Color Spaces and Regions of Interest in Camera Based Heart Rate Estimation*

Hannes Ernst, Matthieu Scherpf, Hagen Malberg, and Martin Schmidt

***Abstract*—Camera-based heart rate (HR) estimation is used with different color spaces and regions of interest (ROIs). In this work, seven commonly used ROIs were determined via facial landmarks focusing on forehead, cheeks and glabella, and one ROI was derived by level set segmentation. For each ROI, 25 signals of eight color spaces were derived and HR estimation accuracy was determined. Both color channel and ROI showed significant effects ($p < 0.001$) on the accuracy of HR estimation. Glabella or forehead ROI combined with HSV-H or NTSC-Q color channel (accuracy: up to 73.3 %) as well as level set ROI with NTSC-Q color channel (accuracy: 68.7 %) performed best.**

I. Introduction

Camera-based photoplethysmography (cbPPG) is a non-contact measurement technique tracking pulsatile skin color changes in videos to assess cardiovascular parameters. It is most frequently applied to the face. The most investigated parameter is heart rate (HR). The processing pipeline for HR estimation consists of region of interest (ROI) definition, color transformation, signal processing, and HR estimation. [1]

ROI tracking in real life applications imposes high demands on motion robustness and independence from skin pigmentation. A poorly chosen ROI decreases the signal's HR-related power fraction or may even contain no physiological information at all. This hinders exact HR estimation. Therefore, facial markers are used to identify ROIs (e. g. forehead, cheek) that contain physiological information. However, findings on signal origin imply benefits from fine-tuning towards homogeneous regions. Trumpp et al. [2] proposed such a method which utilizes a Bayesian classifier and level set segmentation to find homogenous skin areas. [1]

Standard cameras record RGB color videos. In RGB color space, the green channel delivers the highest signal quality. Transformations into other color spaces have been used to improve motion robustness and HR estimation accuracy of cbPPG. For example, it has been reported that the hue channel from HSV space outperforms the green channel. [1]

This contribution compares the accuracy of cbPPG-based HR estimation from 25 color channels of 8 color spaces in combination with 7 feature-based ROIs and the level set ROI.

II. Data Material and Methods

A. Data Basis

For the comparison 330 RGB-videos (1392x1040 pixels, 3x8 bit, 25 fps) of 33 female subjects from the "Binghamton-Pittsburgh-RPI Multimodal Spontaneous Emotion Database" (BP4D+) [3] were used.

Subjects differ in skin type and were stimulated to trigger various emotions eliciting mimics and talking (detailed description in [3]). The BP4D+ contains facial feature coordinates and reference HR signals sampled at 1000 Hz. [3]

B. Regions of Interest

The level set method described in [2] starts off with Bayesian skin classification. For the first image of every video, two probabilities $p(c|skin)$ and $p(c|\neg skin)$ are estimated for each pixel c from its red, green and blue color values. By using the threshold θ, a pixel is finally labeled as skin if:

$$p(c|skin) \,/\, p(c|\neg skin) \geq \theta \qquad (1)$$

To improve classification results and extend for tracking over time, region-based level set segmentation is applied to the initial skin mask. The skin mask is modelled by the level set function Φ (contour: $\Phi = 0$, inside region (skin) Ω_1: $\Phi > 0$, outside region (non-skin) Ω_2: $\Phi < 0$) that is re-adjusted by a region-based gradient descent approach with every image, thus tracking a homogenous skin ROI over time. [2]

