Robust Remote Heart Rate Determination for E-Rehabilitation
A Method that Overcomes Motion and Intensity Artefacts

Christian Wiede, Jingting Sun, Julia Richter and Gangolf Hirtz
Department of Electrical Engineering and Information Technology,
Chemnitz University of Technology, Reichenhainer Str. 70, 09126 Chemnitz, Germany

Keywords: Remote Heart Rate Determination, rPPG, Vital Parameters, E-Rehabilitation.

Abstract: Due to an increasing demand for post-surgical rehabilitations, the need for e-rehabilitation is continuously rising. At this point, a continuous monitoring of vital parameters, such as the heart rate, could improve the efficiency assessment of training exercises by measuring a patient's physical condition. This study proposes a robust method to remotely determine a person's heart rate with an RGB camera. In this approach, we used an individual and situation depending skin colour determination in combination with an accurate tracking. Furthermore, our method was evaluated by means of twelve different scenarios with 117 videos. Altogether, the results show that this method performed accurately and robustly for e-rehabilitation applications.

1 INTRODUCTION

In recent years, the number of rehabilitation as a part of post-surgical care is continuously rising. Especially for surgeries of the musculoskeletal system, rehabilitation is a key factor for recovering. In order to prevent too light training or over-training a continuous monitoring of the patient is necessary. One possibility to evaluate a person's physical condition is to measure his or her vital parameters, such as the heart rate, the respiration rate or the oxygen saturation. In this work, we focus on the remote determination of the heart rate by means of an RGB camera.

This contact-less working principle has the advantage that the patients are not required to wear additional devices during the training, which is inconvenient for the patients and, in addition to that, increases the effort for the rehabilitation centres. Furthermore, more significant information about the patient's rehabilitation performance can be obtained. For example, a sudden change of a patient's physical condition can be detected by monitoring the heart rate. In that case, the training can be stopped and medical personnel can be informed. Afterwards the training intensity can be adapted.

However, e-rehabilitation is not the only application field of remote heart rate determination. In the field of ambient assisted living (AAL) such a remote heart rate determination could contribute to a long-term observation of the health status and assure a fast response time in cases of emergencies. Furthermore, such as system can also be applied for monitoring a driver's well-being in the context of autonomous driving and take control in emergency cases, e. g. a heart attack.

For remote heart rate determination there exist two general principles, which are intensity based methods, such as proposed by Poh et al. (Poh et al., 2010), and motion-based methods proposed by Balakrishnan (Balakrishnan et al., 2013). There are enhanced approaches as well, which combine the advantages of both principles (Wiede et al., 2016b). However, these methods encounter problems with motion and intensity artefacts, which poses challenges with regard to the application in e-rehabilitation: When a person moves during an exercise, the determined heart rate will be less accurate due to motion artefacts. Similarly, intensity artefacts, such as reflections and shadows, reduce the accuracy as well. In order to overcome these issues, we propose a robust, remote heart rate determination algorithm with an accurate pixel tracking and a situation and person dependent skin colour model. A database with reference data was recorded for the evaluation of this method.

This work is structured as follows: In Sect. 2, the related work in the field of remote heart rate determination is outlined and the research gap is highlighted. Based on this, our new method, which overcomes intensity and motion artefacts, is presented in Sect. 3. This is followed by the experimental results with a

Wiede, C., Sun, J., Richter, J. and Hirtz, G.
Robust Remote Heart Rate Determination for E-Rehabilitation - A Method that Overcomes Motion and Intensity Artefacts.
DOI: 10.5220/0006537504910500
In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 4: VISAPP*, pages 491-500
ISBN: 978-989-758-290-5

variety of evaluated scenarios in Sect. 4, which is accompanied by a discussion. Finally, we summarise our findings and outline future work.

2 RELATED WORK

The development of e-rehabilitation systems is continuously rising because of a higher demand and a lack of personnel resources. With the release of the Microsoft Kinect, cost-effective depth sensors became affordable and e-rehabilitation applications that employ the Kinect made its breakthrough. In the last years, several Kinect-based e-rehabilitation systems were developed, such as proposed by Su et al. (Su et al., 2014) or Gal et al. (Gal et al., 2015). However, to date there is no study that evaluates a patient's performance during exercises based on remotely determined vital parameters.

