

Dieses Dokument ist eine Zweitveröffentlichung (Postprint) /

This is a self-archiving document (accepted version):

Hannes Ernst, Matthieu Scherpf, Hagen Malberg, Martin Schmidt

Optimal color channel combination across skin tones for remote heart rate measurement in camera-based photoplethysmography

Erstveröffentlichung in / First published in:

Biomedical Signal Processing and Control. 2021, 68, Art.-Nr. 102644 [Zugriff am: 16.09.2021]. Elsevier. ISSN 1746-8108.

DOI: <https://doi.org/10.1016/j.bspc.2021.102644>

Diese Version ist verfügbar / This version is available on:

<https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-806683>



Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung – Nicht kommerziell – Keine Bearbeitungen 4.0 International Lizenz](https://creativecommons.org/licenses/by-nc-nd/4.0/).

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Optimal color channel combination across skin tones for remote heart rate measurement in camera-based photoplethysmography

Hannes Ernst^{a, b, *}, Matthieu Scherpf^b, Hagen Malberg^b and Martin Schmidt^b

^a Research Training Group Conducive Design of Cyber-Physical Production Systems, TU Dresden, Dresden, Germany

^b Institute of Biomedical Engineering, TU Dresden, Dresden, Germany

ARTICLE INFO

Keywords

imaging photoplethysmography
color channel combination
color space transformation
heart rate
region of interest

ABSTRACT

Objective: The heart rate is an essential vital sign that can be measured remotely with camera-based photoplethysmography (cbPPG). Systems for cbPPG typically use cameras that deliver red, green, and blue (RGB) channels. The combination of these channels has been proven to increase signal-to-noise ratio (SNR) and heart rate measurement accuracy (ACC). However, many combinations remain untested, the comparison of proposed combinations on large datasets is insufficiently investigated, and the interplay with skin tone is rarely addressed. **Methods:** Eight regions of interest and eight color spaces with a total of 25 color channels were compared in terms of ACC and SNR based on the Binghamton-Pittsburgh-RPI Multimodal Spontaneous Emotion Database (BP4D+). Additionally, two systematic grid searches were performed to evaluate ACC in the space of linear combinations of the RGB channels. **Results:** Glabella and forehead regions of interest provided highest ACC (up to 74.1 %) and SNR (> -3 dB) with the hue channel H from HSV color space and the chrominance channel Q from NTSC color space. The grid searches revealed a global optimum of linear RGB combinations (ACC: 79.2 %). This optimum occurred for all skin tones, although ACC dropped for darker skin tones. **Conclusion:** Through systematic grid searches we were able to identify the skin tone independent optimal linear RGB color combination for measuring heart rate with cbPPG. Our results proved on a large dataset that the identified optimum outperformed conventionally used color channels. **Significance:** The presented findings provide useful evidence for future considerations of algorithmic approaches for cbPPG.

1. Introduction

Vital sign monitoring is an essential part of diagnostics and clinical health surveillance. Over the years its application expanded to other fields, such as telemedical home usage [1], fitness tracking [2], and occupational protection [3]. In some of these applications, requirements arise which contact measurement methods can only meet to a limited extent, e.g. operation freedom which is hindered by cables, ease of use which is hindered by fixating sensors or electrodes and connecting hardware, and reusability as well as low disinfection effort which are hindered by subject contact. In these cases, contactless vital sign monitoring methods provide suitable alternatives. [4,5]

Camera-based photoplethysmography (cbPPG), also called imaging or remote photoplethysmography, is a non-contact optical measurement technique that uses cameras to assess cardiovascular vital signs such as heart rate (HR). Blood volume pulsation in the vessels and ballistocardiographic effects lead to pulsatile alterations of optical parameters in the skin tissue and therefore modulate the number of reflected photons reaching the camera.

The skin tissue is illuminated either with ambient light or with dedicated light sources. While conventional photoplethysmography utilizes a photo detector directly attached to the skin to record an intensity time series, cbPPG works remotely and additionally provides spatial resolution, i.e. the camera records videos. [6]

Most cbPPG systems include cameras that deliver a red (R), green (G), and blue (B) channel for each pixel, so-called RGB cameras [7]. The intensity time series of the RGB channels differ due to 1) the spectral composition of the illumination, 2) wavelength dependent optical parameters of skin tissue and blood, and 3) the camera sensor specific spectral sensitivity characteristics [8]. It has been shown that cbPPG benefits from the combination of RGB channels [6]. However, many of these combinations remain untested and the interplay with skin tone is rarely addressed. The scope of this work is to identify beneficial color combinations for camera-based HR measurement across all skin tones. In order to acquire reliable results, our investigation is based on a large dataset.

* Corresponding author. E-mail address: hannes.ernst@tu-dresden.de (Hannes Ernst).

2. Related work

It is well known that the G channel provides the best cbPPG signal in RGB color space [6]. The absorption peak of hemoglobin in the green spectral range leads to strong pulsation of the signal and thus to a higher signal-to-noise ratio (SNR) compared to the red and blue channel [9]. Also, green photons penetrate deeper into the skin than blue photons [10] so that pulsating vessels, i.e. upper dermal arterioles, are better reached [11]. Finally, the Bayer filter mosaic of standard RGB cameras provides twice as many green sensor cells as red or blue ones, which increases SNR of the G channel due to reduced impact of sensor noise after spatial averaging.

However, multiple comparative studies [12–15] show that the hue (H) channel outperforms the G channel. The H channel represents the color information in a single scalar that can be interpreted as the angle of a spectral color wheel. Several color spaces (e.g. HSL or HSV) decouple this color information from other appearance parameters such as saturation and lightness by a nonlinear RGB color transformation [13]. This reduces the influence of unsteady lighting conditions and small motion artifacts on the cbPPG signal [13].

Apart from the frequently utilized G and H channels, other color channels derived from the RGB color space are occasionally used. Recent findings indicate that the chrominance channel Q from NTSC color space yields similar accuracy (ACC) in camera-based HR measurement as the H channel [15]. Other color spaces applied in the context of cbPPG are for example CIE XYZ [13], CIE L*a*b* [16], and CIE L*u*v* [17].

It has been reported that skin pigmentation influences the cbPPG measurement [6]. According to the Fitzpatrick scale, skin types can be classified into six groups with skin tones ranging from pale white (I) to dark black (VI) [18]. Higher epidermal melanin concentration of darker skin increases the constant photon absorption in skin tissue. Thus, fewer photons contribute to the pulsatile signal component, which diminishes the cbPPG signal amplitude. Consequently, darker skin tones are associated with lower signal intensity and reduced signal quality (e.g. SNR) [19]. In [19], it is stated that for white light skin tones can be normalized by

$$[R_s, G_s, B_s] = [R, G, B] / \sqrt{R^2 + G^2 + B^2} \quad (1)$$

leading to a standardized skin color model $[R_s, G_s, B_s] = [0.7682, 0.5121, 0.3841]$ for all skin tones. The spectral absorption characteristics of hemoglobin are not altered by the melanin concentration of skin tissue, which implies that the pulsatile component of the cbPPG signal can be extracted for all

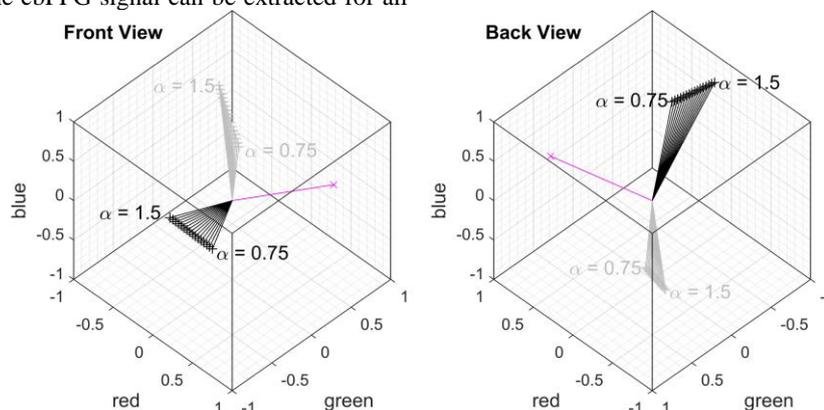


Fig. 1. Visualization of CHROM (black) [19] and POS (gray) [8] color combinations in RGB color space for different $\alpha \in \{0.75, 0.8 \dots 1.5\}$. The visualized combinations are normalized. The pink vector represents the standardized skin color model $[R_s, G_s, B_s] = [0.7682, 0.5121, 0.3841]$ from [19].

skin tones with the same color combination (stable illumination presumed) [20].

