Distant pulse oximetry based on skin region extraction and multi-spectral measurement

Christian Herrmann[a,b], Jürgen Metzler[a], Dieter Willersinn[a], and Jürgen Beyerer[a,b]

[a]Fraunhofer IOSB, Karlsruhe, Germany
[b]Vision and Fusion Lab, Karlsruhe Institute of Technology KIT, Karlsruhe, Germany

ABSTRACT

Capturing vital signs, specifically heart rate and oxygen saturation, is essential in care situations. Clinical pulse oximetry solutions work contact-based by clips or otherwise fixed sensor units which have sometimes undesired impact on the patient. A typical example would be pre-term infants in neonatal care which require permanent monitoring and have a very fragile skin. This requires a regular change of the sensor unit location by the staff to avoid skin damage. To improve patient comfort and to reduce care effort, a feasibility study with a camera-based passive optical method for contactless pulse oximetry from a distance is performed. In contrast to most existing research on contactless pulse oximetry, a task-optimized multi-spectral sensor unit instead of a standard RGB-camera is proposed. This first allows to avoid the widely used green spectral range for distant heart rate measurement, which is unsuitable for pulse oximetry due to nearly equal spectral extinction coefficients of saturated oxy-hemoglobin and non-saturated hemoglobin. Second, it also better addresses the challenge of the worse signal-to-noise ratio than in the contact-based or active measurement, e.g., caused by background illumination. Signal noise from background illumination is addressed in several ways. The key part is an automated reference measurement of background illumination by automated patient localization in the acquired images by extraction of skin and background regions with a CNN-based detector. Due to the custom spectral ranges, the detector is trained and optimized for this specific setup. Altogether, allowing a contactless measurement, the studied concept promises to improve the care of patients where skin contact has negative effects.

Keywords: distant pulse oximetry, multi-spectral, skin segmentation

1. INTRODUCTION

There exist care situations where contact-based capturing of vital signs is undesired, impacting the patient or causing additional effort. Here, we focus on heart rate and oxygen saturation (SpO_2) measurement by pulse oximetry. Typical examples are neonatal care units where pre-term infants have a very fragile skin but need permanent monitoring. Currently, sensors are switched e.g. from one leg to the other every few hours by the nursing staff to avoid skin damage. Besides neonatal care, further situations include emergency medicine, such as burn victims. To reduce the staff effort and to improve patient comfort, we study a contactless method to determine heart rate and oxygen saturation by pulse oximetry. For replacing the contact-based pulse oximetry with a passive contactless solution working from a distance, several challenges have to be addressed. While the basic measurement principle remains, the signal-to-noise ratio is expected to be significantly worse. First, environment illumination is the major part of the captured signal in the camera, whereas the measurement signal is only a small fraction. Second, patient motion has to be handled because the camera is not moving with the patient in comparison to, e.g., contact-based finger clips. Third, contact-based pulse-oximetry works with active illumination which is impractical and disturbing if used for a distant setup, thus the ambient illumination in the room has to suffice. All these challenges are addressed by according image and signal processing methods.

The proposed sensor unit consists of three separate monochrome cameras, each mounted with an according optical filter to distinguish the spectral absorption of saturated (HbO_2) and unsaturated hemoglobin (Hb). The combination of three spectral ranges is optimized with regard to the signal-to-noise ratio. The finally selected configuration is depicted

Further author information:
C. Herrmann: christian.herrmann@iosb.fraunhofer.de
J. Metzler: juergen.metzler@iosb.fraunhofer.de
D. Willersinn: dieter.willersinn@iosb.fraunhofer.de
J. Beyerer: juergen.beyerer@iosb.fraunhofer.de

