IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 26, 2020, accepted July 5, 2020, date of publication July 13, 2020, date of current version July 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3008687

Fusion-Based Approach for Respiratory Rate Recognition From Facial Video Images

MARC-ANDRÉ FIEDLER, MICHAŁ RAPCZYŃSKI, AND AYOUB AL-HAMADI
Neuro-Information Technology Group, Institute for Information Technology and Communications, Otto von Guericke University Magdeburg, 39106 Magdeburg, Germany

Corresponding author: Marc-André Fiedler (marc-andre.fiedler@ovgu.de)

This work was supported by the Federal Ministry of Education and Research of Germany (BMBF) within the Zwanzig20 Alliance 3Dsensation Consortium (Project RoboAssist and Project HuBa) under Grant 03ZZ0448L and Grant 03ZZ0470.

ABSTRACT The respiratory rate is an important vital parameter that provides information about persons' physical condition. In clinical practice it is currently only monitored using contact-based techniques, which can have negative effects on patients. In this study, a new algorithm for remote respiratory rate recognition is presented using photoplethysmographic signals derived from facial video images in the visible light spectrum. The effects of different implementation steps in the presented algorithm are investigated in order to optimize the approach and gain new findings in this research field. In addition, a detailed examination of already implemented procedures is performed and the results are compared on two different databases. We show that by fusing the results of seven different respiratory-induced modulations in combination with other processing steps, very good estimates for the respiratory rate on both moving and non-moving data are achieved. The obtained detection rates of 72.16 % and 87.68 % are significantly higher than those of the best comparison algorithm with 37.37 % and 59.13 %. The comparison algorithms developed so far are not competitive with the newly designed method, especially for video recordings involving persons in motion. This paper provides important new findings in the field of facial video-based respiratory rate recognition for the research community. A new method has been created that delivers significantly better estimates of the respiratory rate than previously developed techniques.

INDEX TERMS Facial videos, non-contact monitoring, remote photoplethysmography, remote PPG, respiratory rate, visible light spectrum, vital signs.

I. INTRODUCTION

The respiratory rate (RR) is an important diagnostic parameter that can provide information about persons' physical condition, notably because it contains prognostic information and can point to initial indications of a later case of illness [1]. Therefore, it serves hospitals as a highly sensitive value which is capable of mapping a patient's state of health [2].

It is often measured via the sensors of an electrocardiogram on the test person's upper body or via photoplethysmography (PPG) techniques, e.g. a finger pulse oximeter. Both signals contain the modulation of respiration activity, which makes it possible to derive the target RR from them [3]. Respiration is linked to heart rate (HR) mainly through the natural phenomenon of respiratory sinus arrhythmia (RSA). This causes an increase in heart frequency during inhalation and a decrease during exhalation [4]. Spirometry, pneumography or a chest belt are also used for measurement of RR.

Thus all conventionally used systems have in common that their sensors must be attached directly to the person's body. This can have several negative effects on the health [5]–[7] and well-being [8], [9] of the person being monitored. In addition, they are only partially suitable for long-term monitoring or early detection of disease symptoms [9].

In 2008, Verkruysse *et al.* [10] were able to show that PPG signals can be captured non-invasively from a distance using a digital camera. This opened up new possibilities for the acquisition and monitoring of vital parameters, which extended the classic concept of the original photoplethysmography by a camera-based approach using photo sensors such as CCD sensors [11].

This new approach allows to significantly increase the available measurement periods [12], enabling early detection and prevention of a wide range of diseases. As a result,

The associate editor coordinating the review of this manuscript and approving it for publication was Yuan-Pin Lin.

a diagnostic device with great medical benefits can be created. The need for such a system and ubiquitous health surveillance is also rising due to chronic diseases and an ageing population [8].

Due to this potential, research in the area has been heavily intensified in the last few years. Researchers attempted to improve the accuracy of vital parameter estimates using various methods such as principal component analysis (PCA) [13], independent component analysis (ICA) [14], single-channel ICA (SCICA) [15], auto-regressive (AR) models [9], wavelet filtering [16] or a combination of ICA and AR models [17]. Also, methods for the conversion of color channels into higher quality PPG signals have been developed, such as an adaptive green red difference color model [18], a chrominance-based method [19] or the use of hue channels after an HSV color space transformation [20].

Most research papers which explore the extraction of camera-based physiological vital parameters limit the methods and findings to the detection of HR frequencies. If respiratory signals were derived, these were often regarded as ''by-products'' and therefore not subjected to a more precise analysis.

Additionally, there are research results from the neighbouring field of RR detection from conventional PPG signals, which do not originate from video images, but from contact PPG recording devices.

The current state of the art in RR detection is based on modulations of various signal parameters that are influenced by respiration. These respiration-induced variations are mainly the amplitudes, intensities and frequencies of the PPG signal [1], [11], [21]. For example, Hernando *et al.* [22] extract the amplitudes from the PPG signal, which are then modulated using spline interpolation. After that the signal is filtered and subjected to a frequency analysis to obtain the RR.

In this paper, the two research areas shall be brought together and an overview of already developed methods will be given. A new algorithm for the detection of the RR from video sequences of human faces was designed. PPG measurements can be realized on many parts of the body [23], but the forehead and other facial areas are particularly well suited for this purpose. Nilsson [24] were able to show that the extracted respiratory energies on the forehead were six times higher than those on the finger, because it has a high density of blood vessels and the skull is covered by a comparatively thin skin [8]. In addition, the face is exposed to occlusions much less frequently than other body regions and is therefore best suited for non-contact monitoring. This work therefore focuses on this approach.

The designed algorithm will then be tested in combination with multiple pre- and post-processing steps and its results will be compared with four other algorithms from the literature on two available databases with a large amount of signals. Thus, meaningful and generally valid results could be created. Furthermore, to the best of our knowledge, this is the first comprehensive examination and comparison of the methods developed so far on a large amount of uniform data.

Section II first explains the proposed method starting with PPG signal extraction from the videos. Then the following processing steps for RR estimation are delineated. Finally, the comparison algorithms found in the literature are described. Section III reports the used databases and experimental results, which are discussed in section IV. In section V a final conclusion is derived.

II. METHODS

This section describes first how the PPG signals are derived from the video sequences. Then our proposed algorithm for estimating the RR with is presented in detail. The processing pipeline of a single modulation is shown in Fig. 1. At the end the comparison algorithms used as a benchmark are presented.

A. PPG SIGNAL EXTRACTION FROM VIDEO SEQUENCES

The extraction of the PPG signal from video sequences can be broken down into two steps. First, a suitable Region-of-Interest (ROI) in the face has to be selected and detected over the entire duration of the video in every frame. The desired signals can then be derived from the color information of these ROIs. Often an RGB signal is first extracted and then eventually converted into a PPG signal.

1) GENERATION OF RGB SIGNAL

First a ROI must be defined in the video frame. This specific region is usually determined by a face tracking algorithm and/or facial landmarks [25]. From this areas in the face, for example the forehead [13], [26], can be calculated. Other approaches used a skin detection algorithm, applying it on the image or the detected face [19], [27].

Rapczynski *et al.* [25] comprehensively investigated the impact of ROI on contact-free HR estimation. They tested nine different algorithms on four different ROIs. The two best ROIs (skin detection, forehead) of their study were used for this paper because it can be assumed that the better the estimation of the HR works, the more probable it is that a correct RR can be derived due to the fact that the correct derivation of the pulse signal is the basis for our algorithm.