CHROM [19] and POS [8], two algorithms for cbPPG extraction, build on skin color information. In CHROM, two orthogonal chrominance signals are used to cancel out specular reflections (induced e.g. by motion artifacts). To correct for non-white illumination, the approach implements skin tone standardization and α -amplitude tuning, which leads to a time dynamic color combination [19]. POS follows the same idea, but projects orthogonal to the skin tone [8]. Table 1 summarizes the key elements of color combination of CHROM and POS, and Fig. 1 visualizes CHROM and POS color combinations in the RGB color space. For further details we refer to the original publications (CHROM: $X_{smin}\alpha Y_s$ in [19], POS: [8]). Since time dynamic color combinations change the shape of the signal as it progresses, they eliminate low-frequency components (e.g. from breathing, vasoconstriction or vasodilation) and can distort investigations of cbPPG signal morphology (e.g. pulse amplitudes). It is claimed that CHROM and POS work for all skin tones [8,19]. However, the results of CHROM are not reported by skin tone but only in total. Zaunseder *et al.* [6] note that the skin tone validation of POS bases only on a small database (15 subjects with a total recording duration of less than 19 min at 20 fps). The need for grouping the test subjects according to their skin tone further limits the significance of results from such small databases.

Table 1. Comparison of CHROM [19] and POS [8] color combinations.

Variable	CHROM	POS
X	$3R_n - 2G_n$	$G_n - B_n$
Y	$1.5R_n + G_n - 1.5B_n$	$-2R_n + G_n + B_n$
α	$std(X_f)/std(Y_f)$	$std(X)/std(Y)$
$S(\alpha)$	$X_f - \alpha Y_f$ $= 3(1 - \alpha/2)R_f$ $-2(1 + \alpha/2)G_f$ $+3(\alpha/2)B_f$	$X + \alpha Y$ $= -2\alpha R_n$ $+(1 + \alpha)G_n$ $-(1 - \alpha)B_n$
$S(0.75)$	$1.875R_f - 2.75G_f + 1.125B_f$	$-1.5R_n + 1.75G_n - 0.25B_n$
$S(1)$	$1.5R_f - 3G_f + 1.5B_f$	$-2R_n + 2G_n$
$S(1.25)$	$1.125R_f - 3.25G_f + 1.875B_f$	$-2.5R_n + 2.25G_n + 0.25B_n$

X and Y denote the chrominance signals that are calculated from the signal channels red R , green G , and blue B . std is the standard deviation. The subscript n marks temporally normalized signals. The subscript f marks temporally normalized signals that are band-pass filtered (only CHROM). The parameter α corrects for illumination changes and is recalculated every 1.6 s. $S(\alpha)$ denotes the cbPPG signal derived from the chrominance signals X and Y . Both algorithms work with sliding window overlap adding of $S(\alpha)$ to merge consecutive signals with different α . Three examples with $\alpha = 0.75$, $\alpha = 1$, and $\alpha = 1.25$ are presented for each algorithm.

3. Investigation goals

This work consists of three consecutive studies. In order to provide a clear understanding of the research objectives of the three studies, they will be presented individually.

Study #1: Our previous work presented in [15] reports that the ACC of the Q channel is comparable to the ACC of the H channel and bases on the results from 33 out of 140 subjects of the Binghamton-Pittsburgh-RPI Multimodal Spontaneous Emotion Database (BP4D+). To validate our findings, the analysis (see Fig. 2) was extended to the complete BP4D+ with all 140 subjects. Objective of study #1 is the comparison of camera-based HR measurement ACC and SNR in eight regions of interest (ROIs, see Fig. 3) and eight color spaces with a total of 25 color channels (see legend of Fig. 4).

Study #2: In [15], the Q channel from NTSC color space reached comparable ACC as the H channel from HSV color space. The Q channel is a linear combination of the RGB channels [21]:

$$Q = 0.211R - 0.523G + 0.312B. \quad (2)$$

Since this combination is rather arbitrary for the application in cbPPG, further analysis proves if an even better linear combination of RGB color channels exists. Therefore, a grid search was set up to systematically evaluate the camera-based HR measurement ACC of different linear RGB combinations. Objective of study #2 is the identification of the optimal linear RGB combination that provides the highest ACC.

Study #3: As indicated earlier, a standardized skin color model [19] and the independence of color combinations from skin tone [20] have been postulated. Therefore, the RGB compositions of all Fitzpatrick skin tones were investigated. Also, another grid search was set up, but this time for each skin tone separately. The goal is to compare the performance of RGB combinations across skin tones.

In summary, by investigating ACC and SNR of the Q and H channels in a more comprehensive dataset (study #1), we sought to verify previous findings about the impact of different ROIs and color spaces on camera-based HR measurement. Proceeding from these results, a novel grid search approach was implemented to determine an optimal linear RGB color combination to further improve the ACC of camera-based HR measurements (study #2). To analyze the influence of the skin tone, our grid search approach was structurally refined and extended by a skin tone dependent identification of the optimal RGB combination (study #3).

4. Materials and methods

4.1. Database

The Binghamton-Pittsburgh-RPI Multimodal Spontaneous Emotion Database (BP4D+) [22] consists of 140 subjects (82 female, 58 male). The subjects vary in skin tone (see Table 2) and declared to be healthy. For each subject 10 videos (1392x1040 pixels, 25 fps, 3x8 bit RGB) under different emotional stimuli (happiness, surprise, sadness, startle, skepticism, embarrassment, fear, physical pain, anger, and disgust) are provided leading to 1400 videos of different length. This represents an extraordinarily large cohort for state-of-the-art cbPPG research [7]. The BP4D+ also delivers facial landmark coordinates for each video frame and a synchronized heart rate reference signal (1000 Hz) derived from inter-beat intervals. The spectral composition of the illumination and the sensitivity curves of the camera sensor chip are assumed to remain constant throughout the recordings.

Table 2. Subject and segment distribution of BP4D+ [22] by skin tone.

Skin Tone	I	II	III	IV	V	VI	Total
Subjects	16	44	45	19	5	11	140
in %	11.4	31.4	32.1	13.6	3.6	7.9	100
Segments	787	2617	2735	1111	309	648	8207
in %	9.6	31.9	33.3	13.5	3.8	7.9	100

Skin tones were annotated manually following the Fitzpatrick scale [18].

4.2. Study #1: Comparison of heart rate measurement accuracy in multiple color channels and regions of interest

Aim of study #1 was to validate the investigation of 33 female subjects presented in [15] by extending our analysis to all 140 subjects of BP4D+. Fig. 2 summarizes the workflow of study #1.

4.2.1. Regions of interest and color spaces

Seven ROIs from the facial landmarks and one ROI from the Level-Set algorithm [23] were computed similar to [15]. The landmark ROIs focused on forehead and cheeks as these regions are known to provide the best signal quality [6]. Since the orientation towards the camera and the supplying vessels of the glabella are similar to the forehead ROI [24], the glabella area was also included. For details on the generation of the Level-Set ROI, see [23]. Fig. 3 gives an example of all ROIs.

