Skin Detection and Tracking for Camera-Based Photoplethysmography Using a Bayesian Classifier and Level Set Segmentation

Alexander Trumpp[1], Stefan Rasche[2], Daniel Wedekind[1], Martin Schmidt[1], Thomas Waldow[3], Frederik Gaetjen[2], Katrin Plötze[2], Hagen Malberg[1], Klaus Matschke[2], Sebastian Zaunseder[1]

[1]Institute of Biomedical Engineering, TU Dresden
[2]Department of Cardiac Surgery, Herzzentrum Dresden
[3]Center for Wound Therapy in Medical Rehabilitation, Klinik Bavaria Kreischa
alexander.trumpp@tu-dresden.de

Abstract. Camera-Based Photoplethysmography is a measuring technique that permits the remote assessment of vital signs by using cameras. The face is the preferred area of measurement (region of interest: ROI) that has to be selected automatically for convenient application. Most works use common face detection algorithm for this purpose. However, these approaches often fail if the face is partly occluded or distorted. In this work, we propose an automatic method for ROI detection and tracking that does not rely on facial features. First, a Bayesian skin classifier was applied. Second, the detected areas were refined and tracked by level set segmentation. We tested our method on videos of 70 patients. The determined ROIs were used for signal extraction and heart rate (HR) estimation. The results showed that our method can detect and track suitable skin regions. We achieved a median HR detection rate of 80 % which was only 6 % lower than when applying manually defined ROIs.

1 Introduction

Camera-based photoplethysmography (cbPPG) is an optical measuring technique that permits the contactless derivation of cardiorespiratory signals by using normal video cameras [1]. Similar to the common reflective photoplethysmography, backscattered light from superficial skin layers is captured over time and converted into a signal which relates to the cardiac cycle. In addition to the advantage of a remote application, cbPPG can be operated with only ambient light and can allow a spatial assessment of the recorded area [2].

For a variety of reasons, the face is the preferred region of measurement for cbPPG. Therefore, the automatic detection and tracking of suitable facial areas are essential for a convenient use of this technique. Existing approaches mostly rely on detection algorithms for the whole face or facial landmarks and subsequently involve a tracking based on either re-detection or by using determined image features [3, 4, 5]. However, faces that are partially occluded, rotated or

not in a frontal position towards the camera might hinder any valid detection. To our knowledge, there are only few works related to cbPPG which propose an automatic detection and tracking method that does not require the initial registration of facial areas.

In this paper, we present a fully automated approach for the detection and tracking of face regions that are most suitable for cbPPG. For this approach, no prior knowledge about the image content is needed. We tested our method for camera recordings of 70 patients in an intensive care unit. The evaluation was performed using cbPPG's detection rate of the heart rate (HR). We compared the results of our method to results obtained by applying manually annotated regions of interest (ROIs).

2 Material and methods

2.1 Data and technical setup

For our tests, we analyzed measurements of a study that featured 70 patients (50 male, 20 female, ages 70.3±11.4 years) after cardiac surgery [6]. All participants were recorded for about 30 min using a two-camera system. In this work, we only considered the RGB camera (IDS UI-3370CP-C-HQ) which was set to a resolution of 420x320 pixels, a frame rate of 100 fps and a color depth of 12 bit. During the recording, the patients were usually not conscious but sometimes woke up and moved their head. The illumination conditions were mainly defined by the indoor light source and outdoor sunlight. Synchronously to the videos, we captured reference signals from the clinical monitor system. The study was approved by the Institutional Review Board of the TU Dresden (IRB00001473, EK168052013) and each patient gave written consent.