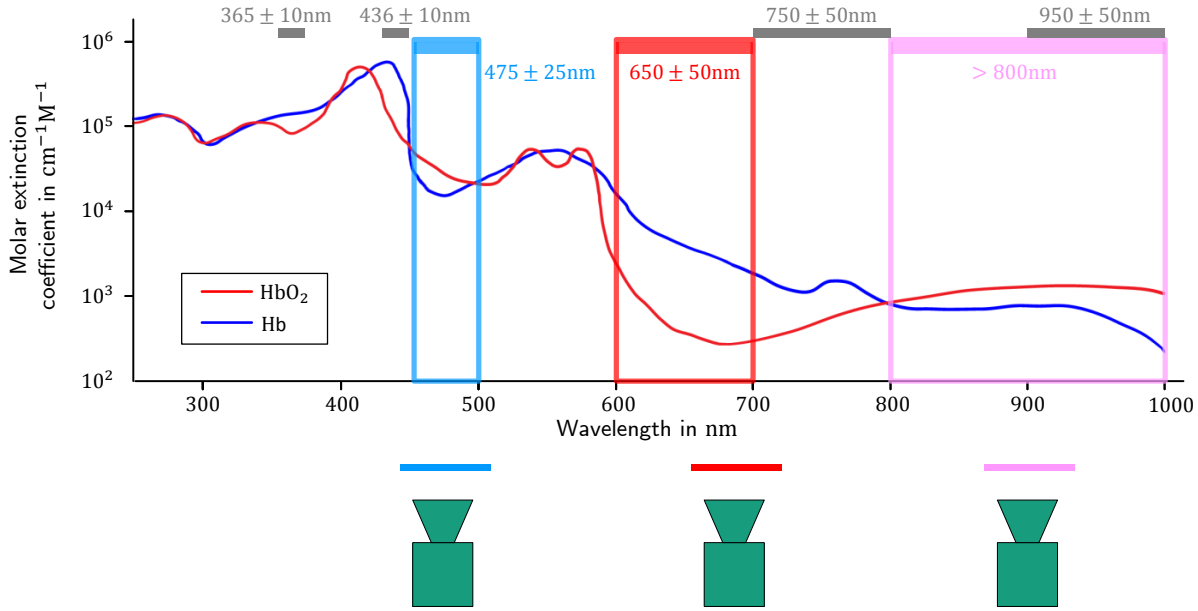

Figure 1. Schematic drawing of selected spectral ranges to distinguish Hb and HbO_2. Ranges of finally selected spectral filters are shown with colored boxes. Further studied filter options are indicated by gray bars.

in figure 1. Besides using the common 650 and 900 nm range, one additional blue wavelength range around 475 nm is used. The contribution of this work is a multi-spectral camera-based system for contactless and passive pulse oximetry to measure the heart rate and oxygen saturation. Two key aspects are considered:

- Fitting tailor-made optical filters increases the signal-to-noise ratio compared to regular RGB-cameras which are widely applied for this task. This avoids the overlapping and non-sharp RGB spectral ranges which are not designed to match the task. The proposed three channel setup increases measurement stability in comparison to the common two channel option.
- Signal and image processing strategies are proposed to measure the desired vital signs in the distant pulse oximetry setup. This includes an automated skin and background region extraction based on CNNs.

2. RELATED WORK

As contactless measurement of vital signs is a desirable goal, research addresses this continuously and steadily tries to relax the required conditions. A currently widespread application is the heart rate measurement with inexpensive cameras, such as smartphones. Usually, the face regions is detected and used for pulse detection. Methods to extract the heart rate from the intensity signal include FFT, ICA, or auto-regressive models[1,2] . Further vital signs in the focus of contactless measurement include respiration and oxygen saturation[3] .

Estimating oxygen saturation based on optical pulse measurement is usually referred to as pulse oximetry. In pulse oximetry, the oxygen saturation (SpO_2) measurement is based on different spectral extinction coefficients between saturated oxy-hemoglobin (HbO_2) and non-saturated hemoglobin (Hb). For measuring the oxygen saturation, at least two wavelengths with different extinction coefficients are required. Common contact-based solutions use 660nm and 950nm LED light sources because of their cheap availability. Absorption at the respective spectral range is measured by alternate switching of the LEDs and a simple intensity sensor. Thus, the spectral ranges are selected on illumination side.