The ROI using skin detection [27] performed best. They use the lookup table approach from Jones and Rehg [28], which provides for each color pixel c the relative frequency p:

$$p(c) = \frac{n(c, X_{\text{skin}})}{n(c, X)} \tag{1}$$

where $n(c, X_{\text{skin}})$ represents the number of observations of the color c in the skin training dataset and $n(c, X)$ the number of observations of the color c in the entire training dataset. After calculating the lookup table, the relative frequencies can then be used for segmentation of skin. To avoid a hard threshold and binary clustering, they used the skin probability p_{i} for each pixel i from the total number of pixels in the image n as a weighting factor. With this percentage, the respective color value c_{i} is now included in the mean value calculation of the

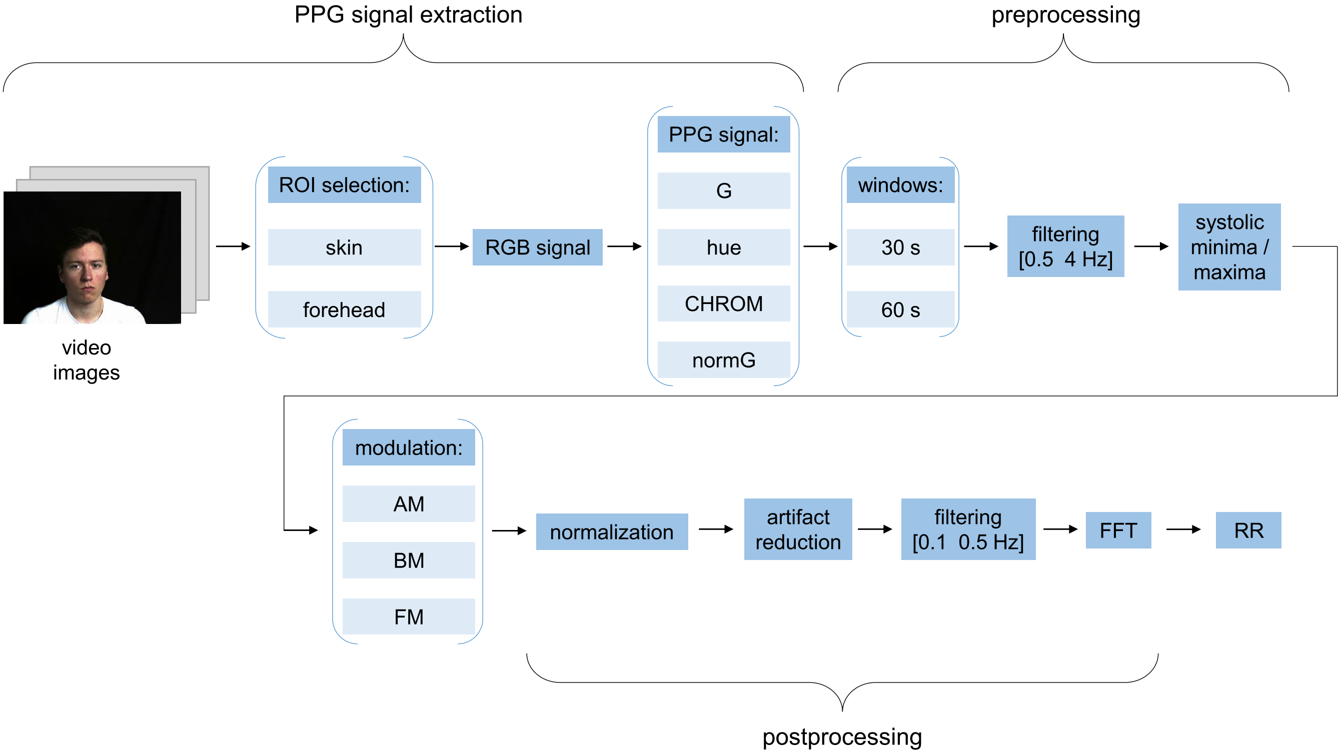


FIGURE 1. Processing pipeline of a single modulation: Main processing steps (blue boxes) with possible options (light blue).

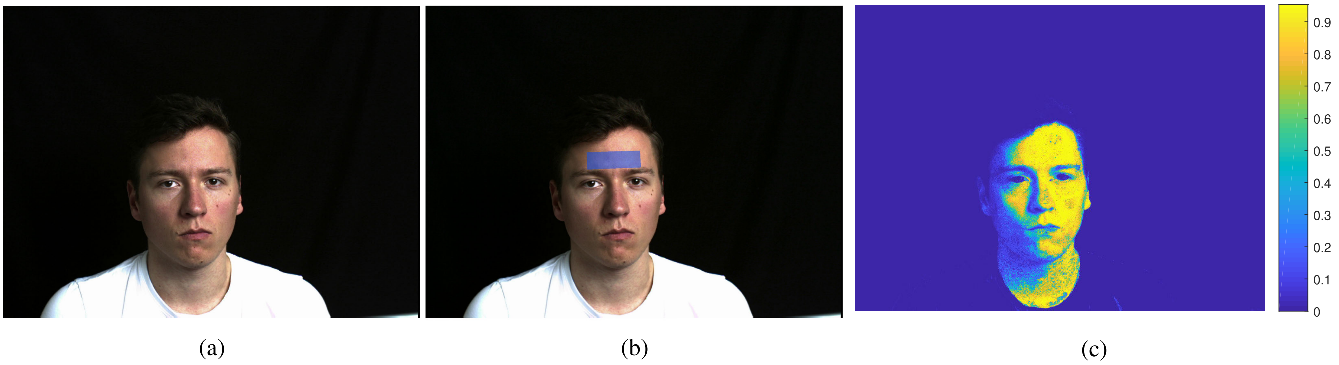


FIGURE 2. The (a) original image with (b) blue marked forehead and (c) detected skin probabilities.

PPG signal S_{PPG} for each color channel:

$$S_{\text{PPG}} = \frac{1}{\sum_{i=1}^{n} p_{\text{i}}} \cdot \sum_{i=1}^{n} p_{\text{i}} \cdot c_{\text{i}} \tag{2}$$

The forehead was used as another ROI. This is calculated from the distance between the eye corners and is placed above the eyebrows. Further information on the exact calculation of the ROI can be found in [25]. The mean value of all pixels for each color channel in the ROI for every frame is calculated to generate the RGB signal.

The forehead and skin ROI are shown in Fig. 2.

The comparison algorithm of Van Gastel [29] additionally uses a ROI of the face, which is divided into 30 subregions. It is only used by the method itself (see description in section II-C).

2) RGB TO PPG SIGNAL CONVERSION

The single RGB color channels can be used directly to calculate the RR, but this is not necessarily achieving the best results. For this reason, the research community has developed a variety of methods to make pulsatile and respiratory components in the signal better accessible. These include methods of image and signal processing. Several of these techniques were implemented and their effects investigated.

The general knowledge is that the green (**G**) channel contains the strongest photoplethysmographic information [8], [10], [18], [30]. As the wavelength of light increases, it penetrates deeper into the human tissue, enabling more accurate measurements. However, as the wavelength of light increases, the signal also becomes more susceptible to movement artifacts caused by changes in the inner tissue, such as muscle movements [31]. Besides, the absorption coefficient of the erythrocytes in the region of blue-green light is seven times higher than for red light [32]. For these reasons, G light has proven to be the best wavelength for remote PPG recordings [33]. This could also be verified in the results of this work. Therefore, the conversion methods for the generation of higher-quality PPG signals are evaluated to see whether they can contribute to an improvement in the estimation of RRs compared to the signal from G values.

Sanyal and Nundy [20] reported that a color space transformation from RGB to HSV color space produced better results for both HR and RR estimation. They transformed the individual pixels of the forehead into HSV values and then used the averaged **hue** channel. In this work, the averaged RGB signal is already calculated and is then converted afterwards.

In 2013, De Haan and Jeanne [19] introduced a chrominance-based approach (**CHROM**) to enhance robustness in HR detection from remote PPG. The reflected light captured by the camera consists of two different components according to the dichromatic reflection model. The first component is the diffuse light reflected directly from the body's surface, which contains color changes in the skin that are periodically associated with each heartbeat. The second is a reflective component that mirrors the color of the light source and contains no photoplethysmographic information. By adding a white light specular fraction an equal change of the respective diffuse reflection component could be observed on all RGB channels. This specular reflection component is now to be eliminated by the calculation of differences of the color channels, namely the chrominances. The final pulse signal S results as follows:

$$S = 3(1 - \frac{\alpha}{2})R_\mathrm{f} - 2(1 + \frac{\alpha}{2})G_\mathrm{f} + \frac{3\alpha}{2}B_\mathrm{f} \tag{3}$$

where R_f, G_f and B_f stand for the normalized filtered color channels and α is used to equalize the chrominance amplitudes. The signal S is used in this project for further processing. More detailed information can be found in the original paper.