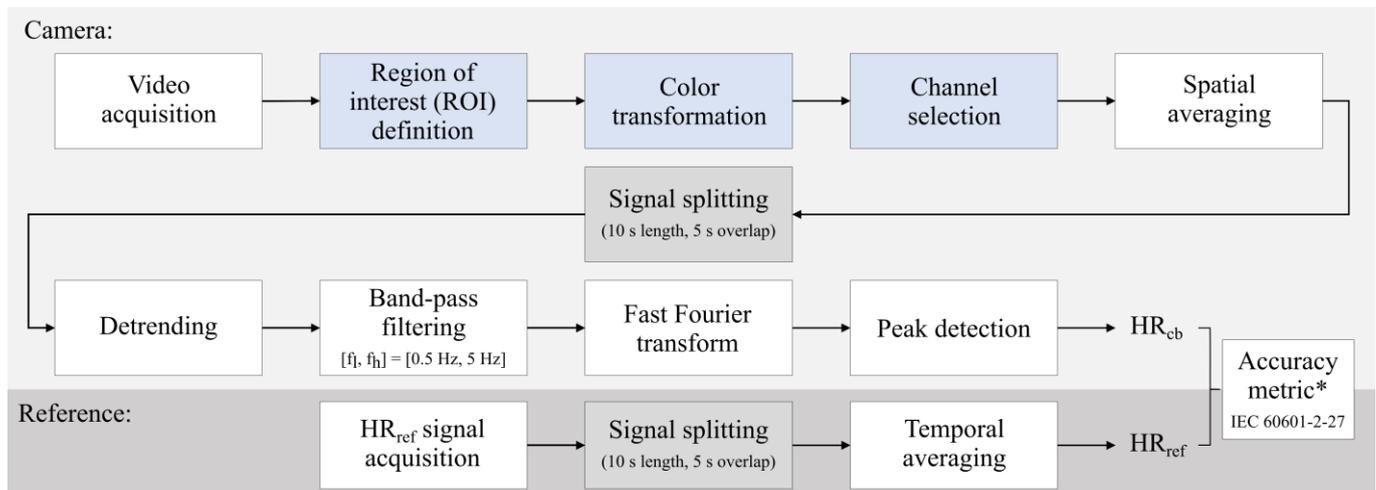


Fig. 2. Workflow of camera-based heart rate (HR_{cb}) measurement and BP4D+ reference heart rate (HR_{ref}) derivation. The steps highlighted in blue constitute the core research topic of this work. * HR_{cb} was accepted as correct if the condition in (3) was fulfilled.

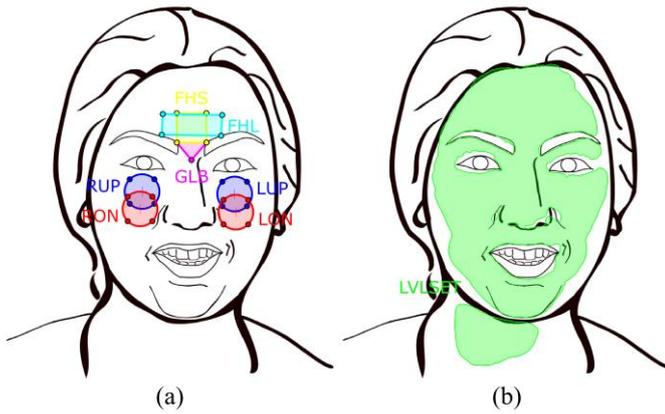


Fig. 3. Regions of interest included in the comparison. (a) GLB: Glabella, FHS: Forehead small, FHL: Forehead large, LUP/RUP: Upper left/right cheek, LON/RON: On left/right cheek. (b) LVLSET: Level-Set. Head contours were re-painted from a BP4D+ image [22].

Eight color spaces that are used to varying degrees in the literature were selected: RGB, HSV, YCbCr, NTSC (also known as YIQ), CIE XYZ, CIE $L^*a^*b^*$, CIE $L^*u^*v^*$, and CMYK. They provide a total of 25 color channels (see legend of Fig. 4). For each frame and ROI, the RGB values of the corresponding pixels were transformed into the aforementioned color spaces.

4.2.2. Signal processing

Frame-wise averaging of the transformed pixel values yielded a raw cbPPG signal for each color channel and ROI. The raw cbPPG signals were divided into segments of 10 s length with an overlap of 5 s. This resulted in 8207 segments per color channel, whose distribution by skin tone is presented in Table 2. Each segment i was detrended and band-pass filtered (10th order Butterworth with cutoff frequencies 0.5 Hz and 5 Hz) to diminish the impact of unwanted components such as breathing or noise. The i^{th} filtered cbPPG signal segment is called $S_{f,i}$. A fast Fourier transform mapped each zero-padded segment to the frequency domain leading to a frequency spectrum $X(S_{f,i})$ with a spectral resolution of 98 mHz (5.9 bpm) for each segment in each color channel and each ROI.

4.2.3. Heart rate measurement and evaluation

The frequency of the dominant peak in the amplitude spectrum $|X(S_{f,i})|$ was associated with the camera-based heart rate $HR_{cb,i}$. Corresponding to [25] and in accordance with the precision requirement for medical ECG defined in IEC 60601-2-27, all $HR_{cb,i}$ that fulfilled the criterion

$$|HR_{cb,i} - HR_{ref,i}| \leq \max(5 \text{ bpm}, 0.1 \cdot HR_{ref,i}) \quad (3)$$

were accepted as correct. The proportion of correct camera-based HR yielded ACC. To acquire the reference heart rate $HR_{ref,i}$ for each $HR_{cb,i}$, the reference signal was split into intervals identical to the filtered signal (10 s length, 5 s overlap). Subsequently, $HR_{ref,i}$ was acquired as the mean of the i^{th} interval of the heart rate reference signal.

The second evaluation parameter was SNR which was calculated as described in [26]. SNR has been found to indicate signal quality and imply the correctness of camera-based HR measurement [27].

4.3. Study #2: Cube grid search for optimal static linear combination in RGB color space

The goal of study #2 was to systematically examine linear RGB combinations regarding their suitability for cbPPG. Therefore, a

grid search was performed. Three color weighting coefficients c_r , c_g , and c_b for the channels R, G, and B respectively define the grid axes. The value of each color weighting coefficient varied from -1 to +1. The grid granularity, i.e. the step size of the color weighting coefficients, was set to 0.01. This resulted in a cube grid with 201 points on each of the three axis and $8'120'601$ points in total.

The grid search was performed only for the in study #1 identified most suitable ROI (see FHL in Fig. 3). For each grid point, the raw cbPPG signals (see Section 4.2.2) from the R, G, and B channels of a video were linearly combined to a signal S_c :

$$S_c = c_r R + c_g G + c_b B. \quad (4)$$

The combined signals S_c were then processed further as described in Section 4.2.2 and Section 4.2.3 starting with the generation of 10 s segments (signal splitting in Fig. 2). In this manner, an ACC was determined for each point of the cube grid, i.e. for each color combination defined by c_r , c_g , and c_b . The computation was vectorized and parallelized in order to process the full BP4D+ more than 8 million times in a reasonable amount of time.

For the results, two properties of the cube grid were to be expected:

1. A point symmetry with respect to the origin of the grid since

$$|X(S_{f,i})| = |X(-S_{f,i})|. \quad (5)$$

2. Identical results along lines through the origin due to

$$|a| \cdot |X(S_{f,i})| = |X(a \cdot S_{f,i})|. \quad (6)$$

This includes the point symmetry of (5) as the special case $a = -1$.

The three-dimensionality of the cube grid impedes a simple visualization of the results. For this reason, the evaluation parameters were visualized multiple times by sequentially removing poor results (unfavorable RGB combinations) via ACC threshold filtering.

4.4. Study #3: Skin tone investigation and hemispherical surface grid search for optimal static linear combination in RGB color space

Study #3 targeted at the influence of the skin tone: 1) The RGB distribution of the skin pixels and the standardized skin color model were examined, and 2) another grid search was set up for each skin tone individually.

At first, the color distributions of the skin tones were investigated. For the first frame of each 10 s segment, all pixels of the in study #1 identified most suitable ROI (see FHL in Fig. 3) were extracted and summarized by skin tone in RGB histograms (256 bins for 8 bit resolution). Additionally, skin pixels were normalized according to (1) and also plotted by skin tone in RGB histograms (256 bins).

To find out whether the skin tone influences the optimal linear RGB combination, another grid search was set up. Due to the expected properties of the cube grid described in Section 4.3, the grid points for study #3 were placed on a hemispherical surface with radius 1. This requires fewer grid points as the calculation of the inner structure becomes obsolete (identical results along each line through the grid origin) and only one half of the three-dimensional space needs to be evaluated (point symmetry with respect to grid origin). The vacated computational capacities allowed for a higher resolution of the hemispherical surface grid. The hemispherical surface grid was generated with Icosahedron-based S^2 Decomposition [28] and consisted of $1'311'745$ points.