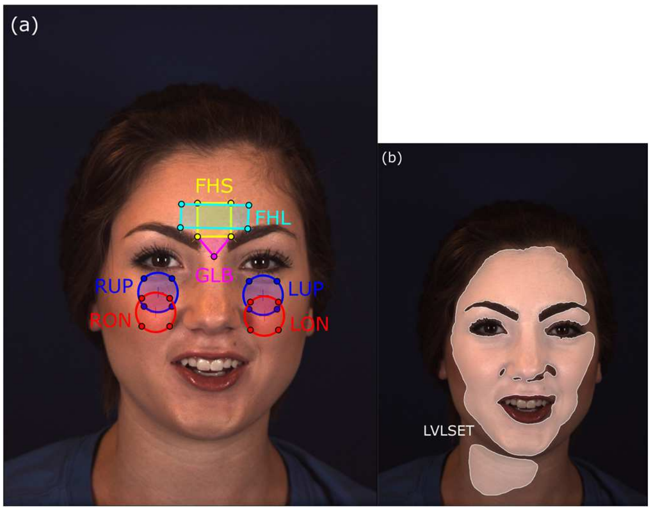

Fig. 1: Compared ROIs: (a) GLB – Glabella, FHS – Forehead small, FHL – Forehead large, LUP/RUP – Higher left/right cheek, LON/RON – On left/right cheek; (b) LVLSET – Level set. Modified image from BP4D+ [3].

Because it is known that forehead and cheek regions provide the best signal quality [1], several ROIs in these regions were derived from the facial landmarks. Fig. 1 gives an overview over all compared ROIs.

C. Color Transformations

For each ROI, all pixels were transformed from RGB color space to HSV, YCbCr, NTSC (also known as YIQ), XYZ, L*a*b, L*u*v, and CMYK color spaces. This led to a total of 25 color channels which are listed in the legend of fig. 2.

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

H. Ernst, M. Scherpf, H. Malberg, and M. Schmidt are with the Institute of Biomedical Engineering, TU Dresden, Dresden, Germany (e-mail: hannes.ernst@tu-dresden.de, matthieu.scherpf@tu-dresden.de, hagen.malberg@tu-dresden.de, martin_schmidt@tu-dresden.de)

978-1-7281-5751-1/20/$31.00 ©2020 IEEE

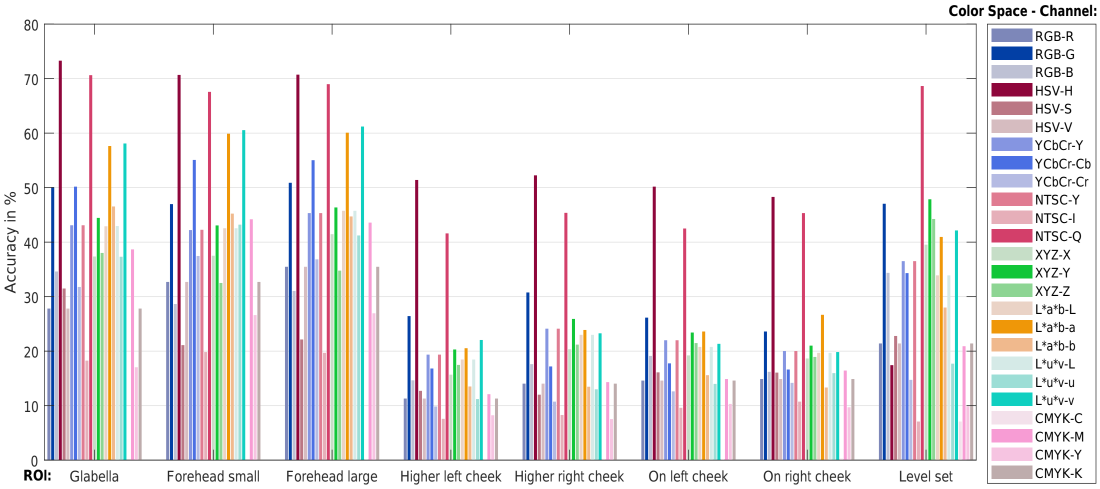

Fig. 2: Accuracy of heart rate estimation from 25 color channels of 8 color spaces for 8 different ROIs. NTSC(YIQ)-Q color channel worked across all ROIs. HSV-H color channel consistently performed best for all facial landmark ROIs, but not for the level set ROI. Cheek ROIs showed reduced accuracy.