Normalized green (**normG**) was used by Stricker *et al.* [34] and Rapczynski *et al.* [27] as a signal extraction method. The green channel is normalized by the sum of all channels to compensate for different or changing spatial and temporal light intensity levels in the video:

$$PPG_\mathrm{i} = \frac{G_\mathrm{i}}{R_\mathrm{i} + G_\mathrm{i} + B_\mathrm{i}} \tag{4}$$

The variable *i* stands for each frame in the video. Due to the significantly stronger weighting of the green value, the other two color channels play a subordinate role in the function and thus do not lead to a negative influence on the correctly extracted signal components from G. According to the theory, a PPG signal can be constructed which is superior to the pure use of G values.

Furthermore the methods PCA [13], ICA [14], JADE [35], Inverse Fast Fourier Transformation (IFFT) [36] and adaptive Green-Red-Difference (aGRD) [18] were tested. But compared to the detection performance of the algorithm when using the G channel, no improvements or even worsening occurred. For this reason, they will not be further considered in this article.

B. RR ESTIMATION

The proposed algorithm for estimating RR is based on the modulation of respiration-induced variations in the signal extracted from video sequences. The estimation of the RR is divided into preprocessing, modulation, postprocessing and fusion of results.

1) PREPROCESSING

The preprocessing prepares the signals for the modulation. For the estimation of instantaneous RRs it is the norm to divide the signal into smaller sections of a certain length and to shift them for continuous calculation by a defined time interval. So far there is no consensus in the literature what length and step size of the signal window is to be regarded as optimal. Most studies use window durations between 30 and 90 seconds [1]. However, the lower limit must be at least 20 seconds, assuming that the minimum possible respiration of a normal person cannot fall below six breaths per minute (bpm), even in extreme cases. In order to ensure detection for this case, at least two complete breaths must be detectable in the signal [20].

For this paper, the main window length was set to 30 seconds. For testing and evaluation a length of 60 seconds was added as a benchmark to analyse the performance of the algorithms on longer windows. Depending on the application and use, one of these two may be the more suitable solution, e. g. 30 seconds for an accurate instantaneous estimate of the RR and 60 seconds for long-term monitoring, where short change periods should not be considered strongly. Shorter and longer window lengths are in our view not appropriate for respiration detection. The step length for both window sizes was set to ten seconds. This fulfils the criterion that at a minimal RR of six bpm at least one of them is completely captured.

The individual PPG signal windows are first filtered with a second order zero-phase Butterworth bandpass. A lower cut-off frequency of 0.5 Hz and an upper cut-off frequency of 4 Hz are used. Atfer that all systolic maxima (max) and minima (min) are detected.

2) MODULATIONS

The respiration signal modulates the PPG signal based on different physiological causes and can be derived using the PPG signals minima and maxima.

The amplitude modulation (AM) can be attributed to a reduced stroke volume during inhalation caused by changes in intrathoracic pressure, which causes a dropping of the pulse amplitude [37]. The baseline modulation (BM) is caused by changes in tissue blood volume. These are a result of changes in intrathoracic pressure and vasoconstriction of the arteries during inhalation when blood flows into the veins [38]. The frequency modulation (FM) is caused by the natural phenomenon of RSA (see section I). This causes an increase in the HR during inhalation and a decrease during exhalation [4].

The signal parameters amplitude, baseline and frequency are processed using three different modulations to estimate the RR. In total seven varieties were implemented for this purpose: one AM, three BMs, and three FMs.

The AM is based on changes in the peak-to-peak amplitude. The coordinates of the amplitudes result from the difference between the two related extreme values and the mean of their time distance:

$$y_{\text{i, amplitude}} = y_{\text{i, max}} - y_{\text{i, min}} \tag{5}$$

$$x_{\text{i, amplitude}} = \frac{x_{\text{i, max}} + x_{\text{i, min}}}{2} \tag{6}$$

where *i* is the number of the maxima and minima. This is shown exemplarily in Fig. 3.

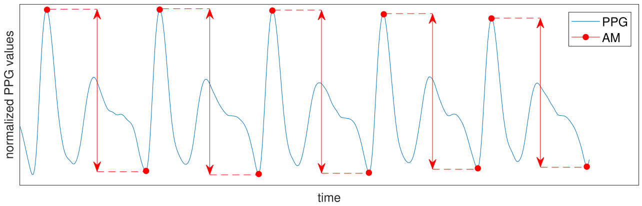


FIGURE 3. Example for the calculated amplitudes (red arrows) for the AM of the PPG signal (blue).

Three different types of implementation of BM have been developed (halfway, maxima, and minima). The y-coordinate of the first one results in:

$$y_{\text{i, baseline}} = \frac{y_{\text{i, max}} + y_{\text{i, min}}}{2} \tag{7}$$

where the x-coordinate is the point on the original signal between the two extrema closest to the calculated y-coordinate. In addition, once all maxima and once all minima were used to modulate the baseline. The different BMs are exemplarily pictured in Fig. 4.

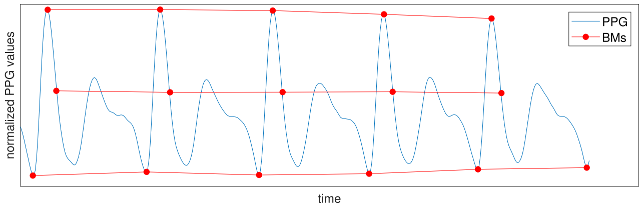


FIGURE 4. Example for the calculated baselines (red) for the BMs of the PPG signal (blue).

Different types of frequency-dependent parameters are used to calculate three different FMs (time interval maxima, time interval minima, and HR maxima). Two use the time periods between successive extreme values once between the maxima and once between the minima.

$$y_{\text{i, frequency}} = x_{\text{i+1, max}} - x_{\text{i, max}} \tag{8}$$

$$y_{\text{i, frequency}} = x_{\text{i+1, min}} - x_{\text{i, min}}. \tag{9}$$

The third FM is based on the instantaneous frequency of the HR in beats per minute. For the estimation of the HR the maxima are used:

$$y_{\text{i, frequency}} = 60 \cdot \frac{1}{x_{\text{i+1, max}} - x_{\text{i, max}}} \tag{10}$$

As x-coordinates the ones of the respective higher extrema are adopted. Fig. 5 shows the period durations of the FM exemplary.

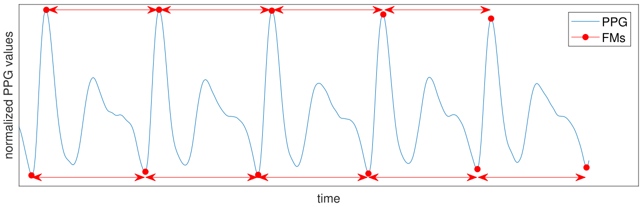


FIGURE 5. Example for the calculated period durations (red arrows) for the FMs of the PPG signal (blue).

The determined respiration-induced parameters are then linear interpolated. As frequency their original frame rate is used, which was 25 Hz for the videos of both databases.

At this point it should again be noted that the AM, the three BMs and the three FMs run completely independend from each other and have only the same pre- and post-processing steps.

3) POSTPROCESSING

Postprocessing is performed in the last steps after modulation to calculate the final RR. For this purpose, a normalization is performed, in which the mean value of the entire signal μ is subtracted from each data point y_i of the signal to remove the strong DC components from the signal, which remain in the frequency spectrum after the bandpass:

$$y_{\text{i, normalized}} = y_i - \mu \tag{11}$$

After the normalization a method for removing artifacts is performed and a differentiation of the signal is inserted. Further information can be found in subsection II-B4.

The modulated signal is filtered again with a Butterworth bandpass this time using the range of human respiratory frequencies. Therefore a lower cut-off frequency with 0.1 Hz and an upper one with 0.5 Hz is selected. There is no consensus in research which range is optimal for determining plausible respiratory frequencies, since these can also vary greatly depending on the group of persons considered [17]. The RR of adult healthy persons who are at rest and are not exposed to physical exertion is normally between 12 and 18 bpm. RR below 12 bpm are considered as slow, above 18 bpm as fast [39]. The selected range therefore allows to capture as much of the RR of adults as possible at rest and during moderate physical exertion without running into danger of capturing pulsatile instead of respiratory frequencies, which can occur with a further increase in the upper cut-off frequency.