Signal combination, processing, and evaluation for each grid point were identical with study #2 (see Section 4.3). However, in contrast to study #2 ACC was calculated separately for each skin tone.

5. Results

5.1. Study #1: Comparison of heart rate measurement accuracy in multiple color channels and regions of interest

Fig. 4 presents the resulting camera-based HR measurement ACCs of all 200 combinations of the investigated eight ROIs and 25 color channels. Forehead, glabella, and Level-Set ROIs outperformed cheek ROIs across all color channels with exception of HSV-H and NTSC-I in the Level-Set ROI. The highest ACCs

were reached in the HSV-H channel and glabella (74.1 %) or forehead (large and small both 73.4 %) ROIs. The NTSC-Q channel also yielded high ACC, especially in the Level-Set ROI (71.1 %), where HSV-H performance dropped to 25.3 %. For glabella and forehead ROIs the color channels RGB-G, HSV-H, YCbCr-Cb, NTSC-Q, XYZ-Y, $L^*a^*b^*-a^*$, $L^*u^*v^*-v^*$, and CMYK-M performed best in their color space. The results for SNR (see Fig. 5) confirm the findings on ACC, which means that color channels with high ACC also exhibited high SNR. Best median SNR (> -2 dB) were reached in the HSV-H channel of glabella and forehead ROIs, followed by the NTSC-Q channel (-3 dB $<$ median SNR < -2 dB) of the same ROIs and additionally the Level-Set ROI.

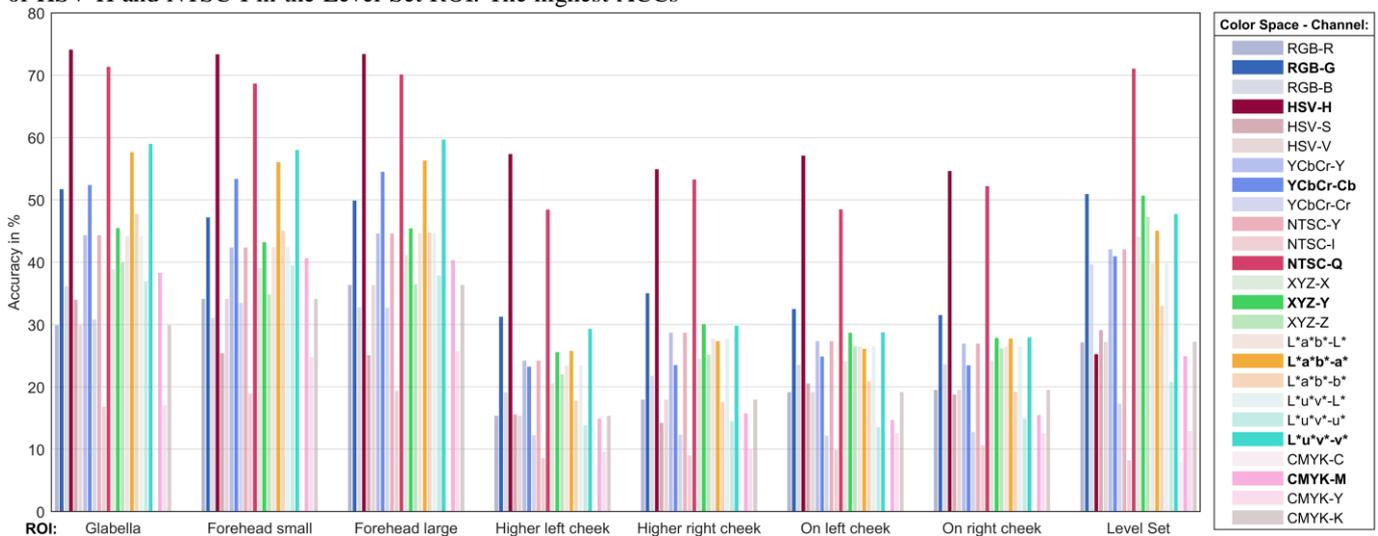


Fig. 4. Accuracy of camera-based heart rate measurement by region of interest (ROI) and color channel. Highest accuracies were reached in HSV-H channel and forehead or glabella ROIs. The NTSC-Q channel also yielded high accuracies in these ROIs and additionally in the Level-Set ROI, where the performance of the HSV-H channel dropped. Apart from the latter, the cheek ROIs generally yielded lower accuracies. For the glabella ROI and the two forehead ROIs the color channels RGB-G, HSV-H, YCbCr-Cb, NTSC-Q, XYZ-Y, $L^*a^*b^*-a^*$, $L^*u^*v^*-v^*$, and CMYK-M performed best in their color space (bold font in legend). The plot shows that ROI and color channel have to be chosen carefully for remote heart rate measurement with camera-based photoplethysmography.

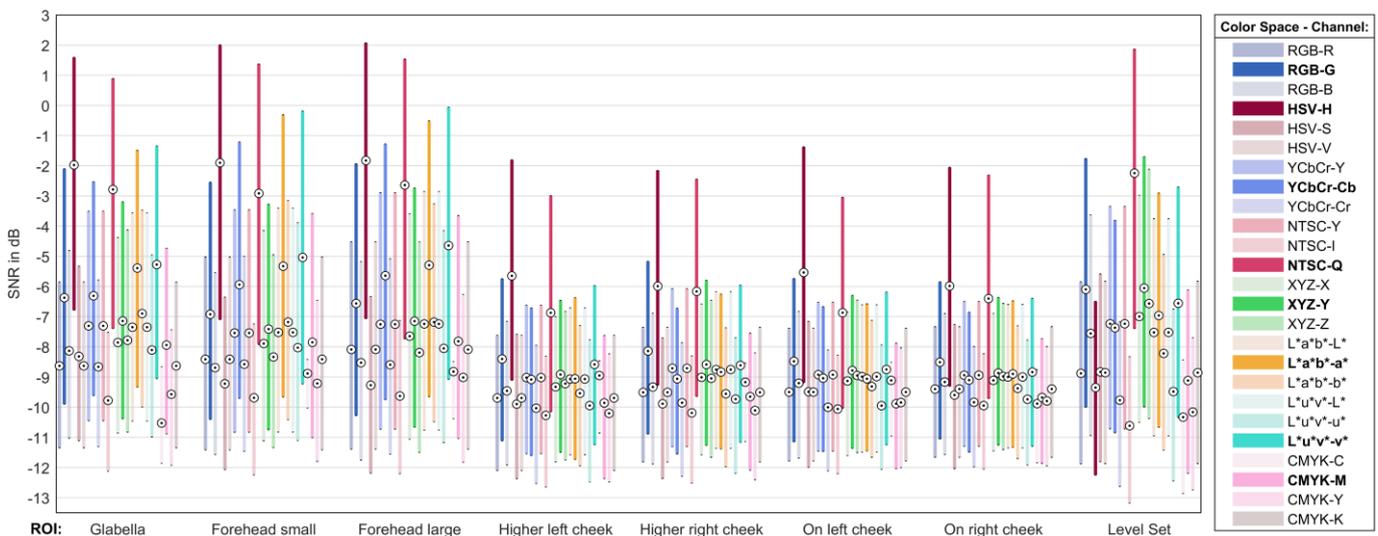


Fig. 5. Signal-to-noise ratio SNR of camera-based photoplethysmograms by region of interest (ROI) and color channel in compact boxplots (no whiskers). With exception of the Level-Set ROI, the HSV-H color channel provided the highest SNR, followed by NTSC-Q. Generally, the forehead, glabella, and Level-Set ROIs outperformed the cheek ROIs. However, SNR dropped in the HSV-H color channel of the Level-Set ROI. The SNR results confirmed the accuracy results presented in Fig. 4 and underline the importance of proper ROI and color channel selection to achieve high signal quality with camera-based photoplethysmography.