Existing contactless approaches can be separated in two main categories: active and passive illumination. Active approaches define a setup with an illumination source[4] , either in a transmissive[5,6] or a reflective setup[7,8] . The transmissive setup usually generates a better signal to noise ratio but requires a more complex physical measurement setup because the illumination has to be on the opposite side of the inspected tissue with respect to the sensor device. In general,

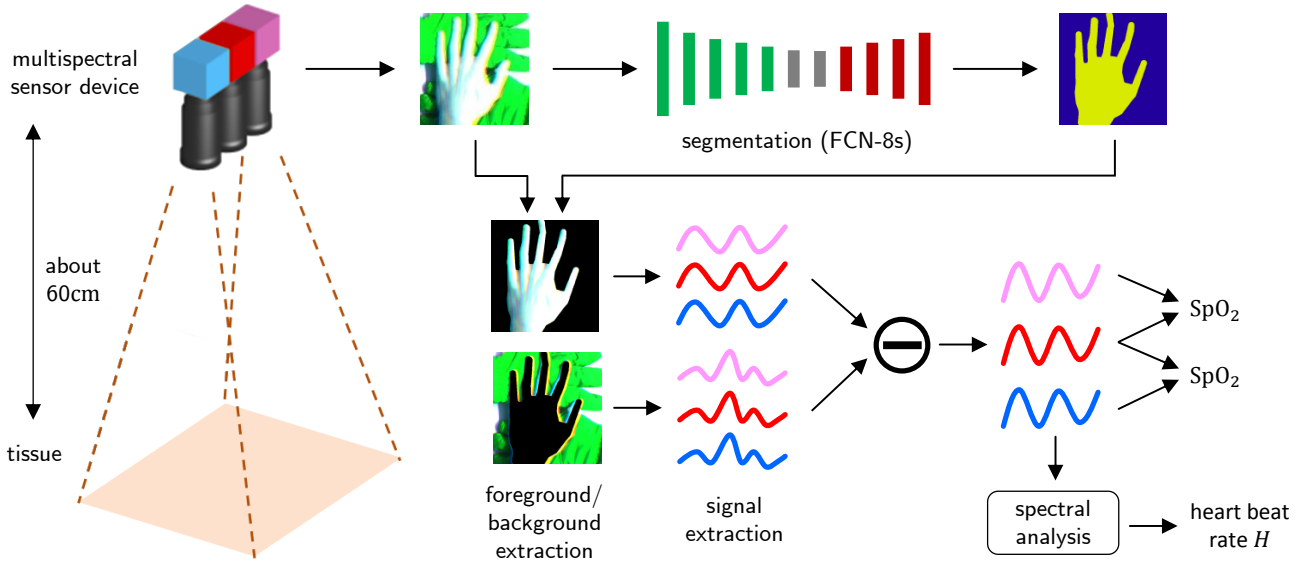

Figure 2. Proposed concept with three spectral cameras observing the tissue and several signal processing steps.

the active strategy means employing selected LED light sources as spectral illumination and combining it with a distant monochrome camera as sensor device. Recently, passive approaches were proposed where the environment illumination serves as light source to determine the oxygen saturation. This allows a more flexible setup, as only the sensor device remains in the system. However, the spectral ranges have to be selected on sensor side. As sensor device, either a regular RGB camera[3,9,10] or a dual camera setup with tailor-made spectral filters[11,12] is applied.

We follow the passive option as it offers a simplified installation. Because the RGB spectral ranges are non-optimal, we suggest replacing it with a task-optimized multi-spectral sensor unit. This avoids the most obvious shortcoming of the RGB setup which is the green spectral range. The spectral extinction coefficients of saturated oxy-hemoglobin and non-saturated hemoglobin are nearly equal there making it unsuitable for pulse oximetry. In contrast to other approaches working with tailor-made spectral filters, the proposed system consists of three instead of two spectral ranges to increase measurement stability.