Finally, a Fast Fourier Transformation (FFT) is used for frequency analysis and the dominant frequency is determined, which is subsequently converted into the RR with the unit bpm.

4) ARTIFACT REDUCTION

Over all modulation types a dominant frequency was occasionally found in the lower range of the bandpass filtered frequency spectrum, which overlaid the searched one of the ground truth. The wrongly recorded RRs were mostly between 6 bpm and 10 bpm, in few cases an incorrect rate up to 12 bpm was detected. Thus the frequency range of these repeatedly measured interference frequencies is in the interval from 0.1 Hz to 0.2 Hz.

This type of interference can also be found in the literature. Some researchers mention motion artifacts as the cause of this phenomenon. Moraes *et al.* [40] and Lee *et al.* [41] write in their work that movements performed by the test person often have frequencies around 0.1 Hz or a bit higher. This theory also agrees with the research results of Ram *et al.* [42], who performed a comprehensive frequency analysis of human movements. In particular, normal and natural body movements, for example during speaking or due to changes in facial features, were examined. The strongest frequency component in their evaluation was a rate of approximately 0.1 Hz, but minimally higher frequencies were also present on a large scale.

As another hypothesis for the origin of these interferences, Charlton *et al.* [1] named the Traube-Hering-Mayer waves. These waves are constant in humans at a frequency value of about 0.1 Hz. They are generated by the sympathetic nervous system and enable a conscious regulation of organ activity [43].

The exact origin and mode of formation of these frequencies could not be finally clarified, but a method was developed to suppress them in respiration detection. For this a differentiation of the signal after the normalization and before the last bandpass filtering is inserted. The gradients $y_{\text{i, gradient}}$ of the normalized signal $y_{\text{i, normalized}}$ are calculated as follows:

$$y_{\text{i, gradient}} = \frac{y_{\text{i+1, normalized}} - y_{\text{i-1, normalized}}}{2} \tag{12}$$

where i represents the respective frame or element of the signal vector.

5) FUSION OF RESULTS

The aim of the fusion of the results for the different modulations in our proposed method (**FuseMod**) is to increase the robustness of a final estimate of the RR by merging the results of several individual methods. In the literature this technique is described as a useful tool to improve the results for the detection of human RR [1], because the individual methods show a too high susceptibility to interference, especially if the test persons move additionally during monitoring. In addition, the single respiration-induced variations are differently strong depending on the person to be examined. Influencing factors are e. g. individual RR [44], gender [45] or age [6].

The applied fusion strategies realize this combination by calculating the mean or the median. In order to reduce the influence of outliers in averaging, the α-trimmed mean was additionally used. Here, the results are first sorted according to the RR estimation and then the mean of the middle three (for $\alpha = 0.29$) or five elements (for $\alpha = 0.14$) is calculated, excluding possible outliers. More detailed explanations of the α-trimmed mean can be found in [46].

C. COMPARISON ALGORITHMS

In the last few years several methods for HR detection from video images have been developed, but only a handful number of methods for RR estimation. Four of them were selected for the validation of our method. This section gives a brief overview of the re-implemented algorithms. Detailed explanations of the exact implementation and all processing steps can be found in the respective original paper. The algorithms are listed hereafter.

Poh *et al.* [14] used an ICA based on Joint Approximation Diagonalization of Eigen-matrices (JADE) [35] in combination with some other processing steps for the RR estimation. As a small modification our skin and forehead ROI (for more information see section II-A1) was utilized. This makes the performance of the actually algorithm better comparable.

Sun [15] used every RGB channel as a single input and utilized a SCICA to derive the RR. In divergence to the original paper the green channel of our skin and forehead ROI (see section II-A1) were used.

Van Gastel *et al.* [29] calculated the pixel differences of 30 subregions of the ROI to derive the weights of the linear combination for the pulse signal. These weights are then applied to pixel differences which exclusively contain respiratory frequencies and thus generate the respiratory signal. For the calculation of the weights in our re-implementation only the chrominance-based method [19] was used, because this method achieved the best results in the original paper for videos recorded in the visible light spectrum.

Sanyal and Nundy [20] used the forehead as ROI and transformed the pixel RGB values into the HSV color space, using only the hue channel and the pixels with values in the range [0 0.1] to average the PPG signal. Then the signal is bandpass filtered and the RR is determined.

III. RESULTS

Comparison of algorithms for RR estimation is difficult because each study is tested on different datasets and often uses varying statistical error measures. In order to be able to compare the previous findings in this research field with the developed method, four algorithms from the literature were re-implemented and validated on the two available databases.

In this section we first describe the used databases. Then the error measures used with the associated results are presented.

A. DATABASES

Most papers in the area of respiration detection use their own inaccessible datasets with often only a small number of signals. The comparison algorithms used for validation databases containing between 12 and 32 signals [14], [15], [20], [29]. For better reproducibility two available databases have been used in our study. But suitable databases are rare and difficult to find. In addition, they must meet certain standards for the application of non-contact vital signs monitoring, such as a low level of video compression, because the small color changes that are required for the HR detection can thus be reduced or even completely eliminated [47].

The first database we used is the **BP4D+** by Zhang *et al.* [48]. It was originally designed for emotion recognition and therefore contains 2D videos and respiration signals as well as many other ground truths. A total of 140 test persons were recorded while they each had to complete 10 different tasks. These tasks include interviews, physical activities or watching video clips. As the test persons move and speak without restrictions, the database is challenging for non-contact monitoring of vital parameters. This is also due to the fact that the database contains only natural respiratory cycles and no artificial ones. The respiration signals were recorded via a chest belt. Further details on the data and their acquisition can be found in the original paper.

As the ground truths of the respiratory signals were temporarily very strongly affected by artefacts and disturbances, some of them had to be sorted out because it was not possible to determine exact estimates for the correctly underlying RR with certainty (see Fig. 7). Therefore, the signals were preprocessed with a Butterworth bandpass with cut-off frequencies of 0.1 Hz and 0.5 Hz and then normalized to a value range of [−1 1]. The peaks and troughs were then determined using a designed peak detector. Two standard deviations were then selected as parameters for estimating the quality of the respiratory signal. Signals are rejected if the standard deviation of the difference between the peak intervals of a signal is greater than 1 second or the standard deviation of the heights of the minima of a signal is greater than 0.2. In addition, all videos shorter than 30 seconds are also rejected. This results in 269 remaining signals, which are, comparatively to other studies, a large number of signals on which the algorithms are validated.

An example for each respiratory signal class is shown below: Fig. 6 shows a remaining signal and Fig. 7 shows a rejected signal where the ground truth cannot be determined.

The second database was acquired by ourselves. The goal was to generate robust, uncompressed data with only few errors in order to use them as a good benchmark for RR detection. It includes recordings of 12 adults ages ranging from 23 to 36 years. There are 10 men and 2 women. Four videos per person were recorded with an RGB camera (model Pike F-145) from a distance of about 1.5 meters for 3 minutes per sequence with a frame rate of 25 frames per second. Four different RR scenarios were recorded for every subject. In the first scenario the spontaneous respiration of the test person is recorded. For the next three measurements, the subjects should try to follow a given respiratory pattern, which was displayed on a monitor in front of them. The respiratory frequencies for the single videos were 10, 15 and 20 bpm. Over all measurements the participants were asked not to move heavily. The reference signals were recorded via a chest belt (model NB-RSP1A) with a sample rate of 512 Hz using a trigger signal for synchronization with the camera.

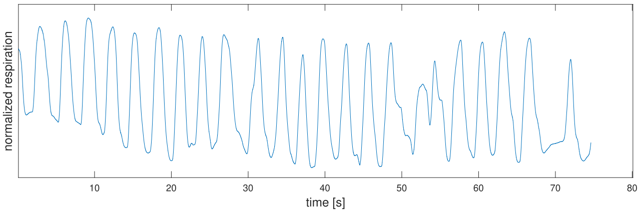


FIGURE 6. Example for a remaining respiratory signal where the ground truth can be determined.