5.2. Study #2: Cube grid search for optimal static linear combination in RGB color space

Fig. 6 shows the camera-based HR measurement ACC results of the cube grid search for the large forehead ROI (FHL in Fig. 3) for different ACC thresholds and with several RGB combinations of CHROM (black) and POS (gray). The results were point-symmetric to the grid origin, i.e. ACC did not alter along the lines through the grid origin. However, gaps were observed near the origin, which became particularly apparent in the cube grid with the maximum ACC threshold (see Fig. 6). In general, the ACC converged monotonically in form of two elliptic cones. The maximum achieved ACC was 79.2 %. The line along the cone axis at maximum ACC threshold, further denoted as convergence line L_C , was determined as

$$L_C: [c_r, c_g, c_b]^T = t[0.25, -0.83, 0.5]^T, t \in \mathbb{R}. \quad (7)$$

It appeared that unfavorable RGB color combinations exist relatively close to the RGB color combinations with the highest ACC. For example at $[c_r, c_g, c_b]^T = [0.3, -1, 0.6]^T$ a strong falling ACC gradient towards the cube corner $[c_r, c_g, c_b]^T = [1, -1, 1]^T$ was observed (see Fig. 6 with no threshold).

5.3. Study #3: Skin tone investigation and hemispherical surface grid search for optimal static linear combination in RGB color space

Fig. 7 shows the RGB intensity distributions of the large forehead ROI skin pixels, while Fig. 8 shows the via (1) normalized version. Darker skin tones led to lower RGB intensities. Saturation was observed in the R channel of skin tones I and II, i.e. peaks occurred at the maximum intensity level in these histograms (see Fig. 7). Skin tone normalization unified the histogram shapes of all skin tones, even though minor variations remained. The standardized skin color model from [19] deviated slightly from the peaks of the normalized RGB histograms, particularly for darker skin tones (see Fig. 8).

Fig. 9 visualizes the ACC results of the hemispherical surface grid search for all skin tones. Across all skin tones, the highest ACC was found close to the color combination

$$S_{c,opt} = 0.25R - 0.83G + 0.5B, \quad (8)$$

which confirms L_C from (7). As shown in Table 3, the maximum ACC dropped for skin tones V and VI. Compared to the other skin tones, high ACC was achieved over a vast range of RGB combinations for skin tone II (see top view and front view in Fig. 9). In all cases, unfavorable RGB combinations occurred close to the optimal RGB combinations and manifested in the form of a band across the hemispherical surfaces.

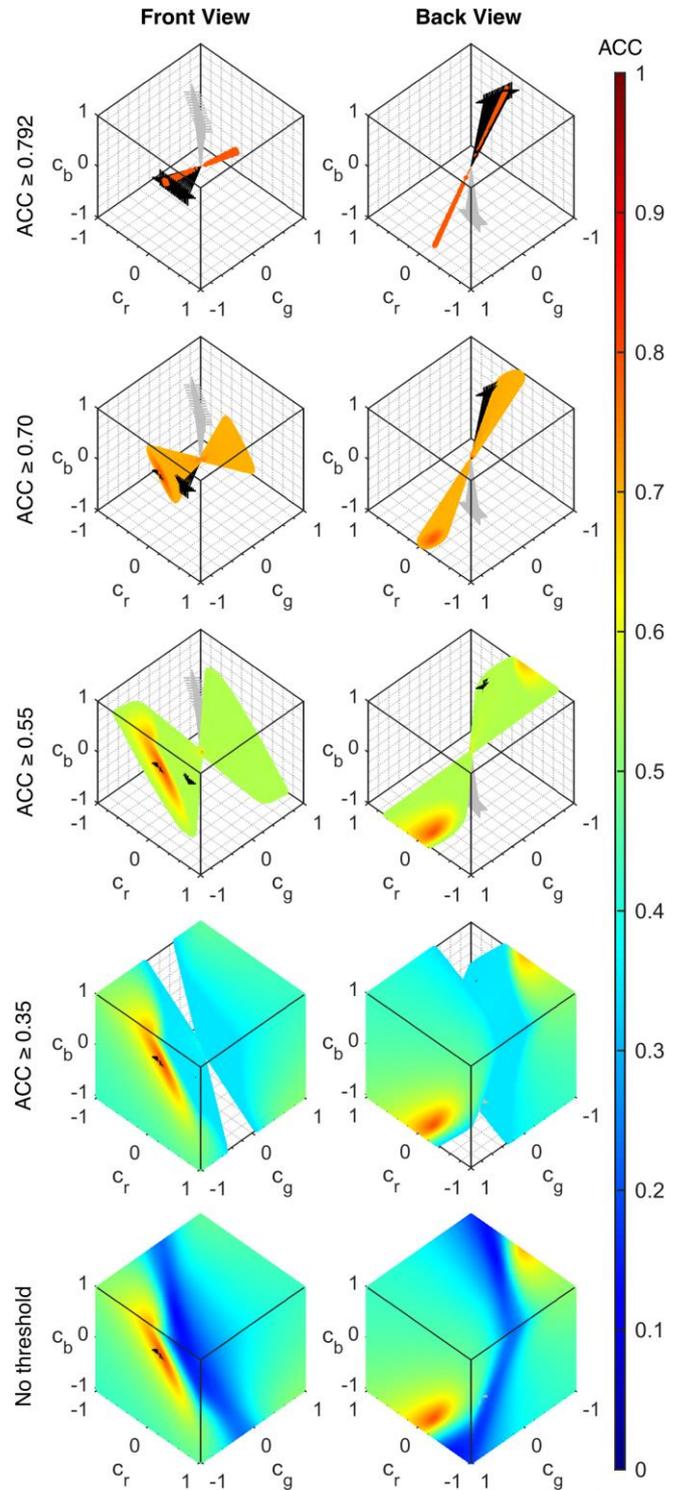


Fig. 6. Accuracy ACC results of the RGB combination cube grid search for the large forehead ROI (see FHL in Fig. 3). c_r , c_g , and c_b denote the color weighting coefficients of the red, green, and blue channel from (4). Each grid axis consists of 201 equidistant steps from -1 to +1 (grid granularity of 0.01) leading to $8 \cdot 120 \cdot 601$ grid points. For better perception, the grid is always shown from the front and back perspective. The black and gray lines mark RGB combinations of CHROM and POS for different $\alpha \in \{0.75, 0.8, \dots, 1.5\}$ (see Table 1 and Fig. 1). The accuracy converged in form of two elliptic cones that are point-symmetric to the grid origin. No isolated point clouds of high accuracies were found. To uncover the convergence axis of the elliptic cones (definition in (7)), grid points with an accuracy smaller than a certain threshold were hidden. The results reveal that an optimal static linear color combination for camera-based heart rate measurement exists and that it lies along the convergence axis.