3. MULTI-SPECTRAL CONTACTLESS CONCEPT

The whole proposed concept is illustrated in figure 2. This section focuses on the basic principle and the hardware design. The signal processing will be discussed in the following sections.

The designed concept has two key aspects: passive and contactless measurement. This avoids the complexity of any active illumination source which would restrict the geometric setup. The sensor device consists of three monochrome CMOS cameras with quantum efficiency fading out just above 1000 nm. Each camera is equipped with a spectral filter (usually a bandpass) to focus on certain spectral ranges. The cameras have a spatial resolution of 1280×1024 pixels and the capturing distance between the sensor device and the inspected tissue is about 60cm. As oxygen saturation measurement works best on extremities, measurement on the hand and forearm is performed in the feasibility study.

While our studied sensor unit consists of three separate monochrome cameras, each mounted with an according optical filter, a production system can easily combine the filters directly with the sensor chip, making it no more complex than a single RGB-camera. The combination of the three spectral ranges is optimized with regard to the signal-to-noise ratio. The focus has to be on spectral areas where the extinction coefficients of HbO_2 and Hb are as different as possible and the influence by surrounding disturbances is minimized.

Ambient effects are handled in several ways. First, following the completely passive approach, we synchronize the camera frame rate with the grid power frequency to reduce potential aliasing effects with the artificial background illumination. The selected frame rate is 10 fps which aligns with the European power grid frequency of 50 Hz. To further reduce effects from varying background illumination, patient location is determined by skin regions detection with a pose-tolerant

Table 1. Studied spectral filters with achieved sensor SNR in the test setup.

filter	SNR in dB
365±10 nm	20.4
436±10 nm	81.2
475±25 nm	81.6
650±50 nm	81.8
750±50 nm	70.6
800 nm longpass	74.1
950±50 nm	30.5
none	81.8

CNN-based skin and body part segmentation which is adapted to the custom spectral ranges of the proposed system. Then, subtracting a reference measurement from non-skin background regions within the same image maintains only the intensity modulations caused by the patients pulse in the signal. Each of the three spectral signals (channel) is then bandpass filtered with regard to the expected pulse frequency range. Heart rate measurement is performed by autocorrelation within one channel which is proven more stable than a heartbeat peak detection or a frequency domain analysis for the present signal-to-noise ratio. The oxygen saturation is measured by the ratio of ratios method on the preprocessed signals. The three independent channels allow two independent oxygen saturation measurements. Their combination then increases the measurement stability. This is an improvement to a regular RGB-camera where only the red and blue channel are suitable for pulse oximetry resulting in only a single measurement.

Candidates for spectral filters are selected based on the difference between $\mathrm{HbO_2}$ and Hb extinction as well as off-the-shelf availability of high-quality bandpass filters. Table 1 lists all reasonable filter choices. Note that the captured spectrum for the 800 nm longpass filter is limited by the sensitivity of the camera sensor to slightly above 1000 nm, thus making it to behave similar to a 900±100 nm bandpass. Filters are selected based on the resulting signal to noise ratio of the captured images:

$$SNR = \frac{\mathrm{mean}(G+H)}{\mathrm{std}(G-H)}\,. \tag{1}$$

Where G and H denote consecutively captured images with fixed camera settings of an appropriate fixed probe scene. As expected, table 1 indicates that narrow filters and filters at the spectral edges of the sensor sensitivity result in the worst signal to noise ratios. Consequently, the 475±25 nm and 650±50 nm filters are a clear choice for the first two filters due to filter width and SNR. As third filter, the 800 nm longpass is preferred over the 436 nm bandpass because of the relation of the $\mathrm{HbO_2}$ and Hb curves in the range as indicated in figure 1. The 650 nm bandpass already includes a wide spectral range where Hb extinction is significantly higher than $\mathrm{HbO_2}$ extinction. Thus, a second range with this order is less important because oximetry requires ranges with opposite extinction coefficients. The 475 nm range is narrower and less distinctive than the 650 nm one, thus we consider it more important to have an additional range having the same extinction coefficient order as the 475 nm range, which results in choosing the 800 nm one.