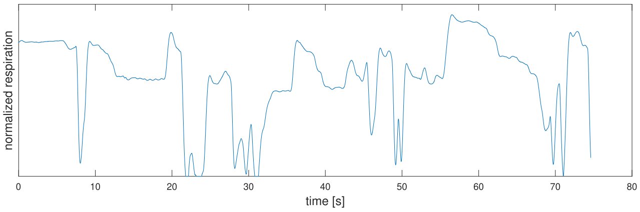


FIGURE 7. Example for a rejected respiratory signal where the ground truth cannot be determined.

Consequently, a large amount of data from a total of 318 videos, which corresponds to over 7.5 hours of recordings, could be analyzed for the validation of the algorithms, which cover both important types of conditions, including and excluding movements of the test persons.

B. ERROR MEASURES AND RESULTS

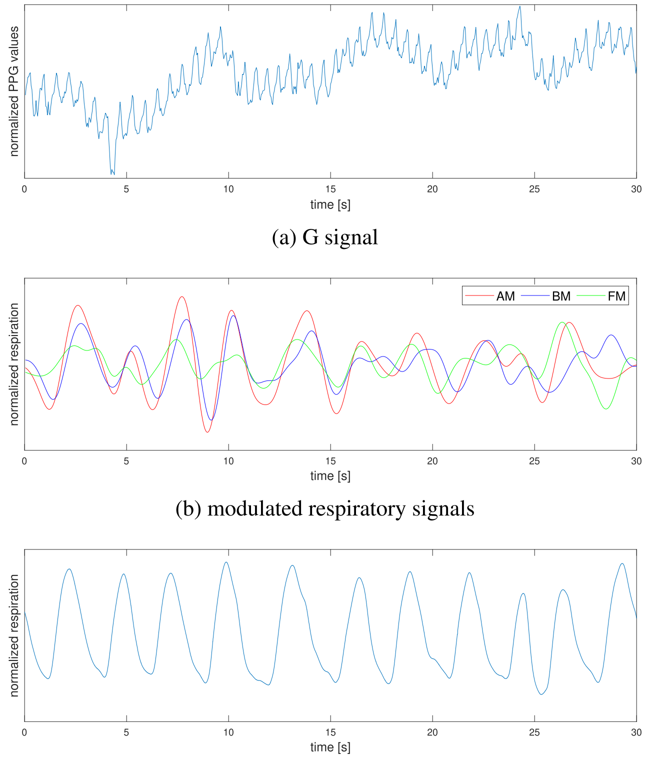
Several error measures were used to compare the accuracy of the different algorithms and implementation approaches. The detection rate (**DR**) was introduced, which describes how many percent of windows are correctly detected:

$$DR\,[\%] = \frac{n_{\text{correct windows}}}{n_{\text{all windows}}} \tag{13}$$

where n is the respective number of windows. A window is assumed to be correct if the error between detected value and ground truth is less or equal than 2 bpm. It is the most important measure in our study and was similarly proposed by Charlton *et al.* [1] as a statistical error measure for RR detection.

In addition, the mean μ and the standard deviation σ of the errors are given.

Fig. 8 shows an exemplary G signal, the resulting modulated respiratory signals (prior to FFT analysis) for each of the three modulation types (AM, BM, and FM) and the ground



(a) G signal

(b) modulated respiratory signals

(c) ground truth respiratory signal

FIGURE 8. **Exemplary illustration of (a) the G signal, (b) the resulting modulated respiratory signals (prior to FFT analysis) for each of the three modulation types and (c) the ground truth respiratory signal.**

truth respiratory signal in order to illustrate the processed signals.

An overview of the results of our FuseMod algorithm can be found in Tables 1 - 4. For this purpose, two tables are mapped for each database, each one for the skin and forehead ROI. The results are shown for the different preprocessing PPG conversions and fusion methods. The window length is fixed to 30 seconds.

Tables 5 and 6 show the comparisons of the estimation performance of the FuseMod algorithm (30s windows) with and without artifact reduction. Only the results for the dominant PPG conversion and fusion combinations with the best results are listed.

An additional error measure was calculated to determine the percentage of incorrectly detected windows in the low-frequency range, to assess the impact of the artifact reduction. The false detection rate (**FD**) indicates the number of incorrectly detected windows in percent, which have a dominant low-frequency component in the area between 0.1 Hz and 0.2 Hz compared to total number of windows:

$$FD\,[\%] = \frac{n_{\text{incorrect windows [0.1 0.2 Hz]}}}{n_{\text{all windows}}} \qquad (14)$$

where n is the respective number of windows. These unwanted artifacts are caused by movement or Traube-Hering-Mayer waves (see subsection II-B4).

Table 7 and 8 compare the results of our best FuseMod implementations for facial video-based RR recognition with those of the comparison algorithms, showing one table per database. All were tested on a consistent window length of 30 seconds for better comparison. To show the performance of the implemented algorithms in the same way as in the original papers, the results with their original window specifications were calculated. For our methods a window length of 60 seconds was additionally utilized to investigate whether the window length has an influence on the detection performance.

IV. DISCUSSION

This section discusses the achieved results. First, the different specifications of the FuseMod implementations are considered. Then the influence of artifact reduction is examined and finally the performance is benchmarked against the comparison algorithms.

A. FuseMod

For the FuseMod implementations there are numerous combinations of different parameters (see Tables 1 - 4).

Various PPG signal conversions have been tested. The results show a clear improvement in detection performance for hue, CHROM and normG compared to the G channel. This is reflected in a significant increase in the DR of up to 19 %, a further approximation of the μ to zero and a reduction of the σ. This increase is relatively constant on the BP4D+ database across all fusion types and both ROIs. For our own database, the performance of the hue channel declines compared to CHROM and normG, whereby CHROM again performs best.

Procedures which combine the color values of all three RGB channels in order to derive the final PPG signal achieve better results in our experiments. The calculation steps of all three methods (hue, CHROM and normG) have this in common. The three approaches perform similarly well with minimal advantages for CHROM, although it is also the most computation-intensive method.

Of the fusion techniques, the median performs better than the mean on both databases. This result can be attributed to the fact that the individual modulations are temporarily subject to strong variations, which are better eliminated by the median. No significant differences can be seen between 0.14-trimmed mean and 0.29-trimmed mean. Averaging of the middle results leads to a significant increase in DR compared to the mean, but cannot compete with the detection performance of the median.

When examining the influence of the ROI, clear advantages can be seen for the use of the skin algorithm over the forehead. A static ROI has certain weaknesses, as it is not possible to ensure that the calculated area contains only of human tissue. Part of the ROI can be covered by hair, a beard, glasses or headgear, which leads to interfering pixels in the calculation of the PPG signal. It is therefore useful to perform an individual assessment for each pixel to determine whether it is skin or not.

TABLE 1. Results for FuseMod (skin ROI, 30s windows) using different PPG and fusion methods on the BP4D+.

	fusion type	mean				0.14-trimmed mean				0.29-trimmed mean				median			
	PPG signal	G	hue	CHROM	normG	G	hue	CHROM	normG	G	hue	CHROM	normG	G	hue	CHROM	normG
error measures	DR [%]	53.11	60.81	63.63	61.06	56.75	65.70	66.45	65.95	54.52	65.29	65.87	65.45	62.97	**70.84**	**72.16**	**71.00**
	μ [bpm]	-0.52	-0.39	-0.22	-0.48	-0.56	-0.44	-0.25	-0.49	-0.56	-0.44	-0.25	-0.47	-0.56	-0.44	-0.26	-0.49
	σ [bpm]	3.90	3.38	3.35	3.47	3.99	3.42	3.37	3.52	4.09	3.49	3.38	3.58	4.23	3.62	3.52	3.78

TABLE 2. Results for FuseMod (forehead ROI, 30s windows) using different PPG and fusion methods on the BP4D+.

	fusion type	mean				0.14-trimmed mean				0.29-trimmed mean				median			
	PPG signal	G	hue	CHROM	normG	G	hue	CHROM	normG	G	hue	CHROM	normG	G	hue	CHROM	normG
error measures	DR [%]	46.39	59.67	56.10	52.86	49.46	**63.82**	60.25	56.93	47.22	60.50	59.75	55.85	54.19	**67.22**	**67.88**	61.99
	μ [bpm]	-0.69	-0.53	-0.34	-0.72	-0.76	-0.55	-0.35	-0.69	-0.76	-0.49	-0.34	-0.63	-0.65	-0.40	-0.28	-0.63
	σ [bpm]	4.01	3.57	3.65	3.88	4.12	3.65	3.69	3.99	4.25	3.71	3.73	4.11	4.44	3.83	3.93	4.35