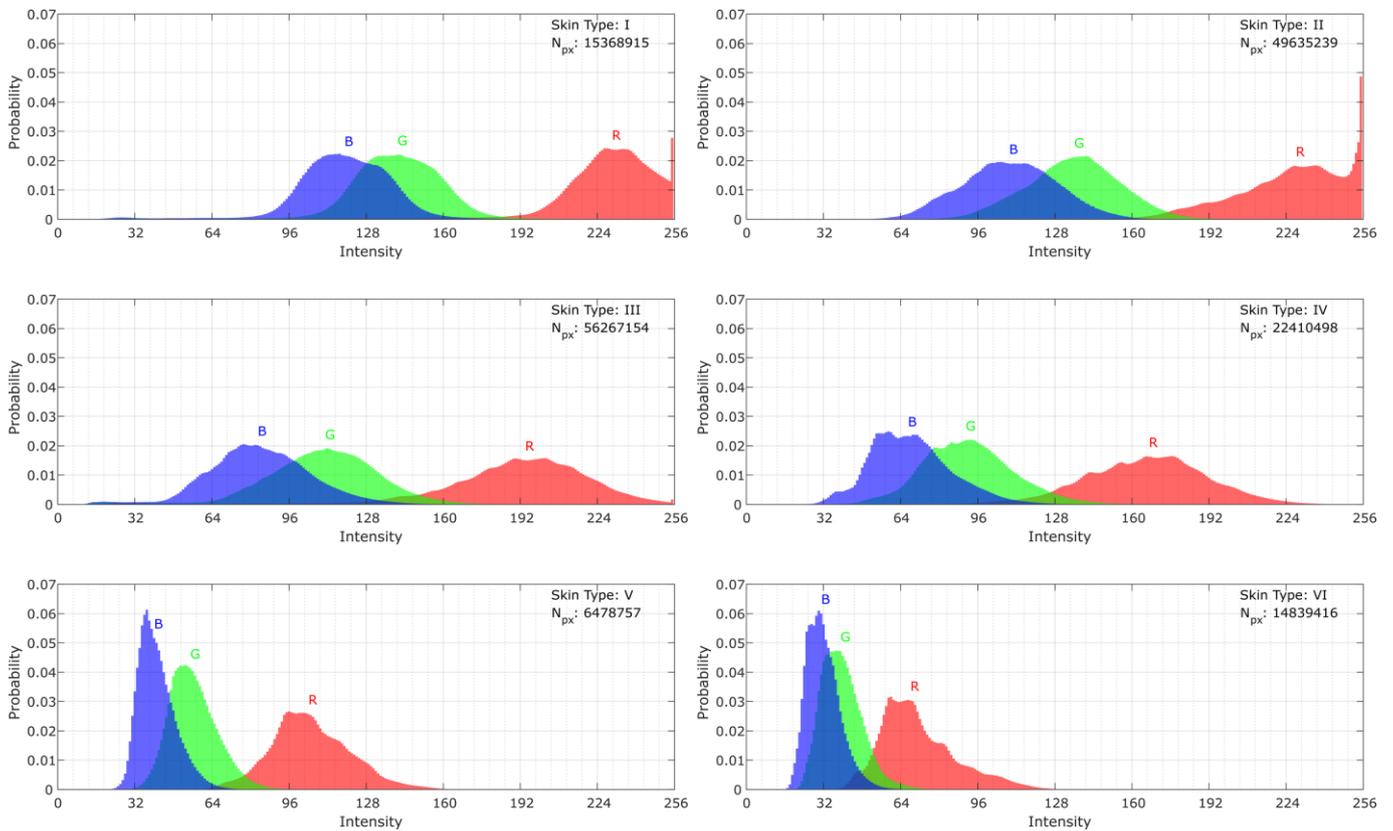


Fig. 7. RGB intensity histograms of forehead pixels (FHL in Fig. 3) from the first image of each 10 s segment by skin tone. Skin tone distributions across subjects and segments are listed in Table 2. N_{px} denotes the number of pixels in each histogram. Darker skin tones led to lower intensities.

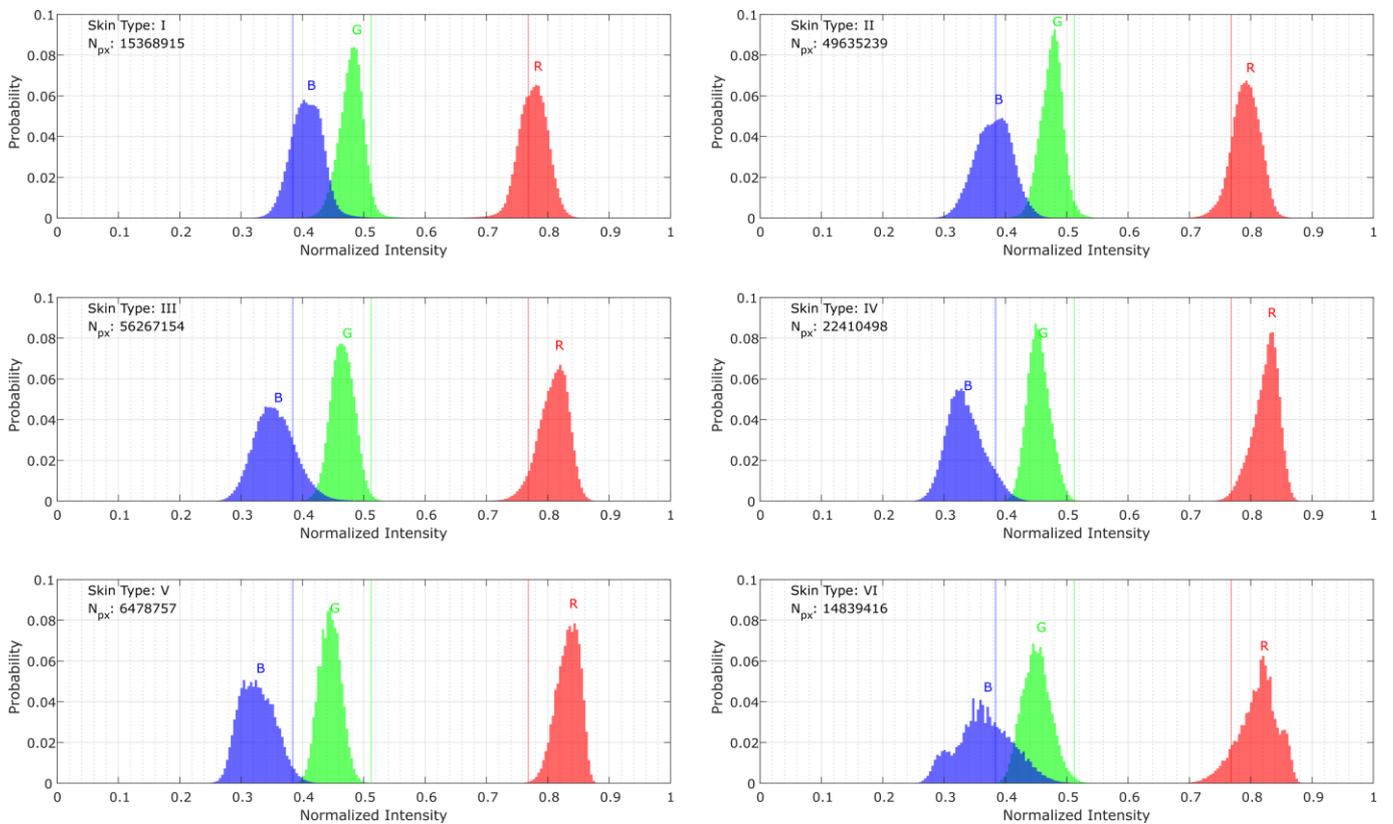


Fig. 8. Normalized RGB intensity histograms of forehead pixels (FHL in Fig. 3) from the first image of each 10 s segment by skin tone. Normalization with (1) projected the different intensity histograms of Fig. 7 onto a similar representation for all skin tones. Vertical lines mark the standardized skin color model $[R_s, G_s, B_s] = [0.7682, 0.5121, 0.3841]$ from [19], which differed slightly from the distribution centers of the normalized skin color histograms, particularly for darker skin tones. That the distributions show similar characteristics after normalization implies that the proportions between the RGB intensities remain similar across different skin tones. This is an essential prerequisite for finding a single optimal RGB combination for camera-based photoplethysmography across all skin tones. Otherwise, the optimal RGB combination would have to differ between skin tones.