4. SKIN SEGMENTATION

Figure 3 shows an exemplary image as recorded in the feasibility study. It results from a calibrated merge of the three camera views. Besides the desired target, which in this case is the person's hand, additional cluttered background is included in the image. To ignore impacts of the background on the measurements, vital signs should only be extracted from the target tissue region. A semantic segmentation of the image content serves this purpose. We rely on a state-of-the-art CNN-based approach to solve this task. Pre-trained networks are inapplicable because of the custom spectral ranges of the proposed sensor device. Consequently, a segmentation dataset consisting of 155 images focusing on hands and forearms is recorded. A fixed random set of 8 images is used for validation, the remaining ones are used for training. Refer to figure 3 for annotated examples.

For segmentation, an FCN-8s[13] is employed. To benefit as much as possible from the RGB pretrained VGG16 network[14] which serves as base for FCN-8s, the three recorded spectral images are assigned as adequate as possible to the RGB input channels of the network. The selected assignment is: 650 nm $\mapsto$ R, 800 nm $\mapsto$ G, 475 nm $\mapsto$ B.

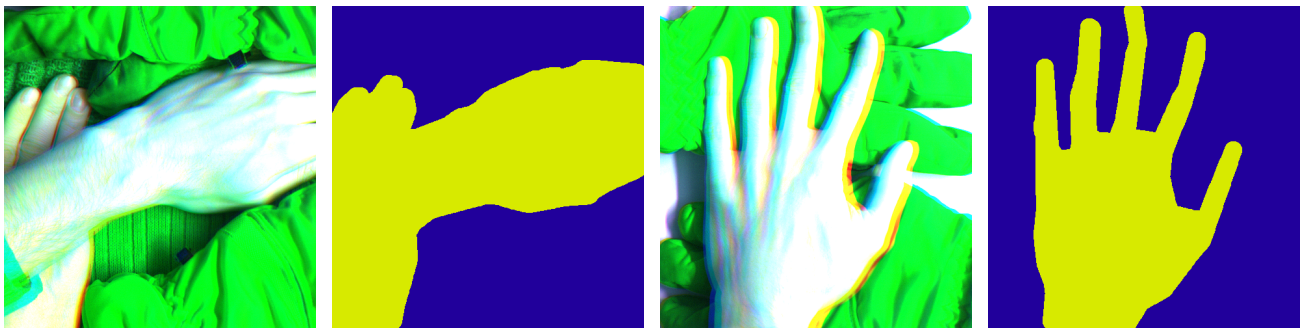

Figure 3. Examples of the recorded skin segmentation dataset including their annotated segmentation.

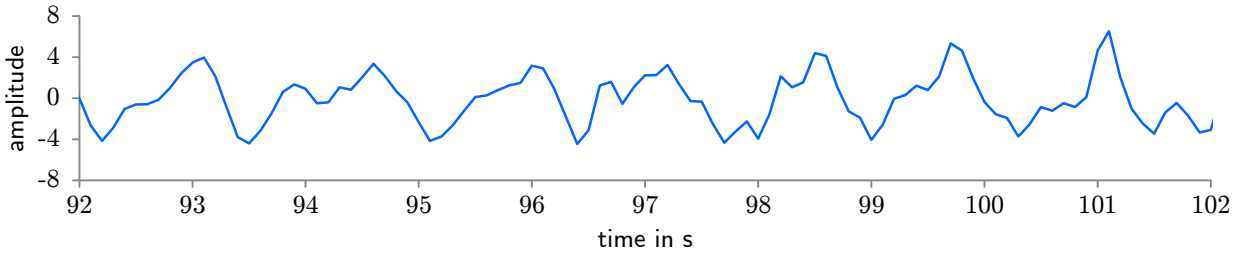

Figure 4. 10 s time series for the preprocessed F_{475} signal.