TABLE 3. Results for FuseMod (skin ROI, 30s windows) using different PPG and fusion methods on our own database.

	fusion type	mean				0.14-trimmed mean				0.29-trimmed mean				median			
	PPG signal	G	hue	CHROM	normG	G	hue	CHROM	normG	G	hue	CHROM	normG	G	hue	CHROM	normG
error measures	DR [%]	51.21	56.69	65.86	67.39	56.82	61.78	74.14	72.74	56.18	62.93	**75.16**	73.76	63.18	68.41	**81.27**	**78.47**
	μ [bpm]	-1.50	-1.57	-1.13	-1.41	-1.73	-1.64	-1.16	-1.49	-1.75	-1.65	-1.15	-1.40	-1.80	-1.69	-1.15	-1.39
	σ [bpm]	3.74	3.45	2.90	3.06	3.79	3.53	2.84	3.08	3.95	3.67	2.87	3.17	4.19	3.89	3.05	3.33

TABLE 4. Results for FuseMod (forehead ROI, 30s windows) using different PPG and fusion methods on our own database.

	fusion type	mean				0.14-trimmed mean				0.29-trimmed mean				median			
	PPG signal	G	hue	CHROM	normG	G	hue	CHROM	normG	G	hue	CHROM	normG	G	hue	CHROM	normG
error measures	DR [%]	44.08	43.82	53.89	51.97	45.61	48.92	57.58	57.32	43.95	49.04	**58.34**	57.71	49.68	54.14	**63.69**	**62.55**
	μ [bpm]	-1.48	-1.90	-1.58	-1.65	-1.72	-2.06	-1.74	-1.79	-1.81	-2.10	-1.79	-1.83	-1.86	-2.10	-1.78	-1.81
	σ [bpm]	4.19	4.13	3.91	3.78	4.30	4.29	4.02	3.90	4.43	4.48	4.11	4.05	4.70	4.64	4.35	4.26

TABLE 5. Results with and without artifact reduction for FuseMod with best performing specifications using 30s window on the database BP4D+.

	artifact	error measures			
algorithm	reduction	DR [%]	μ [bpm]	σ [bpm]	FD [%]
FuseMod (hue, median)	with	70.84	-0.44	3.62	2.32
FuseMod (hue, median)	without	29.66	-7.10	5.63	52.94
FuseMod (CHROM, median)	with	72.16	-0.26	3.52	1.16
FuseMod (CHROM, median)	without	31.40	-6.93	5.65	50.37
FuseMod (normG, median)	with	71.00	-0.49	3.78	2.57
FuseMod (normG, median)	without	27.26	-7.44	5.56	54.68

TABLE 6. Results with and without artifact reduction for FuseMod with best performing specifications using 30s window on our own database.

	artifact	error measures			
algorithm	reduction	DR [%]	μ [bpm]	σ [bpm]	FD [%]
FuseMod (hue, median)	with	68.41	-1.69	3.89	0.64
FuseMod (hue, median)	without	62.17	-2.93	4.37	21.53
FuseMod (CHROM, median)	with	81.27	-1.15	3.05	0.76
FuseMod (CHROM, median)	without	67.13	-2.76	4.36	12.99
FuseMod (normG, median)	with	78.47	-1.39	3.33	0.89
FuseMod (normG, median)	without	69.17	-2.40	4.25	15.16

The difference between the two ROIs is even more significant on our own database than on the BP4D+. With the skin ROI, our non-moving data shows improvements over the BP4D+ moving data, but the opposite is the case for forehead. This can be explained by the fact that the forehead of the subjects in our database often showed occlusions by hair and at times strong light reflections of the skin pixels during the recording, causing them to become over-saturated. In addition, one test person wore a headscarf reducing the visible skin in the forehead ROI.

Thus, the dominance of the skin ROI over other static ROIs of different facial areas can be seen. These results are also consistent with the findings of Rapczynski *et al.* [25].

B. INFLUENCE OF ARTIFACT REDUCTION

Artifact reduction is an fundamental part of the algorithm we have designed. The impact of this is shown in Tables 5 and 6, which compare the recognition of the developed procedure with and without additional artifact reduction.

TABLE 7. Results for the best-performing FuseMod implementations (30s and 60s windows) and the comparison algorithms (see section II-C) (30s and the original proposed window) on the BP4D+.

			error measures		
algorithm	window	ROI	DR [%]	μ [bpm]	σ [bpm]
FuseMod (hue, median)	30 s	skin	70.84	-0.44	3.62
FuseMod (hue, median)	60 s	skin	71.40	-0.20	3.07
FuseMod (CHROM, median)	30 s	skin	**72.16**	-0.26	3.52
FuseMod (CHROM, median)	60 s	skin	71.78	-0.25	2.96
FuseMod (normG, median)	30 s	skin	71.00	-0.49	3.78
FuseMod (normG, median)	60 s	skin	68.22	-0.03	3.12
Poh [14]	30 s	skin	37.37	-3.78	5.63
Poh [14]	60 s	skin	37.01	-4.10	5.53
Poh [14]	30 s	forehead	32.45	-3.66	5.85
Poh [14]	60 s	forehead	36.26	-3.61	5.43
Sun [15]	30 s	skin	20.30	5.39	5.56
Sun [15]	34 s	skin	17.77	5.32	5.58
Sun [15]	30 s	forehead	23.07	5.27	5.56
Sun [15]	34 s	forehead	17.18	5.31	5.75
VanGastel [29]	30 s	VanGastel	22.66	0.19	9.29
VanGastel [29]	8 s	VanGastel	31.27	1.00	9.51
Sanyal [20]	30 s	skin	21.16	-6.33	4.36
Sanyal [20]	20 s	skin	17.99	-5.35	5.00
Sanyal [20]	30 s	forehead	19.09	-6.56	4.32
Sanyal [20]	20 s	forehead	12.56	-5.91	4.76

TABLE 8. Results for the best-performing FuseMod implementations (30s and 60s windows) and the comparison algorithms (see section II-C) (30s and the original proposed window) on our own database.

			error measures		
algorithm	window	ROI	DR [%]	μ [bpm]	σ [bpm]
FuseMod (hue, median)	30 s	skin	68.41	-1.69	3.89
FuseMod (hue, median)	60 s	skin	78.47	-1.08	3.07
FuseMod (CHROM, median)	30 s	skin	81.27	-1.15	3.05
FuseMod (CHROM, median)	60 s	skin	**87.68**	-0.65	2.33
FuseMod (normG, median)	30 s	skin	78.47	-1.39	3.33
FuseMod (normG, median)	60 s	skin	84.87	-0.72	2.46
Poh [14]	30 s	skin	55.54	0.85	4.89
Poh [14]	60 s	skin	59.13	0.88	4.42
Poh [14]	30 s	forehead	47.01	1.18	5.35
Poh [14]	60 s	forehead	50.55	1.09	5.03
Sun [15]	30 s	skin	18.22	4.15	8.90
Sun [15]	34 s	skin	16.91	3.63	8.66
Sun [15]	30 s	forehead	25.86	3.33	8.83
Sun [15]	34 s	forehead	16.51	3.53	8.98
VanGastel [29]	30 s	VanGastel	35.16	5.15	9.01
VanGastel [29]	8 s	VanGastel	32.27	6.34	9.26
Sanyal [20]	30 s	skin	52.99	-2.31	4.15
Sanyal [20]	20 s	skin	41.66	0.21	3.99
Sanyal [20]	30 s	forehead	53.12	-2.03	4.25
Sanyal [20]	20 s	forehead	34.09	-0.06	4.65

Especially the database BP4D+ contains low-frequency interference frequencies from movements or Traube-Hering-Mayer waves on a large scale, which overlay the searched ground truth frequency with their energy in the spectrum and cannot be filtered without removing plausible respiratory frequencies from the filter margins. This can be seen from the μ which is heavily biased towards negative errors (see Tables 5 and 6). In addition, the FD rates show that without artifact reduction more than half of the windows are detected incorrectly and instead are assigned to a frequency in the lower range of the observed spectrum.

