

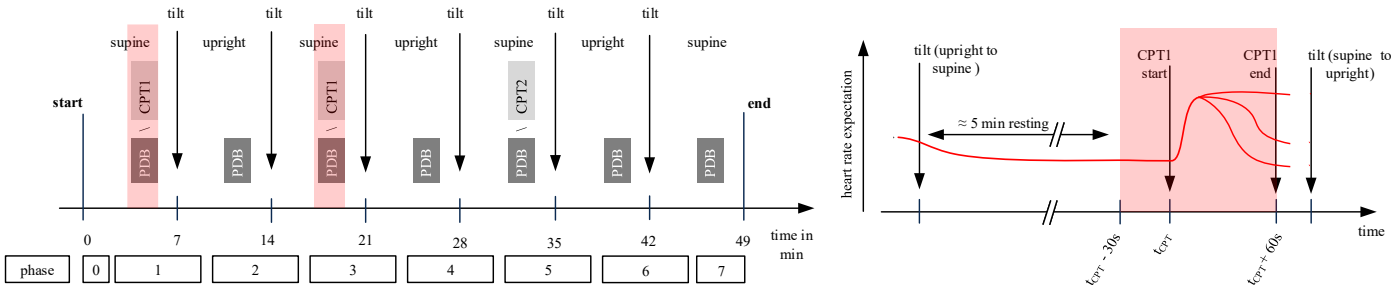


Fig. 1. Overview on the experimental protocol and used data (highlighted in red). Within this work, we use data excerpts of 90 s around CPT1. Left: whole experimental protocol and placing of the data excerpt. Right: data excerpt and expectation on the HR, which is assumed to show intrasubject and intersubject variations upon cold stress.

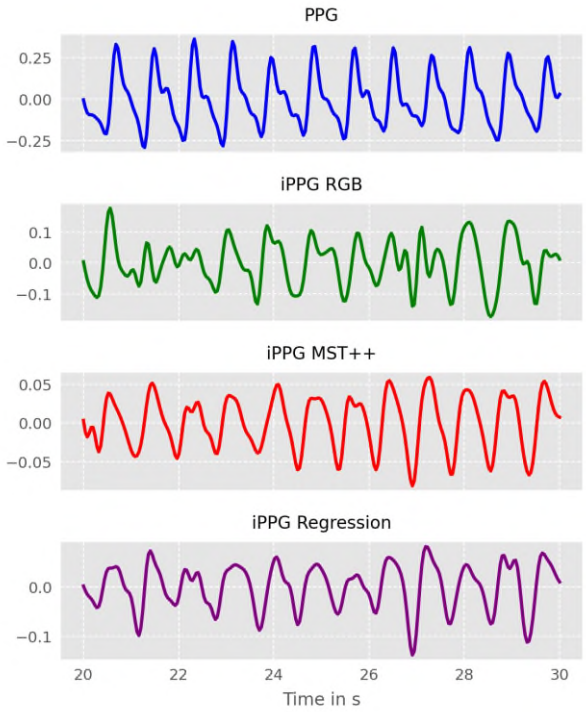
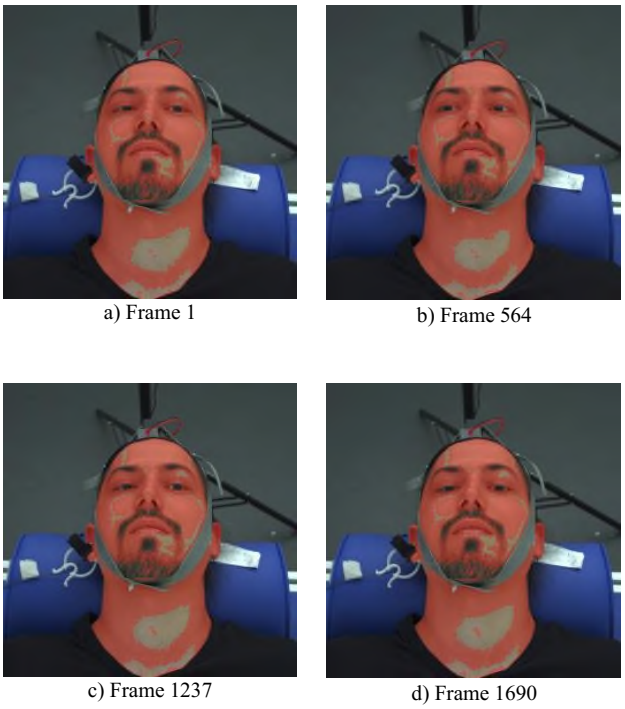