The trained segmentation network generates a foreground/background map for each frame which is used to determine the measurement signal only from the tissue parts of the image. To stabilize segmentation jitter, the foreground is morphologically closed and a moving average strategy is applied to the segmentation mask for temporal stability between frames. The signal is extracted with the stabilized segmentation by averaging all foreground pixel values per channel resulting in three raw signals $\hat{f}_i, i \in \{475, 650, 800\}$. In addition, the raw background signals $\hat{b}_i$, defined as mean of all background pixels, are also collected. These will be used in the next section to compensate for background effects, such as remaining frequency artifacts from artificial light.

5. PULSE OXIMETRY

Pulse oximetry signal processing follows a standard approach consisting of three base parts. After a signal preprocessing, the heart beat rate and the oxygen saturation are extracted from the signals in a separate part.

Preprocessing. Each foreground and background signal is preprocessed by according low- and high-pass filtering, resulting in clean signals f_i and b_i. Lower and higher cutoff-frequencies are 0.5 and 4 Hz respectively which is suitable for heart beat rates between 30 and 240 bpm. Then the difference $d_i = f_i - b_i$ of the filtered foreground and background signals serves for further analysis. Figure 4 includes an example for the preprocessed signal d_{475}.

Heart beat rate. Due to being the signal with the least noise affection, the pulse is determined from f_{475}. We use a moving 5 second window to determine the heart beat rate. This time frame allows for stable estimation as well as reasonably fast adaptation to heart beat rate changes. Updates are made every second. For each window, a Fourier analysis is performed and the heart beat rate is determined by the stable maximum in the Fourier spectrum F_{475} after a non-maximum suppression. If a stable prior heart beat is available, preference for maximums near the previous rate is given by a weak Gaussian windowing of the spectrum. As signal stabilization options, background signal subtraction on signal and spectrum level are considered, where d_{475} and $\widetilde{D}_{475} = F_{475} - B_{475}$ serve as base for heart rate estimation.

Oxygen saturation. Following the *ratio of ratios* method[10] , the oxygen saturation is obtained by

$$\mathrm{SpO}_2 = A - B\,\frac{d_i^{ac}/d_i^{dc}}{d_j^{ac}/d_j^{dc}}\,. \tag{2}$$

A and B denote calibration parameters, d_i^{ac} is the changing and d_i^{dc} the constant signal part. The signal parts d_i and d_j have to originate from two appropriate different spectral ranges. In case of the proposed sensor device, two SpO_2 measurements

Figure 5. Qualitative segmentation results on the test data.

can be made based on the three channels: d_{650} serves as base signal and is combined once with each of the other two spectral signals according to equation (2). The time window in this case is 20 s.

6. EXPERIMENTS

The test recordings are performed with the proposed system as described in section 3. The reference measurements for heart beat rate and SpO_2 are performed with a commercial finger tip pulse oximeter which is connected to the measurement computer for synchronized data acquisition. The setup was designed with regard to a typical monitoring scenario where a person's vital parameters should be observed over a longer period of time.

6.1 Skin Segmentation

The FCN-8s model is trained for 66,000 iterations on the recorded training dataset. The selected network input image resolution is 480×495 pixels. This results from half the resolution after calibrating the cameras' views to each other. Segmentation accuracy is measured by the Jaccard-index (intersection over union, IoU) which denotes how well predicted and actual segmentation overlap. We achieve a IoU of 0.967 on the test samples. Consequently, hardly any errors can be found in the resulting segmentation masks as shown in figure 5. Shadows are segmented robustly as background, only the small gaps between fingers lead to minor artifacts in some cases. Note that minor irregularities are irrelevant due to signal extraction by averaging. Thus, minor artifacts are dominated in the resulting signal by the much bigger amount of correctly segmented image area which then dominates the extracted signal.