D. Signal Processing and Heart Rate Estimation

For each ROI and each color channel, a cbPPG signal was acquired by averaging channel intensities frame-wise over all pixels within the particular ROI. A 5th order FIR band-pass filter then suppressed frequency components below 0.5 Hz and above 4 Hz. The filtered signals were split into 10 s segments with an overlap of 5 s. A fast Fourier transform generated the frequency spectrum of these signal segments. The maximum peak represents the estimated heart rate HR_{cam}.

E. Evaluation

Following [4], HR estimation performance was evaluated in accordance with IEC standard 60601-2-27 for medical ECG. To obtain a single reference heart rate HR_{ref} for each segment, the reference HR signal was averaged over segments with identical structure (10 s length, 5 s overlap). HR estimation was deemed erroneous if the absolute difference between HR_{cam} and HR_{ref} exceeded the greater of either 5 bpm or 10 % of HR_{ref}. Accuracy is the proportion of correctly estimated HR. Significant main effects of ROI and color channel on accuracy were analyzed with two-way ANOVA with repeated measures (rmANOVA).

III. Results

Fig. 2 shows the HR estimation accuracy for all color channels and ROIs. Two-way rmANOVA showed significant effects ($p < 0.001$) on accuracy for both color channel and ROI. The color channels that achieved the highest accuracy in their color space were: RGB-G, HSV-H, YCbCr-Cb, NTSC(YIQ)-Q, XYZ-Y, L*a*b-a, L*u*v-v, and CMYK-M. From all facial landmark ROIs, the cheek regions consistently exhibited lower accuracy than forehead and glabella regions. The level set ROI kept up with the latter in RGB-G, NTSC-Q and XYZ-Y channels. Glabella or forehead ROI combined with HSV-H or NTSC-Q channel delivered the best results. Highest accuracy (73.3 %) was achieved by glabella ROI and HSV-H color channel. The level set ROI using NTSC-Q color channel achieved comparable accuracy (68.7 %).

IV. Discussion

Our findings about the best performing color channels are consistent with the literature [1]. However, the high performance of NTSC-Q was not expected, as the channel has rarely been used and investigated in the context of cbPPG. Furthermore, we found a major performance decline of HSV-H color channel in the level set ROI.

Each ROI showed individual advantages and disadvantages. In case of muscular activity, as perceived during talking or mimics, facial landmarks move, which negatively impacts signal quality. While facial landmark detection is limited to a frontal view, it is fast. The level set ROI in contrast requires more computational power, but can be applied to all skin areas. However, the wide variety of skin types makes all-encompassing skin modelling difficult. The level set ROI is robust to hair that protrudes into the forehead region, but the glabella ROI seems to be an alternative for people not wearing glasses. While the forehead often serves as ROI, glabella rarely receives attention. Our findings indicate that the glabella ROI is a promising option for PPG implementation, e. g. in the context of AR/VR glasses.

The achieved accuracy of camera-based HR estimation on BP4D+ data leaves room for further improvement. Head movements, mimics, talking, 8-bit quantization and different skin types pose the main challenges. However, beneficial ROIs and color channels were identified, which constitutes an encouraging result. Continuing work should focus on overall performance improvement (e. g. by channel combination) and utilize the full data set including male subjects.

References

[1] S. Zaunseder, A. Trumpp, D. Wedekind, and H. Malberg, “Cardiovascular assessment by imaging photoplethysmography – a review,” *Biomed. Eng. / Biomed. Tech.*, vol. 63, no. 5, pp. 617–634, Jun. 2018.

[2] A. Trumpp *et al.*, “Skin Detection and Tracking for Camera-Based Photoplethysmography Using a Bayesian Classifier and Level Set Segmentation,” in *Bildverarbeitung für die Medizin 2017*, Heidelberg: Springer Vieweg, Berlin, Heidelberg, 2017, pp. 43–48.

[3] Z. Zhang *et al.*, “Multimodal Spontaneous Emotion Corpus for Human Behavior Analysis,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 3438–3446.

[4] M. Rapczynski, P. Werner, F. Saxen, and A. Al-Hamadi, “How the Region of Interest Impacts Contact Free Heart Rate Estimation Algorithms,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 2027–2031.