There are four main vital parameters, i. e. heart rate, respiration rate, oxygen saturation and blood pressure. In this study, we focus on remote heart rate determination by means of optical sensors using principles of photoplethysmography (PPG). In clinical environments, the heart rate is normally obtained by electrocardiography (ECG) or pulse oximeters. The basics of PPG were first described by Hertzman and Spealman (Hertzman and Spealman, 1937). They measured the volumetric changes of the blood flow with an optical sensor. The light that transmits through thin body parts, such as fingers or earlobes, is received by an optical sensor (Allen, 2007). This method is called transmissive PPG. Next to transmissive PPG, there exists the reflective PPG as well, which measures the light reflected from a tissue. Due to the reflection, the signal-to-noise ratio (SNR) for this method is decreased by a factor of ten compared to the transmissive PPG. Still, for both of these methods, sensors have to be attached to the body. In order to overcome this issue, Humphreys et al. developed a first concept for remote photoplethysmography (rPPG) (Humphreys et al., 2005). This was followed by first experiments in the infrared spectrum (Garbey et al., 2007) and the visible light spectrum (Verkruysse et al., 2008).

In 2007, Verkruysse et al. recorded probands a with small distance to an RGB camera. These probands were instructed not to move during the recordings in order to avoid motion artefacts. They detected a region of interest (ROI) within a face, perfomed a spatial averaging of the colour channels and determined the heart rate with the Fast Fourier Transform (FFT). This method was followed by the first automated approach by Poh et al. (Poh et al., 2010; Poh et al., 2011). They used an automated face detection and an independent component analysis (ICA). In order to increase the speed, Lewandoska et al. (Lewandowska et al., 2011) suggested to use a principal component analysis (PCA) instead of an ICA. Further works proposed to improve these methods by using temporal filters (van Gastel et al., 2014), autoregressive models (Tarassenko et al., 2014) or an adaptive filtering (Wiede et al., 2016a). All these approaches belong to the group of methods called intensity-based methods.

A different group of approaches are the so-called motion-based methods, which were first proposed by Balakrishnan et al. (Balakrishnan et al., 2013). They made use of small head motions caused by the heart bump triggered blood flow. By using several distinctive feature points in the person's face, small head motions can be tracked over time with a Kanade-Lucas-Tomasi (KLT) point tracker. After that, a PCA determined the principal components of the trajectories of the points. At last, the heart rate was obtained by using a peak detection.

As outlined by Wiede et al. (Wiede et al., 2016b), intensity- and motion-based methods have different advantages and disadvantages. Intensity-based methods are less sensitive to motion artefacts, whereas the motion-based methods suffer from fast motions. This is because the motion artefacts and the heart bump induced motion signal share the same frequency bands. In contrast to that, motion-based methods are less prone to illumination artefacts, such as reflections and shadows. The ratio-based method exploits these facts by using an intensity-based method when less intensity artefacts occur and a motion-based method when less motion artefacts are present (Wiede et al., 2016b). Consequently, the ratio-based method can not completely eliminate such artefacts, because it only chooses the method with the smallest amount of artefacts. Thus, the main problems originate from the underlying sources of artefacts. If these sources can be reduced or eliminated, the accuracy will increase significantly. For that, we propose an intensity-based method, which can overcome the motion artefacts by an accurate tracking and which significantly reduces intensity artefacts with a skin colour model.

3 METHODS

3.1 Overview

The major steps for the proposed robust remote heart rate determination are shown in Figure 1. After acquiring an RGB image, white balancing was app-