Fig. 2. Example PPG and iPPG of subject015



a) Frame 1 b) Frame 564

c) Frame 1237 d) Frame 1690

Fig. 3. Examplary images of the iPPG data. The red area reflects the ROI used to derive the iPPGsignal by averaging over the whole ROI

5) Heart Rate Estimation

HR is estimated in windows of 10 s by simple peak detection in the frequency domain. Thereto, we calculate from our input data (native or reconstructed green channel) spectra using Fast Fourier Transformation for windows of 10 s. HR is identified by the highest peak in the range between 40 and 180 bpm. The window displacement was set to 1 s resulting in a HR estimate for every second. As reference for the HR estimation from videos, we extract a heart rate from the ECG, which is considered as ground truth.

B. Employed Hyperspectral Reconstruction

1) Overview

For hyperspectral reconstruction we use a deep learning approach and a regression-based approach. The following chapter provides a description of the deep neural network utilized for hyperspectral reconstruction along with an explanation of the training strategy. The regression-based approach is less complex (has less parameters and is not at risk of overfitting) and serves as a reference for hyperspectral reconstruction.

2) Hyperspectral Reconstruction by MST++

MST++ [19] is a multi-stage transformer model for hyperspectral reconstruction from RGB images. The model employs transformer blocks to capture both spatial and spectral information, thereby ensuring that the reconstruction is gradually refined in several stages. Furthermore, MST++ utilizes positional coding to place pixels in the context of the image and residual connections to retain information from earlier stages. A comprehensive account of the model's architectural nuances can be found in the paper by Cai et al. [19].

The MST++ was trained with the Hyper-Skin database over 200 epochs and a batch size of 20. All RGB images are linearly rescaled to the range [0, 1] and the HIS samples randomly cropped to 128×128 pixels. We implement four training sessions with different seeds. Furthermore, a cosine annealing scheduler was utilized and the training objective metric is the mean relative absolute error (MRAE). After training, the weights with the lowest loss were selected for hyperspectral reconstruction of the iPPG database. To validate the MST++ during the training phase, three subjects (p049, p050 and p051) were selected as validation data. To evaluate the performance of the MST++ on previously unseen data, four subjects (p006, p021, p037 and p042) were randomly selected as test data.

3) Hyperspectral Reconstruction by Polynomial Regression

To compare the results of the MST++ with a less complex hyperspectral reconstruction, we employ a regression-based approach. Lin and Finlayson [20] compared different regression methods for hyperspectral reconstruction in their work and according to their findings, the most promising results can be achieved using polynomial regression. Polynomial regression represents an extension of linear regression, wherein the model describes a curve represented by a higher-order polynomial. The objective of polynomial regression is to construct a model that describes the relationship between independent variables x and dependent variables y by raising the input values to powers.

Because of the computational limits, the training data was downsampled to 256×256. To evaluate the performance of the polynomial regression on previously unseen data, also four subjects (p006, p021, p037 and p042) were selected as test data.