2.2 Image processing

To verify our assumption, we tested common face detection algorithms like the Viola-Jones method on the videos. These algorithms mostly failed due to patients' head position and occlusion caused by fixation tapes as well as intubation tubes. Therefore, we applied a Naive Bayes classifier to detect suitable skin areas instead of faces. Following the description of Jones and Rehg [7], we built two RGB histograms, one for skin color, and one for non-skin color. We used labeled skin and non-skin images for this purpose that were made available by this group. The normalization of the histograms on the total number of entries provided the probability density functions (PDFs) for the classes skin and ¬skin: $p(\mathbf{c}|\text{skin})$, $p(\mathbf{c}|\neg\text{skin})$. The classifier could then be formulated using the Bayesian decision rule [8]. An RGB pixel $\mathbf{c}$ was set as skin if

$$\frac{p(\mathbf{c}|\text{skin})}{p(\mathbf{c}|\neg\text{skin})} \geq \theta \tag{1}$$

We applied the classifier on the first image of each RGB video. For every patient, the threshold θ was automatically adapted between 0.1 and 10 based on the

number of detected skin pixels. This adaption allowed us to compensate varying conditions among the patients. The classifier is generally a good choice due to its simplicity, the free access of the training data, and because it was often validated. However, in our case, the results did usually not represent the skin regions well enough (Fig. 2). Therefore, we applied a segmentation algorithm based on level set methods to refine the outcome.

Level set methods are able to describe the propagation of a segmentation contour using an implicit representation [9]. This representation is accomplished by the level set function $\Phi(x, y, t)$, where $\Phi = 0$ defines the contour, $\Phi > 0$ the inside region Ω_1 and $\Phi < 0$ the outside region Ω_2. For each segmentation procedure, the contour propagates from an initialization point to an optimal state that minimizes an appropriate energy function. The minimization is performed by a gradient descent. We applied a region-based approach by Brox et al. [10] in which the gradient descent reads to

$$\frac{\partial \Phi}{\partial t} = H'(\Phi) \left[\sum_{j=1}^{3} \log \frac{p_{1j}(I_j)}{p_{2j}(I_j)} + \nu \cdot \mathrm{div} \frac{\nabla \Phi}{|\nabla \Phi|} \right] \tag{2}$$

For the considered image I (j: color channels) in our study, the segmentation was initialized by setting Ω_1 to the skin classification result. During the segmentation, the first sum term allows to separate Ω_1 and Ω_2 due to their local intensities. The PDFs p_{1j} and p_{2j} are modeled using Gaussian distributions and describe the probabilities that an regional image pixel belongs to Ω_1 or Ω_2, respectively. For an optimal outcome, p_{1j} is maximal in Ω_1 and p_{2j} is maximal in Ω_2. The latter term in the equation is the curvature term which allows to control the smoothness of the contour (ν is a weighting factor). $H(\Phi)$ is a Heaviside function. We used 300 iteration steps to achieve the final segmentation result which was defined by Ω_1 and represents our ROI (Fig. 2).

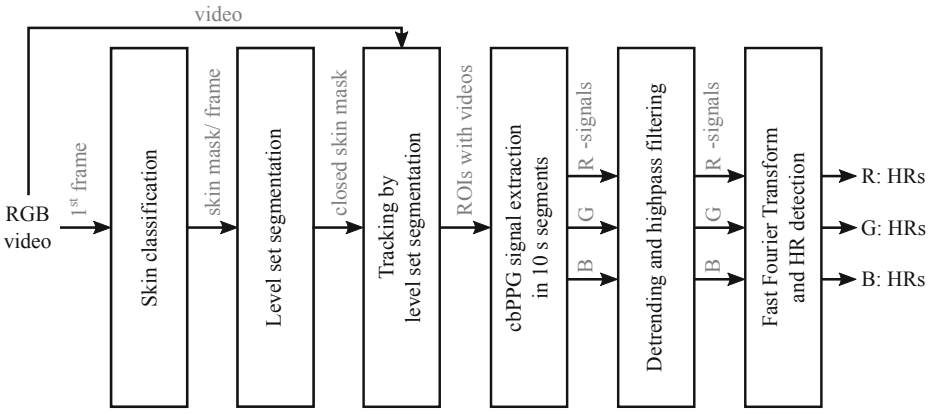

Fig. 1. Program sequence of the proposed procedure (ROI: region of interest, HR: heart rate, R: red, G: green, B: blue). They gray terms represent the transferred data.