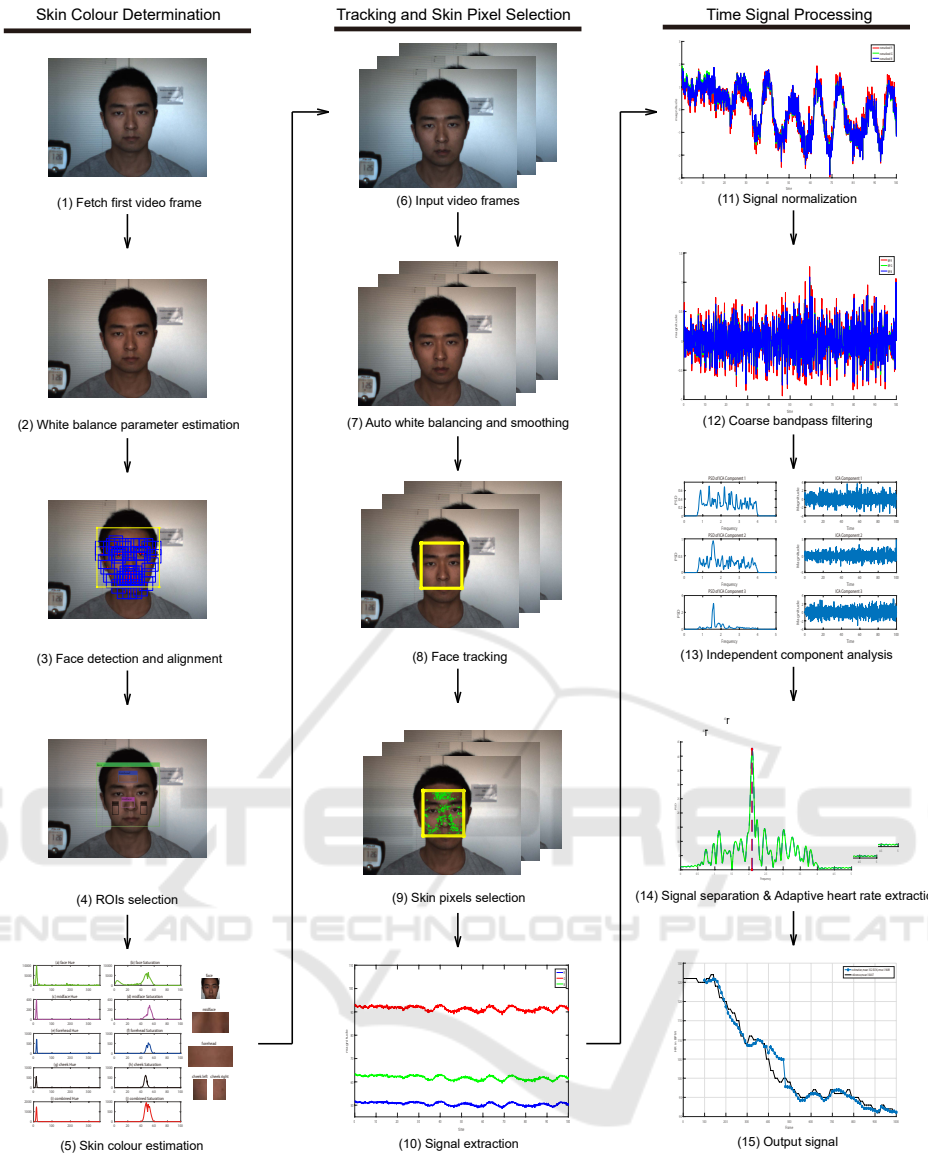

Figure 1: Overview of the proposed remote heart rate determination algorithm.

lied to obtain real world colours. In the first frame of the video, a face detection and alignment was carried out. Based on this, different ROI within the face were sampled and used to determine a proband's individual skin colour model. In the subsequent frames, we applied an auto white balancing, a face tracking, a skin pixel selection and a time signal extraction. This time signal was normalised and bandpass filtered. An ICA determined its independent components and the heart rate was obtained by means of a frequency analysis. An adaptive filtering assured a stable heart rate over time.

3.2 Skin Colour Determination

Due to the fact that different persons have different skin colours and the lighting conditions depend on the location, an individual skin colour model is necessary. For that, the first frame in the video was analysed and the parameters for the skin colour model were determined.

In a first step, a white balancing was applied to adjust the colours of the images by scaling and shifting the intensities in such a way that real white surfaces are finally represented by equally distributed RGB values. With that preprocessing, a bluish white or a yellowish white, for example, can be corrected. We im-

plemented a fast auto white balancing algorithm proposed by Garud et al. (Garud et al., 2014), which is based on the source illuminant values $[\iota_r,\iota_g,\iota_b]$. The correlated colour temperature (CCT) is given and the gain factors κ and the offset values τ can be determined. The gain factors are defined as follows:

$$\kappa_r = \frac{\iota_g}{\iota_r}, \tag{1a}$$

$$\kappa_g = 1, \tag{1b}$$

$$\kappa_b = \frac{\iota_g}{\iota_b}, \tag{1c}$$

where κ_r, κ_g and κ_b are the gain factors for the red, the green and the blue colour channel respectively.

The offset values are calculated as:

$$\tau_r = max(1, \frac{CCT - CCT_{ref}}{100}) \cdot (\kappa_r - 1), \tag{2a}$$

$$\tau_g = 0, \tag{2b}$$

$$\tau_b = max(1, \frac{CCT_{ref} - CCT}{100}) \cdot (\kappa_b - 1), \tag{2c}$$

where CCT_{ref} denotes the CCT of the canonical illuminant.