4) Quality Assessment

Quality assessment relates to both, the performance of hyperspectral reconstruction as well as the reconstructions' performance to retain BVP and extract HR. In order to evaluate the spatial quality of hyperspectral reconstruction, we employ the MSE and Mean relative absolute error (MRAE). The MSE is often used for the training of neural networks due to the advantageous scaling of the error in relation to the optimization process. Conversely, the MRAE is a widely utilized metric in the field of hyperspectral reconstruction due to its capacity to account for varying illumination conditions. Another crucial element is the spectral error. To address this, we employ the SAM, which calculates the angles of pixels to describe a spectral orientation. Subsequently, it compares the angles of the reconstructed images with those of the ground truth. To assess the performance to retain BVP and extract HR, we employ the Pearson correlation coefficient (PCC) in two ways. First, we compare the reconstructed green channels to the original green channel and the contact PPG by means of PCC. Secondly, we compare the extracted HR (from native and reconstructed green channels) to the reference heart rate by means of PCC. In addition, we employ Bland-Altman-Plots together with mean errors and limits of agreements to assess HR estimation accuracy as well as MSE and MAE to compare the extracted heart rates to the reference heart rates.

IV. Results

A. Reconstruction Performance

The reconstruction results on the test data of the Hyper-Skin database are presented in Table 1.

TABLE I. RESULTS OF THE HYPERSPECTRAL RECONSTRUCTIONS USING THE TEST SUBJECTS

Method\Metric	**MSE**	**MRAE**	**SAM**
MST++	0.0008	0.5288	0.1334
Regression	0.0009	0.8232	0.1557

B. BVP Preservation

Table 2 shows the mean PCCs of all ten seconds segments between the BVP signals of all subjects. The results indicate distinct differences in the preservation of BVP.

TABLE II. MEAN PCCS OF ALL SUBJECTS BVP SIGNALS

Methods	**PPG**	**iPPG$_{RGB}$**	**iPPG$_{MST++}$**
PPG	1	-	-
iPPG$_{RGB}$	0.422	1	-
iPPG$_{MST++}$	0.463	**0.898**	1
iPPG$_{Regression}$	**0.647**	0.724	0.780

Figures 3, 4, and 5 show the Bland-Altman plots and scatter plots for HR estimation. Table 2 shows the MSE and MAE between the reference HR and the estimated HR for the different iPPGs. Table 3 summarizes the resulting correlations for heart rate estimation.

TABLE III. MAE AND MSE OF ESTIMATED HR: REFERENCE VS. GREEN CHANNEL

Method	**MAE (STD)**	**MSE (STD)**
RGB	2.649 (1.786)	41.509 (40.330)
MST++	1.647 (1.282)	**12.540 (13.001)**
Regression	**1.570 (1.220)**	16.370 (24.340)