6.2 Heart Beat Rate

We compare several ways to incorporate the background signal in the signal processing. The baseline consists of analyzing simply the foreground signal f_{475}. The second option analyzes the difference d_{475} between foreground and background signal. The last option addresses the issue in the frequency domain by subtracting the background spectrum B_{475} from the foreground spectrum F_{475}. To show the benefit of the proposed foreground skin segmentation, it is compared with a static foreground extraction method, which simply assumes that pixels near the image border belong to the background and a rectangular region in the image to the foreground. Due to the simple camera setup in the experiments, this crude assumption is roughly true and a valid baseline approach. Nevertheless, there is a significant part of background within the foreground region. In more complex setups where the baseline assumption is offended, the difference between the baseline and the dynamic segmentation is expected to grow. Table 2 lists the observed mean absolute difference (MAD) between the finger clip oximeter and the proposed method. It indicates a clear advantage of the segmentation compared with the baseline. Furthermore, background signal subtraction is beneficial in both cases. However, we suspect the spectral subtraction being less effective in the segmentation setup to be caused by no background being included in the segmented foreground region. Thus, the clear spectral disturbances of the background are already excluded and further effects cannot be covered as well with spectral subtraction.

Table 2. Difference of heart beat rate estimation in comparison to contact-based finger clip pulse oximeter.

method	**MAD**	
	baseline	segmentation
only foreground f_{475}	2.8	1.5
signal background subtraction d_{475}	2.0	1.3
spectral background subtraction $\widetilde{D}_{475}$	1.7	4.5

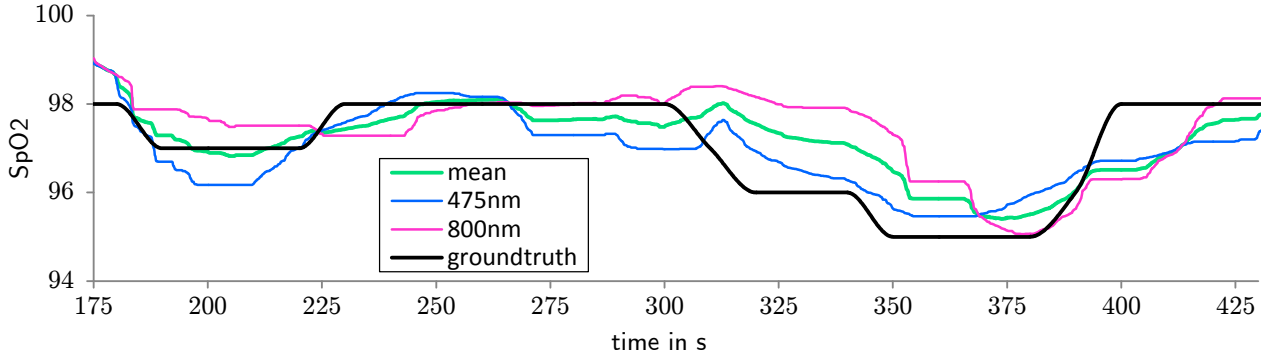

Figure 6. 250 s of finger clip SpO_2 versus both contactlessly measured values which are based on 475 nm and 800 nm signal. Both use 650 nm signal as second signal. Averaging both measurements shows benefits regarding stability.

6.3 Oxygen Saturation

First, parameters A and B are calibrated individually for both SpO_2 measurements. Then, the SpO_2 value is estimated and shown for an example sequence in figure 6. Comparison with the signal from the contact-based finger clip sensor show two aspects:

- The contactlessly measured SpO_2 signals reflect the increases and decreases in the reference SpO_2 value.
- There appears to be a measurement delay compared with the reference signal. We suspect more extensive filtering to be the reason, which is required in the proposed approach due to the worse signal quality compared with the contact-based concept.