With these factors, the white balanced colour channels R_{wb}, G_{wb} and B_{wb} can be determined by the following equation:

$$\begin{bmatrix} R_{wb} \\ G_{wb} \\ B_{wb} \end{bmatrix} = \begin{bmatrix} \kappa_r & 0 & 0 \\ 0 & \kappa_g & 0 \\ 0 & 0 & \kappa_b \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} \tau_r \\ \tau_g \\ \tau_b \end{bmatrix} \tag{3}$$

R, G and B are the original intensity values.

In the next step, the person's face was detected in the image. A common approach for this is the Viola and Jones face detector (Viola and Jones, 2004). However, this approach is not accurate enough for this application so that the face detector by Zhu and Ramanan (Zhu and Ramanan, 2012) was used instead. This detector provides 68 facial landmarks in real-world cluttered images. The provided bounding box is very robustly located around the face. However, we had to adjust the bounding box for our requirements to include the forehead region and to exclude the neck region. For that purpose, the bounding box was enlarged at the left and the right boundary by 10 %, at the upper boundary by 30 % and reduced at the lower boundary by 10 %.

For the skin colour model, there are regions in the face that are certainly skin pixels and not covered by hair or other interfering objects. Under the condition that the face was frontally captured, the regions of the forehead, the two cheeks and the nose were selected by their relative positions with regard to the total face bounding box. One selection of these ROIs is shown in Figure 2. These four ROIs were taken for the following skin colour estimation.

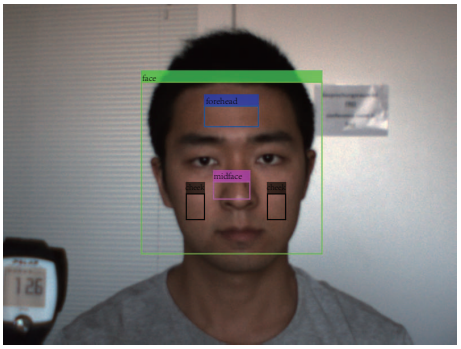

Figure 2: ROIs of the face regions selected for the skin colour model, i. e. forehead, nose and the two cheeks.

The RGB colour space is not suited for determining a skin colour model, because the distribution of the skin pixels does not follow any linear or concentrated coherency. Therefore, a conversion to a different colour space that separates brightness and chrominance is necessary. The HSV colour space containing the hue H, the saturation S and the value V is convenient for this task. In accordance with the conversion rules from Smith (Smith, 1978), the brightness value V can be calculated by:

$$V = \max(R,G,B). \tag{4}$$

The auxiliary variable C, which stands for the chroma value, can be determined as follows:

$$C = V - \min(R,G,B). \tag{5}$$

With these values the saturation S can be calculated by:

$$S = \begin{cases} 0, & \text{if } V = 0, \\ \frac{C}{V}, & \text{otherwise}. \end{cases} \tag{6}$$

The hue H is given by:

$$H = \begin{cases} \text{undefined}, & \text{if } C = 0, \\ 60^\circ \cdot (\frac{G-B}{C}), & \text{if } V = R, \\ 60^\circ \cdot (\frac{B-R}{C} + 2), & \text{if } V = G, \\ 60^\circ \cdot (\frac{R-G}{C} + 4), & \text{if } V = B. \end{cases} \tag{7}$$

The HSV colour space represents a cylindrical colour space. For the further consideration of the skin pixels, the hue-saturation-plane is relevant. In order to define a region in this plane, which represents the skin colour of a certain person, thresholds for the hue and the saturation have to be determined. At this point, an adaption has to be made for the hue: Red is the dominant colour of the face. Since the hue values for the red pixels are in a range around zero, the hue H was shifted by 120 degrees, as shown in the following equation:

$$H^* = \begin{cases} H + 240^\circ, & \text{if } H \leq 120^\circ \\ H - 120^\circ, & \text{otherwise}. \end{cases} \tag{8}$$

As shown in Figure 3, the values for hue and saturation of the previously defined ROIs are located in the same area.

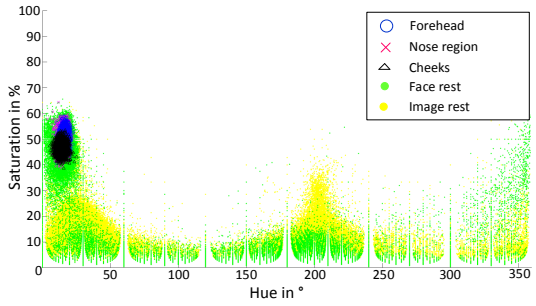

Figure 3: Illustration of the ROI pixels in the shifted hue-saturation-plane in comparison to other pixels in the image and the face.