TABLE III. PCCS OF ESTIMATED HR: REFERENCE VS. GREEN CHANNEL

Method	**PCC**
RGB	0.81
MST++	0.93
Polynomial Regression	0.92

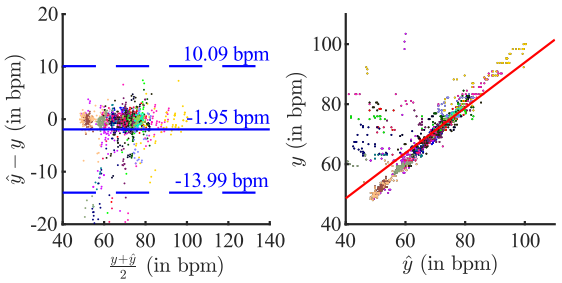


Fig. 4. Bland-Altmann Plot and regression of PPG and iPPG of RGB

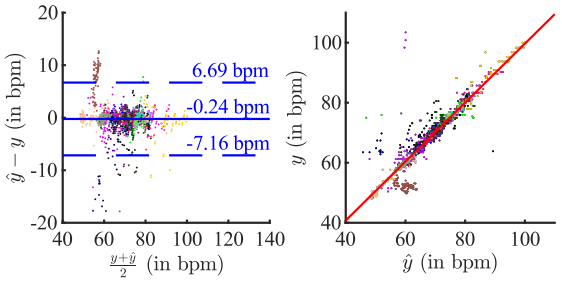


Fig. 5. Bland-Altman Plot and regression of PPG vs iPPG of MST++

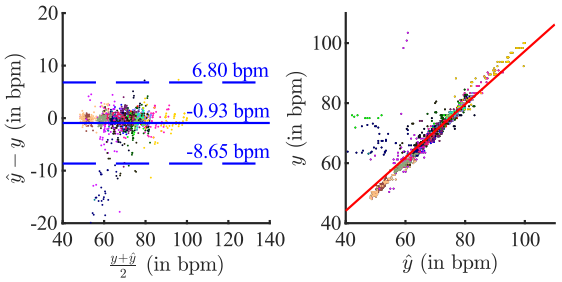


Fig. 6. Bland-Atlman Plot of regression of PPG vs iPPG of Polynomial Regression

V. Discussion and conclusion

A. Discussion

As evidenced in Table 1, the MST++ demonstrates superior quantitative performance compared to polynomial regression. This outcome was anticipated, as neural networks typically demonstrate superior performance in hyperspectral reconstruction, exhibiting enhanced adaptability to specific databases. Furthermore, as shown by Tables 2 to 4 and Figures 4 to 6, hyperspectral reconstruction using polynomial regression and MST++ can retain the BVP. As our results indicate, the reconstruction (and synthesizing a green channel again) can even improve extraction of BVP and HR. However, such finding is true for the current data but might not be generalizable (future investigation will have to confirm). Anyways, hyperspectral reconstruction is intented to be a first step and our future works will direct at methods to combine spectroscopic information thus actually achieving generizable improvements.

B. Limitations

Our study is a preliminary study and there are some limitations. The objective was to assess the impact of hyperspectral reconstruction. The method employed for HR extraction was not optimized, as this would have required the use of alternative techniques for color channel combination, which are known to improve HR detection compared to using the green channel only. In addition, our findings are true for the employed setting (cameras, illumination, collective, ...). While the general statement on the usability of hyperspectral reconstruction, particularly using MST++, is valid, particularly with respect to quantitative results, other settings might differ and more work towards generizable statements are required.

C. Conclusions

Spectroscopic information bears wide information on physiological parameters. Hyperspectral reconstruction is an innovative approach, to combine the advantages of hyperspectral imaging and conventional RGB videos. Our preliminary analysis indicates the potential of hyperspectral reconstruction.

References

[1] N. Molinaro, E. Schena, S. Silvestri, F. Bonotti, and D. e. a. Aguzzi, "Contactless Vital Signs Monitoring From Videos Recorded With Digital Cameras: An Overview," (in eng), *Frontiers in physiology*, vol. 13, p. 801709, 2022, doi: 10.3389/fphys.2022.801709.

[2] V. Selvaraju, N. Spicher, J. Wang, N. Ganapathy, J. M. Warnecke, and et al., "Continuous Monitoring of Vital Signs Using Cameras: A Systematic Review," (in eng), *Sensors (Basel, Switzerland)*, vol. 22, no. 11, 2022, doi: 10.3390/s22114097.

[3] S. Zaunseder, A. Trumpp, H. Ernst, M. Förster, and H. Malberg, "Spatio-temporal analysis of blood perfusion by imaging photoplethysmography," in *Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics*, San Francisco, United States, Jan. 2018 - Feb. 2018, p. 32.

[4] S. Zaunseder and S. Rasche, "Clinical applications for imaging photoplethysmography," in *Contactless Vital Signs Monitoring*: Elsevier, 2022, pp. 149–164.

[5] W. Wang, S. Stuijk, and G. de Haan, "Exploiting spatial redundancy of image sensor for motion robust rPPG," (in eng), *IEEE transactions on bio-medical engineering*, vol. 62, no. 2, pp. 415–425, 2015, doi: 10.1109/TBME.2014.2356291.