By inserting an additional step for differentiating the signal, this percentage can be reduced drastically to approximately 2 %. This phenomenon is thus effectively eliminated increasing the DR by 40 % and resulting μ close to zero.

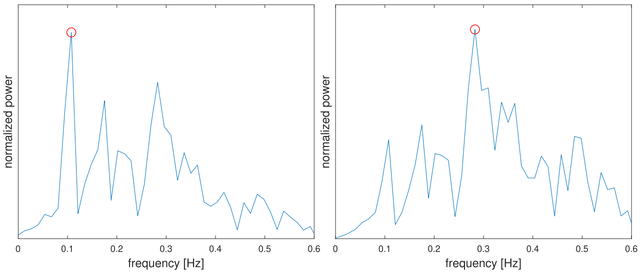
Our database contains less motion-induced interference frequencies, but the DR can be also increased between 6 % and 14 % depending on the PPG signal. The false detections in the lower frequency region fall below 1 %.

For this reason, the artifact reduction procedure has been included as standard in our implementations for RR recognition.

The influence of artifact reduction on the frequency spectrum is illustrated in Fig. 9 using an AM as an example. The issue is analog for BMs and FMs.

C. EVALUATION VERSUS COMPARISON ALGORITHMS

To validate the designed method for RR recognition, four other algorithms from the literature were re-implemented.



(a)

(b)

FIGURE 9. Comparison of the frequency spectra after an AM (a) without and (b) with artifact reduction with a ground truth frequency of 0.28 Hz.

For comparison a consistent window length of 30 seconds and a step size of 10 seconds is used.

The results of all algorithms are shown in Tables 7 and 8. An additional window length, as proposed in the respective original paper was also calculated. A 60-second window was also used for our implementations, as we considered that to be another suitable length.

The results of the comparison algorithms on the BP4D+ database show how challenging the data is for correct RR detection. This can be explained by the fact that the test persons had no restrictions and move strongly. The method by Poh *et al.* [14] still scores best out of the comparison algorithms, as it detects 37 % of the windows correctly. The others only achieve more than 20 %.

All algorithms can increase their DR on our database with the exception of Sun [15], which remains at the same level. This was expected since the signals in the new dataset are less affected by movements. Especially Sanyal and Nundy [20] has to be highlighted at this point, which increases DR by more than 30 % and thus performs almost as well as the best comparison algorithm by Poh *et al.* [14].

For the comparison algorithms, the skin ROI works again better than the forehead ROI.

The changed window length has no negative effect on the comparison algorithms, as their performance with original window parameters is not higher. An exception is Van Gastel [29] with a drop of the DR on BP4D+ of almost 10 %.

The FuseMod algorithm increases its estimation rates on our database when using 60-second windows across all PPG signals and reaches a DR of 87.68 % for CHROM.

In summary, the comparison algorithms are not competitive either on moving or non-moving data with the new developed methods, which achieve a DR of over 70 % on the BP4D+ and over 80 % on our database. In addition, the method is robust to changes in window length, which does not result in a decrease in recognition performance. This is illustrated by the example of 60-second windows, where the values of the error measures remain almost identical or even increase to those at 30 seconds.

V. CONCLUSION

We have shown that the photoplethysmographic information in the PPG signal can be increased by the conversion methods hue, CHROM and normG compared to the green channel. In addition, an efficient technique has been developed that greatly reduces artifacts caused by movement and Traube-Hering-Mayer waves using the differentiation of the signal prior to FFT analysis. Furthermore, the median proved to be the best fusion technique for the results of the single modulations. It was also shown that the use of shorter windows for moving data and longer windows for non-moving data is beneficial. Overall, 30-second windows turned out to be the most appropriate length. The proposed method was evaluated on two available databases with a total of 318 videos. One database contains videos that are strongly characterized by movements of the subjects and the other one consists of videos of persons who were requested not to move heavily. The achieved detection rates of our FuseMod algorithm with 72.16 % and 87.68 % far exceed the rates of the best-performing comparison algorithm of 37.37 % and 59.13 %. This represents an important advance for the research field of non-contact monitoring of respiratory rates. By investigating various algorithms, PPG signal conversion approaches, artifact reduction techniques, fusion procedures, as well as other pre- and post-processing steps a comprehensive examination in the field of facial video-based respiratory rate recognition in the visible light spectrum has been realized, including both existing and newly developed methods.

REFERENCES

[1] P. H. Charlton, D. A. Birrenkott, T. Bonnici, M. A. F. Pimentel, A. E. W. Johnson, J. Alastruey, L. Tarassenko, P. J. Watkinson, R. Beale, and D. A. Clifton, "Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review," *IEEE Rev. Biomed. Eng.*, vol. 11, pp. 2–20, 2018.

[2] M. Cretikos, J. Chen, K. Hillman, R. Bellomo, S. Finfer, and A. Flabouris, "The objective medical emergency team activation criteria: A case–control study," *Resuscitation*, vol. 73, no. 1, pp. 62–72, Apr. 2007.

[3] K. V. Madhav, M. R. Ram, E. H. Krishna, N. R. Komalla, and K. A. Reddy, "Estimation of respiration rate from ECG, BP and PPG signals using empirical mode decomposition," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, May 2011, pp. 1–4.

[4] L. Zhao, S. Reisman, and T. W. Findley, "Derivation of respiration from electrocardiogram during heart rate variability studies," in *Proc. Comput. Cardiol.*, Sep. 1994, pp. 53–56.

[5] E. Sazonov, *Wearable Sensors: Fundamentals, Implementation and Applications*. New York, NY, USA: Academic, 2014.

[6] P. H. Charlton, T. Bonnici, L. Tarassenko, D. A. Clifton, R. Beale, and P. J. Watkinson, "An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram," *Physiol. Meas.*, vol. 37, no. 4, pp. 610–626, 2016.

[7] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Opt. Express*, vol. 18, no. 10, pp. 10762–10774, May 2010.

[8] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, and H. Nazeran, "A review on wearable photoplethysmography sensors and their potential future applications in health care," *Int. J. Biosensors Bioelectron.*, vol. 4, no. 4, p. 195, 2018.

[9] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *Physiol. Meas.*, vol. 35, no. 5, pp. 807–831, 2014.

[10] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Express*, vol. 16, no. 26, p. 21434, Dec. 2008.

[11] P. H. Charlton, M. Villarroel, and F. Salguiero, "Waveform analysis to estimate respiratory rate," in *Proc. 2nd Anal. Electron. Health Records*, 2016, pp. 377–390.

[12] P. H. Charlton, T. Bonnici, L. Tarassenko, J. Alastruey, D. A. Clifton, R. Beale, and P. J. Watkinson, "Extraction of respiratory signals from the electrocardiogram and photoplethysmogram: Technical and physiological determinants," *Physiol. Meas.*, vol. 38, no. 5, pp. 669–690, Mar. 2017.

[13] M. Lewandowska and J. Nowak, "Measuring pulse rate with a Webcam," *J. Med. Imag. Health Informat.*, vol. 2, no. 1, pp. 87–92, Jan. 2012.

[14] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.

[15] Y. Sun, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise," *J. Biomed. Opt.*, vol. 16, no. 7, Jul. 2011, Art. no. 077010.

[16] F. Bousefsaf, C. Maaoui, and A. Pruski, "Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 568–574, 2013.

[17] M. Villarroel, S. Davis, P. Watkinson, A. Guazzi, K. Mccormick, L. Tarassenko, J. Jorge, A. Shenvi, and G. Green, "Continuous non-contact vital sign monitoring in neonatal intensive care unit," *Healthcare Technol. Lett.*, vol. 1, no. 3, pp. 87–91, Jan. 2014.

[18] L. Feng, L.-M. Po, X. Xu, Y. Li, and R. Ma, "Motion-resistant remote imaging photoplethysmography based on the optical properties of skin," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 5, pp. 879–891, Oct. 2015.

[19] G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based RBBG," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878–2886, Feb. 2013.

[20] S. Sanyal and K. K. Nundy, "Algorithms for monitoring heart rate and respiratory rate from the video of a User's face," *IEEE J. Transl. Eng. Health Med.*, vol. 6, pp. 1–11, 2018.