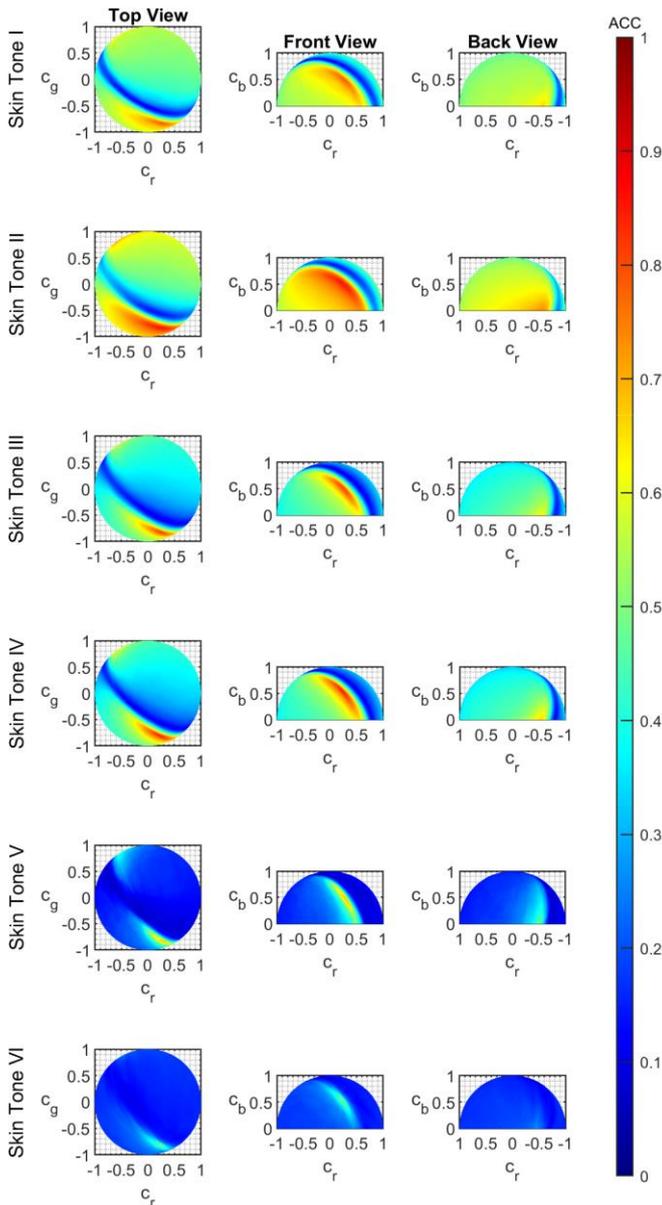


Fig. 9. Accuracy ACC results of the RGB combination hemispherical surface grid search by skin tone. c_r , c_g , and c_b denote the color weighting coefficients of the red, green, and blue channel from (4). The hemispherical surface grid (radius 1) was generated with Icosahedron-based S^2 Decomposition [28] and consists of 1'311'745 grid points. Skin tone distributions across subjects and segments are listed in Table 2. Maximum accuracy was found for all skin tones near the same position: $[c_r, c_g, c_b]^T = [0.25, -0.83, 0.5]^T$. The maximum accuracy for each skin tone is listed in Table 3. The overall performance dropped for skin tones V and VI. Skin tone II provided high accuracies over a larger range of RGB combinations compared to the other skin tones. A band of RGB combinations that yield low accuracies was observed at similar positions for all skin tones (blue diagonal in top view). The figure shows that the optimal RGB combination defined in (8) holds true across all skin tones.

Table 3. Maximum accuracy ACC_{max} of the hemispherical surface grid search by skin tone from Fig. 9.

Skin Tone	I	II	III	IV	V	VI
ACC_{max}	78.9 %	85.6 %	81.2 %	84.5 %	68.6 %	50.6 %

Skin tone distributions across subjects and segments are presented in Table 2.

6. Discussion

Study #1: The results of study #1 confirmed our findings presented in [15]. As expected, the HSV-H channel outperformed the RGB-G channel. The recently validated NTSC-Q channel competed with the HSV-H channel. The performance decline of the HSV-H channel in the Level-Set ROI was traced back to two effects. First, the utilized Bayesian classifier (see [23]) was pre-trained and sometimes falsely included tooth or eye regions. Second, temporary ROI mismatches can occur while the ROI was interpolated between two support frames of the tracking. Both effects led to distortions of the mean color within the Level-Set ROI, which affected especially the non-linear transformation of the HSV-H channel. These drawbacks can be overcome by improving skin detection and decreasing the distance between two support images of the ROI tracking, which requires more computational power. With the NTSC-Q channel, the Level-Set ROI can be considered as an alternative to glabella and forehead ROIs that does not depend on facial landmarks. The reduced performance of the cheek ROIs can be attributed to mimics and talking triggered by the BP4D+ stimuli (see Section 4.1). The stimuli also induced general motion such as head tilts. Motion of any kind still remains the biggest challenge for cbPPG. It strongly affects cbPPG signal quality and leads to erroneous camera-based HR measurement, but techniques to identify such signals exist [27]. Glabella and forehead ROIs represented the best choices. This reinforces the validity of study #2 and study #3, which both relied on the large forehead ROI (see FHL in Fig. 3). The large forehead ROI was chosen because it provided more pixels, i.e. more data for the skin color analysis of study #3.

Study #2: First of all, the expected properties of the cube grid (see Section 4.3) became evident. This confirmed the basis for the design of the hemispherical surface grid in study #3. The convergence line L_C (see (7)) corresponds approximately to the L^2 -normalized CHROM color combination (see Table 1 and Fig. 1) for $\alpha = 1.33$. With an ACC of 79.2 %, color combinations along L_C outperformed all other color channels of study #1 in the large forehead ROI (see Fig. 4). The gap of entries along L_C near the cube grid origin after thresholding with the maximum ACC (see Fig. 6) can be explained by insufficient grid granularity. The problem complexity, and thus the required computational power, increases with the third power of the inverse grid granularity. Therefore, only a limited step size of the color weighting coefficients c_r , c_g , and c_b was feasible, which allowed only rough relative weighting of the RGB channels near the origin. Towards the outside, the relative weighting of the RGB channels became more differentiated. Regarding the illustration of CHROM and POS in Fig. 6, a fundamental difference between those algorithms and the grid search color combinations must be noted: While the grid search color combinations (see (4)) remained static during the whole signal, CHROM and POS (see Table 1) update their tuning parameter α every 1.6 s, which leads to dynamic color combinations. For this reason, the CHROM and POS references in Fig. 6 can only serve as a rough orientation. In contrast to the dynamic color combinations, the static color combinations from the grid search preserve the low-frequency component of the cbPPG signal, which is essential for investigations of respiratory rate or vascular tone.

Study #3: Saturation (e.g. by specular reflection) as found in Fig. 7 is undesirable as it cancels the pulsatile signal component. Fig. 7 shows that this does not affect all color channels equally, but merely the R channel. This leads to an unintentional change in the color combination, which becomes dynamic in case of temporary clipping (fluctuations closely below the maximum intensity going temporarily into saturation). The fact that the

optimal RGB combination (see (8)) was found in the same area of the hemispherical surface grid for all skin tones proves that the pulsatile component of the cbPPG signal can be extracted with the same color combination for all skin tones. The overall loss of ACC in skin tones V and VI confirms the relation to increased melanin concentration. In skin tones I – IV, this effect was not observed.

Finally, two potential limitations of this work should be noted: The proportion of data related to skin tones V and VI is rather low (see Table 2), and, even though the illumination spectrum and camera sensitivity curves can be expected to remain constant within BP4D+, the optimal RGB color combination may depend on these variables and alter with the setup.

7. Conclusion

In the literature, only selected color space transformations have been examined. From these, the best performing color channels were H from the HSV color space and Q from the NTSC color space. In this contribution, for the first time, a systematic investigation of static linear RGB combinations was conducted. By applying a grid search, it was shown that an optimum for camera-based HR measurement exists. For future work, we suggest to refer our method as optimal color channel combination (O3C). O3C outperformed (79.2 % ACC, $S_{c,opt}$) the best color channel from the color space transformations (73.4 % ACC, HSV-H channel). It also yielded best results across all skin tones. However, the performance dropped for darker skin tones (V and VI), which can be attributed to higher melanin concentration and thus a diminished signal strength. The glabella ROI has been proven as a powerful alternative to the forehead ROI. For the Level-Set ROI this only holds for the NTSC-Q channel.

This work bases on an extraordinarily large database for cbPPG research. It helps to select a suitable approach for cbPPG signal extraction and demonstrates the influence of different skin tones. Fundamental research like this strengthens the potential to deploy computational intelligence for remote monitoring in applications where contact measurements pose an obstacle. Further work should address the physical and physiological background of the empirically determined optimal linear RGB combination and investigate issues related to data from different setups.

Acknowledgments

This work was partly supported by grants of the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG 319919706/GRK2323); the European Regional Development Fund (ERDF 100278533); the European Social Fund and the Free State of Saxony (ESF 100339450, "MEDICOS").