The segmented skin regions in the first RGB images were further tracked in order to obtain ROIs for the whole videos. We also applied level set segmentation for this purpose and always used the last segmentation result as initialization point for the next image. Consequently, the contour quickly stabilized even in the case of motion or light variations. If the result changed at all between two images, a maximum of only 50 iteration steps was necessary. The chosen level set method (combined with the classifier) allowed us to track skin regions with similar intensity distributions instead of certain anatomical areas. We hypothesize that the consideration of these distributions provides a benefit for ROI selection.

2.3 Signal processing and evaluation

In order to extract cbPPG signals, the pixel values within the determined ROIs were averaged frame by frame. Due to the three color channels (R,G,B), we obtained three different signals for each patient. All signals were divided into 10 s segments and then detrended and highpass filtered (FIR filter, a cutoff frequency of 30 bpm, an order of 250). Afterward, a Fast Fourier Transform was applied and an amplitude spectrum was built for each segment. We detected the HR in this spectrum by searching for the maximum peak between 30 and 200 bpm. Fig. 1 summarizes the signal and image processing steps.

The detected HRs were compared to reference HRs that were determined from the electrocardiogram. A camera-based HR was considered correct if it differed less than 5 bpm. Using this outcome, we could calculate a mean HR detection rate (HRDR) for each patient and color channel. In a previous work [6], we worked with the same video data but manually annotated ROIs and only calculated HRDR values for the green channel. We used these results as the gold standard to validate our method.

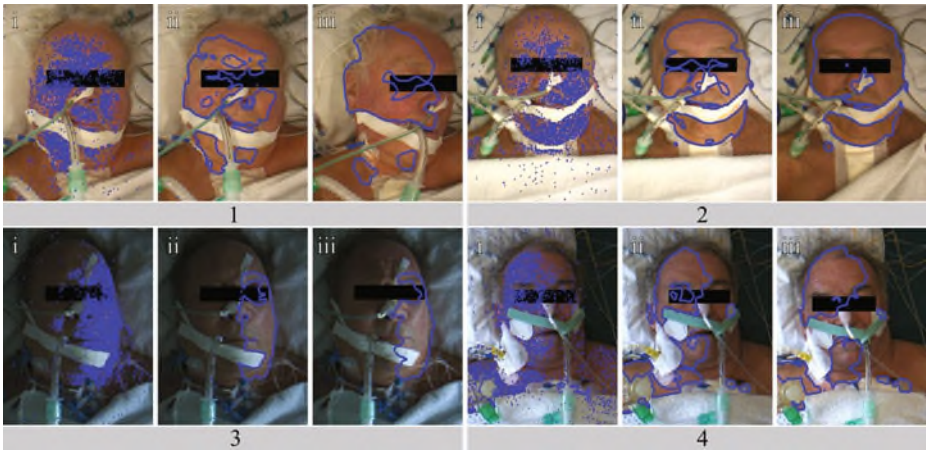

Fig. 2. Skin detection and tracking results for four patients (only contours are shown): i) skin classification, ii) initial segmentation, iii) tracking to a later point in the video.

3 Results

The proposed method proved to be applicable to detect and track suitable skin regions in RGB videos. Besides the face, also regions from the upper body part were occasionally included. Fig. 2 shows four patients where the skin detection, segmentation and one tracking result are visualized. The combination of the classifier and level set segmentation leads to the separation of skin areas (ROIs) with similarly distributed intensities. We can argue that this distribution plays a role for the quality of the extracted cbPPG signals when averaging the ROI pixel values. Patient 3 and 4 in Fig. 2 serve as examples. The light conditions caused the face to be illuminated differently. Our method only selected the appropriate part as ROI and therefore caused a better outcome than using the whole face. We also tested the processing time of our algorithm. Except for the initial detection and segmentation, it can be operated in real-time (MATLAB).

Fig. 3 shows the results for HRDR of the 70 patients (depicted in boxplots). When our ROI selection method was applied, the green channel provided the best rates with a median of 80 %. This median is remarkably high considering that the gold standard, where ROIs were selected manually, yielded 86 %. However, low HRDR values cause a large variance for our method. In comparison to the green channel, the red and blue channel provided a generally poor outcome with a median HRDR of 17 % and 14 %, respectively.