[21] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont, "Multiparameter respiratory rate estimation from the photoplethysmogram," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 7, pp. 1946–1953, Feb. 2013.

[22] A. Hernando, M. D. Pelaez, M. T. Lozano, M. Aiger, E. Gil, and J. Lazaro, "Finger and forehead PPG signal comparison for respiratory rate estimation based on pulse amplitude variability," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 2076–2080.

[23] L. Nilsson, T. Goscinski, S. Kalman, L.-G. Lindberg, and A. Johansson, "Combined photoplethysmographic monitoring of respiration rate and pulse: A comparison between different measurement sites in spontaneously breathing subjects," *Acta Anaesthesiologica Scandinavica*, vol. 51, no. 9, pp. 1250–1257, Aug. 2007.

[24] L. M. Nilsson, "Respiration signals from photoplethysmography," *Anesthesia Analgesia*, vol. 117, no. 4, pp. 859–865, 2013.

[25] M. Rapczynski, P. Werner, F. Saxen, and A. Al-Hamadi, "How the region of interest impacts contact free heart rate estimation algorithms," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2027–2031.

[26] T. Blocher, J. Schneider, M. Schinle, and W. Stork, "An online PPGI approach for camera based heart rate monitoring using beat-to-beat detection," in *Proc. IEEE Sensors Appl. Symp. (SAS)*, Mar. 2017, pp. 1–6.

[27] M. Rapczynski, P. Werner, and A. Al-Hamadi, "Continuous low latency heart rate estimation from painful faces in real time," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 1165–1170.

[28] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *Int. J. Comput. Vis.*, vol. 46, pp. 81–96, Jan. 2002.

[29] M. van Gastel, S. Stuijk, and G. de Haan, "Robust respiration detection from remote photoplethysmography," *Biomed. Opt. Express*, vol. 7, no. 12, p. 4941, Dec. 2016.

[30] L. M. Ruiz, A. Manzo, E. Casimiro, E. Cardenas, and R. Gonzalez, "Heart rate variability using photoplethysmography with green wavelength," in *Proc. IEEE Int. Autumn Meeting Power, Electron. Comput. (ROPEC)*, Nov. 2014, pp. 1–5.

[31] J. Spigulis, L. Gailite, A. Lihachev, and R. Erts, "Simultaneous recording of skin blood pulsations at different vascular depths by multiwavelength photoplethysmography," *Appl. Opt.*, vol. 46, no. 10, p. 1754, 2007.

[32] M. V. Volkov, N. B. Margaryants, A. V. Potemkin, M. A. Volynsky, I. P. Gurov, O. V. Mamontov, and A. A. Kamshilin, "Video capillaroscopy clarifies mechanism of the photoplethysmographic waveform appearance," *Sci. Rep.*, vol. 7, no. 1, Dec. 2017, Art. no. 13298.

[33] N. Sviridova, T. Zhao, K. Aihara, K. Nakamura, and A. Nakano, "Photoplethysmogram at green light: Where does chaos arise from?" *Chaos, Solitons Fractals*, vol. 116, pp. 157–165, Oct. 2018.

[34] R. Stricker, S. Muller, and H.-M. Gross, "Non-contact video-based pulse rate measurement on a mobile service robot," in *Proc. 23rd IEEE Int. Symp. Robot Hum. Interact. Commun.*, Aug. 2014, pp. 1056–1062.

[35] J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Comput.*, vol. 11, no. 1, pp. 157–192, 1999.

[36] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Robust heart rate from fitness videos," *Physiol. Meas.*, vol. 38, no. 6, p. 1023, 2017.

[37] D. J. Meredith, D. Clifton, P. Charlton, J. Brooks, C. W. Pugh, and L. Tarassenko, "Photoplethysmographic derivation of respiratory rate: A review of relevant physiology," *J. Med. Eng. Technol.*, vol. 36, no. 1, pp. 1–7, Mar. 2012.

[38] M. Nitzan, I. Faib, and H. Friedman, "Respiration-induced changes in tissue blood volume distal to occluded artery, measured by photoplethysmography," *J. Biomed. Opt.*, vol. 11, no. 4, 2006, Art. no. 040506.

[39] C. Becker, S. Achermann, M. Rocque, I. Kirenko, A. Schlack, T. Dreher-Hummel, T. Zumbrunn, R. Bingisser, and C. H. Nickel, "Camera-based measurement of respiratory rates is reliable," *Eur. J. Emergency Med.*, vol. 25, no. 6, pp. 416–422, Dec. 2018.

[40] J. Moraes, M. Rocha, G. Vasconcelos, J. V. Filho, V. de Albuquerque, and A. Alexandria, "Advances in photopletysmography signal analysis for biomedical applications," *Sensors*, vol. 18, no. 6, p. 1894, Jun. 2018.

[41] H.-W. Lee, J.-W. Lee, W.-G. Jung, and G.-K. Lee, "The periodic moving average filter for removing motion artifacts from PPG signals," *Int. J. Control, Automat. Syst.*, vol. 5, no. 6, pp. 701–706, 2007.

[42] M. R. Ram, K. V. Madhav, E. H. Krishna, N. R. Komalla, and K. A. Reddy, "A novel approach for motion artifact reduction in ppg signals based on AS-LMS adaptive filter," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 5, pp. 1445–1457, Dec. 2012.

[43] L.-G. Lindberg, H. Ugnell, and P. K. Ãberg, "Monitoring of respiratory and heart rates using a fibre-optic sensor," *Med. Biol. Eng. Comput.*, vol. 30, no. 5, pp. 533–537, 1992.

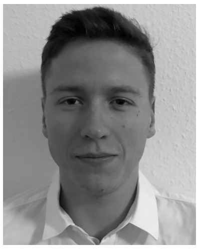
[44] J. Lãzaro, P. Laguna, Y. Nam, K. Chon, and E. Gil, "Respiratory rate influence in the resulting magnitude of pulse photoplethysmogram derived respiration signals," in *Proc. Comput. Cardiol.*, Sep. 2014, pp. 289–292.

[45] J. Li, J. Jin, X. Chen, W. Sun, and P. Guo, "Comparison of respiratory-induced variations in photoplethysmographic signals," *Physiol. Meas.*, vol. 31, no. 3, pp. 415–425, Oct. 2010.

[46] J. Bednar and T. Watt, "Alpha-trimmed means and their relationship to median filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 1, pp. 145–153, Feb. 1984.

[47] M. Rapczynski, P. Werner, and A. Al-Hamadi, "Effects of video encoding on camera-based heart rate estimation," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 12, pp. 3360–3370, Mar. 2019.

[48] 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, and L. Yin, "Multimodal spontaneous emotion corpus for human behavior analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3438–3446.



MARC-ANDRÉ FIEDLER received the B.Eng. degree from the Department of Electrical Engineering and Information Technology, Konstanz University of Applied Sciences, Germany, in 2017, and the M.Sc. degree from Otto von Guericke University Magdeburg, Germany, in 2019, where he is currently pursuing the Ph.D. degree with the Neuro-Information Technology Group. He is a Researcher with the Neuro-Information Technology Group, Otto von Guericke University Magdeburg. His research interests include biomedical signal/image processing, pattern recognition, and computer vision.

MICHAŁ RAPCZYŃSKI received the B.Sc. and M.Sc. degrees from Otto von Guericke University Magdeburg, Germany, in 2011 and 2013, respectively, where he is currently pursuing the Ph.D. degree with the Neuro-Information Technology Group. Since 2013, he has been a Researcher with the Neuro-Information Technology Group, Otto von Guericke University Magdeburg. His research interests include computer vision, image processing, machine learning, and biomedical signal processing.

AYOUB AL-HAMADI received the Ph.D. degree in technical computer science, in 2001, and the Habilitation degree in artificial intelligence and the Venia Legendi degree in pattern recognition and image processing from Otto von Guericke University Magdeburg, Germany, in 2010. He is currently an Adjunct Professor and the Head of the Neuro-Information Technology Group, Otto von Guericke University Magdeburg. He is the author of more than 300 papers in peer-reviewed international journals, conferences, and books. His research interests include computer vision, pattern recognition, and image processing. See http://www.iikt.ovgu.de/al_hamadi.html for more details.