[6] W. Chen, Z. Yi, L. J. R. Lim, R. Q. R. Lim, A. Zhang, and et al., "Deep learning and remote photoplethysmography powered advancements in contactless physiological measurement," (in eng), *Frontiers in bioengineering and biotechnology*, vol. 12, p. 1420100, 2024, doi: 10.3389/fbioe.2024.1420100.

[7] A. Ni, A. Azarang, and N. Kehtarnavaz, "A Review of Deep Learning-Based Contactless Heart Rate Measurement Methods," *Sensors*, vol. 21, no. 11, p. 3719, 2021, doi: 10.3390/s21113719.

[8] M. Dietrich, S. Marx, M. von der Forst, T. Bruckner, F. Schmitt, and et al., "Bedside hyperspectral imaging indicates a microcirculatory sepsis pattern - an observational study," *Microvascular Research*, vol. 136, p. 104164, 2021, doi: 10.1016/j.mvr.2021.104164.

[9] A. Mahmoud and Y. H. El-Sharkawy, "Quantitative phase analysis and hyperspectral imaging for the automatic identification of veins and blood perfusion maps," (in eng), *Photodiagnosis and photodynamic therapy*, vol. 42, p. 103307, 2023, doi: 10.1016/j.pdpdt.2023.103307.

[10] J. Zhang, R. Su, Q. Fu, W. Ren, F. Heide, and Y. Nie, "A survey on computational spectral reconstruction methods from RGB to hyperspectral imaging," *Sci Rep*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-16223-1.

[11] J. He *et al.,* "Spectral super-resolution meets deep learning: Achievements and challenges," *Information Fusion*, vol. 97, p. 101812, 2023, doi: 10.1016/j.inffus.2023.101812.

[12] P. C. Ng, Z. Chi, Y. Verdie, J. Lu, and K. N. Plataniotis, "Hyper-Skin: A Hyperspectral Dataset for Reconstructing Facial Skin-Spectra from RGB Images," 2023.

[13] V. Fleischhauer, J. Bruhn, S. Rasche, and S. Zaunseder, "Photoplethysmography upon cold stress-impact of measurement site and acquisition mode," (in eng), *Frontiers in physiology*, vol. 14, p. 1127624, 2023, doi: 10.3389/fphys.2023.1127624.

[14] B. Arad, R. Timofte, O. Ben-Shahar, Y.-T. Lin, G. Finlayson, and et al., "NTIRE 2020 Challenge on Spectral Reconstruction from an RGB Image," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, 2020, pp. 1806–1822.

[15] B. Arad, R. Timofte, R. Yahel, N. Morag, A. Bernat, and et al., "NTIRE 2022 Spectral Recovery Challenge and Data Set," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, New Orleans, LA, USA, 2022, pp. 862–880.

[16] A. Woyczyk, V. Fleischhauer, and S. Zaunseder, "Adaptive Gaussian Mixture Model Driven Level Set Segmentation for Remote Pulse Rate Detection," (in eng), *IEEE journal of biomedical and health informatics*, vol. 25, no. 5, pp. 1361–1372, 2021, doi: 10.1109/JBHI.2021.3054779.

[17] OpenCV team, *OpenCV 4.1.1 Documentation.* [Online]. Available: https://docs.opencv.org/4.1.1/index.html

[18] T. Chan and L. Vese, "An Active Contour Model without Edges," in *Lecture Notes in Computer Science, Scale-Space Theories in Computer Vision*, G. Goos, J. Hartmanis, J. van Leeuwen, M. Nielsen, P. Johansen, and et al., Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 141–151.

[19] Y. Cai, J. Lin, Z. Lin, H. Wang, Y. Zhang, and et al., "MST++: Multi-stage Spectral-wise Transformer for Efficient Spectral Reconstruction," 2022, doi: 10.48550/arXiv.2204.07908.

[20] Y.-T. Lin and G. D. Finlayson, "On the Optimization of Regression-Based Spectral Reconstruction," (in eng), *Sensors (Basel, Switzerland)*, vol. 21, no. 16, 2021, doi: 10.3390/s21165586.