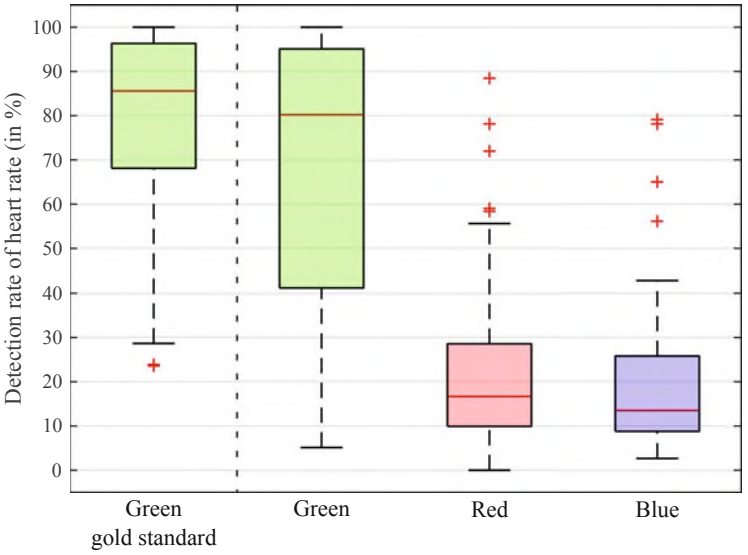

Fig. 3. Results for the detection rate of the heart rate. The first box shows the results using manually annotated ROI. The other boxes show the results for each color channel using our method.

4 Discussion

To some extent, the low HRDR values in the green channel can be explained by a group of patients where the ROI selection process was unsuccessful. Reasons are: (i) insufficient skin classification, (ii) inclusion of non-skin regions in the segmentation, (iii) failed tracking because contour did not stabilize enough. Moreover, the inclusion of non-facial skin regions can also affect the HRDR. Such regions might hold pulse signals with different characteristics and could degrade the outcome if they are combined with facial regions. Although we used the same ROIs for the red and blue channel as we used for the green one, the results differed significantly. This effect can be partly attributed to generally low cbPPG signal amplitudes for red and blue and coincides with the fact that the green channel provides the strongest plethysmographic signal [2]. Nevertheless, signals from the red and blue channel still offer valuable information which should be exploited in future analyses. For example, Poh et al. [3] applied an independent component analysis to the signals of all three channels and, therefore, could achieve a better HR detection than just considering the green channel.

In conclusion, our method allows to automatically determine facial ROIs in videos without using any prior knowledge about the image content. Furthermore, these regions proved to be suitable for cbPPG signal extraction and HR estimation. This work was funded by the "Staatsministerium für Wissenschaft und Kunst (SMWK)" in Saxony, Germany.

References

1. Huelsbusch M, Blazek V. Contactless mapping of rhythmical phenomena in tissue perfusion using PPGI. Proc SPIE. 2002;4683:110–7.
2. Verkruysse W, Svaasand LO, Nelson JS. Remote plethysmographic imaging using ambient light. Opt Express. 2008;16(26):21434–45.
3. Poh MZ, McDuff DJ, Picard RW. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt Express. 2010;18(10):10762–74.
4. Tarassenko L, Villarroel M, Guazzi A, et al. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. Physiol Meas. 2014;35(5):807–31.
5. Kumar M, Veeraraghavan A, Sabharwal A. DistancePPG: robust non-contact vital signs monitoring using a camera. Biomed Opt Express. 2015;6(5):1565–88.
6. Rasche S, Trumpp A, Waldow T, et al. Camera-based photoplethysmography in critical care patients. Clin Hemorheol Microcirc. 2016;64(1):77–90.
7. Jones MJ, Rehg JM. Statistical color models with application to skin detection. Int J Comput Vis. 2002;46(1):81–96.
8. Duda RO, Hart PE, Stork DG. Pattern Classification. John Wiley & Sons; 2001.
9. Osher S, Fedkiw R. Level Set Methods and Dynamic Implicit Surfaces. New York: Springer; 2003.
10. Brox T, Rousson M, Deriche R, et al. Colour, texture, and motion in level set based segmentation and tracking. Image Vis Comput. 2010;28(3):376–90.