7. CONCLUSION AND DISCUSSION

The proposed contactless pulse oximetry concept with heart rate and oxygen saturation measurement promises improvements in care situations where skin contact might cause problems. This includes pre-term infants in neonatal care as well as burn victims or highly infectious patients. The challenges of contactless measurement are addressed by a specifically designed multispectral sensor unit where the combination of three spectral ranges is optimized with regard to the signal-to-noise ratio. While our current sensor unit consists of three separate monochrome cameras, each mounted with an according optical filter, a production system can easily combine the filters directly with the sensor chip, making it no more complex than a single RGB-camera.

Furthermore, according image processing strategies including a CNN-based foreground segmentation for precise measurement and background signal estimation are incorporated. Segmentation as well as background signal subtraction are proven to increase the accuracy of the measurements.

8. ACKNOWLEDGMENT

This study was supported by the German Ministry of Education and Research (BMBF) as part of the TRICORDER program under grant no. 13N13725.

REFERENCES

[1] Kwon, S., Kim, H., and Park, K. S., "Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone," in [*IEEE Conference on Engineering in Medicine and Biology Society*], 2174–2177, IEEE (2012).

[2] Haque, M. A., Nasrollahi, K., and Moeslund, T. B., "Estimation of Heartbeat Peak Locations and Heartbeat Rate from Facial Video," in [*Scandinavian Conference on Image Analysis*], 269–281, Springer (2017).

[3] Aarts, L. A., Jeanne, V., Cleary, J. P., Lieber, C., Nelson, J. S., Oetomo, S. B., and Verkruysse, W., "Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unitA pilot study," *Early human development* **89**(12), 943–948 (2013).

[4] Lugarà, P., "Current approaches to non-invasive optical oxymetry," *Clinical hemorheology and microcirculation* **21**(3), 307–310 (1999).

[5] Fantini, S., Franceschini, M.-A., Maier, J. S., Walker, S. A., Barbieri, B. B., and Gratton, E., "Frequency-domain multichannel optical detector for noninvasive tissue spectroscopy and oximetry," *Optical Engineering* **34**(1), 32–42 (1995).

[6] Humphreys, K., Ward, T., and Markham, C., "A CMOS camera-based pulse oximetry imaging system," in [*IEEE Conference on Engineering in Medicine and Biology Society*], (2005).

[7] Wieringa, F., Mastik, F., and Van der Steen, A., "Contactless multiple wavelength photoplethysmographic imaging: a first step toward SpO2 camera technology," *Annals of biomedical engineering* **33**(8), 1034–1041 (2005).

[8] Humphreys, K., Ward, T., and Markham, C., "Noncontact simultaneous dual wavelength photoplethysmography: a further step toward noncontact pulse oximetry," *Review of scientific instruments* **78**(4), 044304 (2007).

[9] Villarroel, M., Guazzi, A., Jorge, J., Davis, S., Watkinson, P., Green, G., Shenvi, A., McCormick, K., and Tarassenko, L., "Continuous non-contact vital sign monitoring in neonatal intensive care unit," *Healthcare Technology Letters* **1**(3), 87–91 (2014).

[10] Tarassenko, L., Villarroel, M., Guazzi, A., Jorge, J., Clifton, D., and Pugh, C., "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *Physiological measurement* **35**(5), 807 (2014).

[11] Kong, L., Zhao, Y., Dong, L., Jian, Y., Jin, X., Li, B., Feng, Y., Liu, M., Liu, X., and Wu, H., "Non-contact detection of oxygen saturation based on visible light imaging device using ambient light," *Optics express* **21**(15), 17464–17471 (2013).

[12] Verkruysse, W., Bartula, M., Bresch, E., Rocque, M., Meftah, M., and Kirenko, I., "Calibration of contactless pulse oximetry," *Anesthesia and analgesia* **124**(1), 136 (2017).

[13] Long, J., Shelhamer, E., and Darrell, T., "Fully Convolutional Models for Semantic Segmentation," in [*IEEE Conference on Computer Vision and Pattern Recognition*], (2015).

[14] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," in [*International Conference on Learning Representations*], (2015).