References

[1] J. Baker, A. Stanley, *Telemedicine Technology: a Review of Services, Equipment, and Other Aspects*, *Curr. Allergy Asthma Rep.* 18 (2018) 60. <https://doi.org/10.1007/s11882-018-0814-6>.

[2] I. Fister, K. Ljubić, P.N. Suganthan, M. Perc, I. Fister, *Computational intelligence in sports: Challenges and opportunities within a new research domain*, *Appl. Math. Comput.* 262 (2015) 178–186. <https://doi.org/10.1016/j.amc.2015.04.004>.

[3] S. Järvelin-Pasanen, S. Sinikallio, M.P. Tarvainen, *Heart rate variability and occupational stress - systematic review*, *Ind. Health.* 56 (2018) 500–511. <https://doi.org/10.2486/indhealth.2017-0190>.

[4] C. Bruser, C.H. Antink, T. Wartzek, M. Walter, S. Leonhardt, *Ambient and Unobtrusive Cardiorespiratory Monitoring Techniques*, *IEEE Rev. Biomed. Eng.* 8 (2015) 30–43. <https://doi.org/10.1109/RBME.2015.2414661>.

[5] Y. Sun, N. Thakor, *Photoplethysmography Revisited: From Contact to Noncontact, From Point to Imaging*, *IEEE Trans. Biomed. Eng.* 63 (2016) 463–477. <https://doi.org/10.1109/TBME.2015.2476337>.

[6] S. Zauneder, A. Trumpp, D. Wedekind, H. Malberg, *Cardiovascular assessment by imaging photoplethysmography - a review*, *Biomed. Eng. / Biomed. Tech.* 63 (2018) 617–634. <https://doi.org/10.1515/bmt-2017-0119>.

[7] C.H. Antink, S. Lyra, M. Paul, X. Yu, S. Leonhardt, *A Broader Look: Camera-Based Vital Sign Estimation across the Spectrum*, *Yearb. Med. Inform.* 28 (2019) 102–114. <https://doi.org/10.1055/s-0039-1677914>.

[8] W. Wang, A.C. den Brinker, S. Stuijk, G. de Haan, *Algorithmic Principles of Remote PPG*, *IEEE Trans. Biomed. Eng.* 64 (2017) 1479–1491. <https://doi.org/10.1109/TBME.2016.2609282>.

[9] W. Verkruyse, L.O. Svaasand, J.S. Nelson, *Remote plethysmographic imaging using ambient light*, *Opt. Express.* 16 (2008) 21434. <https://doi.org/10.1364/OE.16.021434>.

[10] A.N. Bashkatov, E.A. Genina, V.I. Kochubey, V. V Tuchin, *Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000 nm*, *J. Phys. D. Appl. Phys.* 38 (2005) 2543–2555. <https://doi.org/10.1088/0022-3727/38/15/004>.

[11] A. V Moço, S. Stuijk, G. de Haan, *New insights into the origin of remote PPG signals in visible light and infrared*, *Sci. Rep.* 8 (2018) 8501. <https://doi.org/10.1038/s41598-018-26068-2>.

[12] C. Lueangwattana, T. Kondo, H. Haneishi, *A comparative study of video signals for non-contact heart rate measurement*, in: 2015 12th Int. Conf. Electr. Eng. Comput. Telecommun. Inf. Technol., IEEE, 2015: pp. 1–5. <https://doi.org/10.1109/ECTICon.2015.7206971>.

[13] G.R. Tsouri, Z. Li, *On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras*, *J. Biomed. Opt.* 20 (2015) 048002. <https://doi.org/10.1117/1.JBO.20.4.048002>.

[14] W. Wang, A.C. den Brinker, S. Stuijk, G. de Haan, *Color-Distortion Filtering for Remote Photoplethysmography*, in: 2017 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG 2017), IEEE, 2017: pp. 71–78. <https://doi.org/10.1109/FG.2017.18>.

[15] H. Ernst, M. Scherpf, H. Malberg, M. Schmidt, *Color Spaces and Regions of Interest in Camera Based Heart Rate Estimation*, in: 2020 11th Conf. Eur. Study Gr. Cardiovasc. Oscil., IEEE, 2020: pp. 1–2. <https://doi.org/10.1109/ESGCO49734.2020.9158147>.

[16] Y. Yang, C. Liu, H. Yu, D. Shao, F. Tsow, N. Tao, *Motion robust remote photoplethysmography in CIE Lab color space*, *J. Biomed. Opt.* 21 (2016) 117001. <https://doi.org/10.1117/1.JBO.21.11.117001>.

[17] F. Bousefsaf, C. Maaoui, A. Pruski, *Automatic Selection of Webcam Photoplethysmographic Pixels Based on Lightness Criteria*, *J. Med. Biol. Eng.* 37 (2017) 374–385. <https://doi.org/10.1007/s40846-017-0229-1>.

[18] T.B. Fitzpatrick, *The Validity and Practicality of Sun-Reactive Skin Types I Through VI*, *Arch. Dermatol.* 124 (1988) 869. <https://doi.org/10.1001/archderm.1988.01670060015008>.

[19] G. de Haan, V. Jeanne, *Robust Pulse Rate From Chrominance-Based rPPG*, *IEEE Trans. Biomed. Eng.* 60 (2013) 2878–2886. <https://doi.org/10.1109/TBME.2013.2266196>.

[20] G. de Haan, A. van Leest, *Improved motion robustness of remote-PPG by using the blood volume pulse signature*, *Physiol. Meas.* 35 (2014) 1913–1926. <https://doi.org/10.1088/0967-3334/35/9/1913>.

[21] SMPTE 170M-2004 Composite Analog Video Signal - NTSC for Studio Applications, Society of Motion Picture and Television Engineers, United States of America, 2004. <https://doi.org/10.5594/S9781614823353>.

[22] Z. Zhang, J.M. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, M. Reale, A. Horowitz, H. Yang, J.F. Cohn, Q. Ji, L. Yin, *Multimodal Spontaneous Emotion Corpus for Human Behavior Analysis*, in: 2016 IEEE Conf. Comput. Vis. Pattern Recognit., IEEE, Las Vegas, NV, 2016: pp. 3438–3446. <https://doi.org/10.1109/CVPR.2016.374>.

[23] A. Trumpp, S. Rasche, D. Wedekind, M. Schmidt, T. Waldow, F. Gaetjen, K. Plötze, H. Malberg, K. Matschke, S. Zauneder, *Skin Detection and Tracking for Camera-Based Photoplethysmography Using a Bayesian Classifier and Level Set Segmentation*, in: *Bild. Für Die Medizin 2017*, Springer Vieweg, Berlin, Heidelberg, 2017: pp. 43–48. https://doi.org/10.1007/978-3-662-54345-0_16.

[24] T. von Arx, K. Tamura, O. Yukiya, S. Lozanoff, *The Face - A Vascular Perspective*, *Swiss Dent. J.* 128 (2018) 382–392. <http://www.ncbi.nlm.nih.gov/pubmed/29734800>.

[25] M. Rapczynski, P. Werner, F. Saxon, A. Al-Hamadi, *How the Region of Interest Impacts Contact Free Heart Rate Estimation Algorithms*, in: 2018 25th IEEE Int. Conf. Image Process., IEEE, 2018: pp. 2027–2031. <https://doi.org/10.1109/ICIP.2018.8451846>.

[26] S. Rasche, A. Trumpp, T. Waldow, F. Gaetjen, K. Plötze, D.

- Wedekind, M. Schmidt, H. Malberg, K. Matschke, S. Zaunseder,
Camera-based photoplethysmography in critical care patients, *Clin.
Hemorheol. Microcirc.* 64 (2016) 77–90.
<https://doi.org/10.3233/CH-162048>.
- [27] H. Ernst, H. Malberg, M. Schmidt, More Reliable Remote Heart
Rate Measurement by Signal Quality Indexes, in: 47th Conf.
Comput. Cardiol., IEEE, 2020: pp. 1–4.
<https://doi.org/10.22489/CinC.2020.165>.
- [28] A. Semechko, Suite of functions to perform uniform sampling of a
sphere (version 1.6.0.0), (2019).
<https://www.github.com/AntonSemechko/S2-Sampling-Toolbox>
(accessed June 17, 2020).