According to the overall human skin model, with red as the dominant color, the threshold values for the shifted hue H^* and the saturation S are as follows:

$$\begin{aligned} H^* &\in [186°, 294°] \\ S &\in [20\,\%, 100\,\%] \end{aligned} \tag{9}$$

In our work, however, we applied different thresholds depending on the person's specific skin colour and the lighting conditions. In Figure 4, the selected skin pixels that were chosen according to the adapted thresholds defined in Equation 10 are shown.

$$\begin{aligned} H^* &\in [252°, 259°] \\ S &\in [45\,\%, 56\,\%] \end{aligned} \tag{10}$$

It can be seen that regions of the eyes, hairs, lips, glasses, nostrils, shadows and reflections do not belong the skin colour model.

Figure 4: Skin pixels that were selected with specific hue and saturation thresholds within the face bounding box. Unselected pixels are masked with black.

3.3 Tracking and Skin Pixel Selection

Once the skin colour model was determined based on the first frame, a continuous tracking of the face bounding box and a skin pixel selection were conducted during the following frame sequence. The tracking is necessary to be invariant against different motion artefacts. Our tracking method is based on the optical flow principle. The optical flow method estimates the motion between two consecutive frames at the time t and $t + \Delta t$. This results in the general optical flow equation:

$$I_x V_x + I_y V_y = -I_t \,, \tag{11}$$

where I_x, I_y and I_t are the partial derivatives of the image at the position (x, y) on time t, V_x,V_y are the x and y components of the velocity or the optical flow of $I(x, y, t)$. This equation contains two unknowns and cannot be solved directly. A solution for this is the KLT tracking algorithm (Tomasi and Kanade, 1991). It followed the assumption that the motion is constant in a local neighbourhood of an image patch. For n different patterns in the image, we obtain n equations:

$$\begin{aligned} I_x(p_1)V_x + I_y(p_1)V_y &= -I_t(p_1) \,, \\ I_x(p_2)V_x + I_y(p_2)V_y &= -I_t(p_2) \,, \\ &\vdots \\ I_x(p_n)V_x + I_y(p_n)V_y &= -I_t(p_n) \,, \end{aligned} \tag{12}$$

where $p_1, p_2, \ldots, p_n$ are the pixels inside the image patch. These equations can be written in matrix form $Av = b$, where:

$$A = \begin{bmatrix} I_x(p_1) & I_y(p_1) \\ I_x(p_2) & I_y(p_2) \\ \vdots & \vdots \\ I_x(p_n) & I_y(p_n) \end{bmatrix}, \tag{13a}$$

$$v = \begin{bmatrix} V_x \\ V_y \end{bmatrix}, \tag{13b}$$

$$b = \begin{bmatrix} -I_t(p_1) \\ -I_t(p_2) \\ \vdots \\ -I_t(p_n) \end{bmatrix}. \tag{13c}$$

That equation system can be solved by the least squares principle:

$$A^T A v = A^T b. \tag{14}$$

As a feature, the minimum Eigenvalue features proposed by Shi and Tomasi (Shi and Tomasi, 1993) were selected, because they came up with a large robustness. However, because of projective distortions in the image region of the face, feature points can vanish over time. A solution is to re-detect a subject's face. For that, we used the normalised pixel difference (NPD) face detector proposed by Liao et al. (Liao et al., 2016). The face detector learns NPD features by a classifier with a quadratic tree structure

with a depth of eight. Should the NPD detector fail to detect a subject's face, the KLT tracker is able to track the pixel for a while.

The 2-D geometric transform from one frame to the next frame can be estimated by using the tracked pixels and can be applied in the same manner to the face bounding box to follow the head motion. By combining the NPD face detector and 2-D geometric transform estimation, the subject's face region can be accurately tracked even in case of complex head motions.

Assuming that the lighting conditions do not change completely from the first frame on, the skin colour model can be applied for the total frame sequence. In the tracked face bounding box, all pixels that match the thresholds of the skin colour model were selected. In order to improve the reliability of the skin pixel selection, a distance threshold D was defined. For every skin pixel, the distance to the closest non-skin pixel was calculated. If this distance was smaller than D, this skin pixel was rejected. This procedure is equivalent to an erosion.

For the time signal extraction, all remaining skin pixels were taken into consideration. They were averaged for each frame for all three colour channels R, G and B. Please note that we operate in the discrete time domain and use n instead of the continuous variable t.

$$\bar{R}(n) = \frac{1}{L}\sum_{l=1}^{L} R_l(n) \tag{15a}$$

$$\bar{G}(n) = \frac{1}{L}\sum_{l=1}^{L} G_l(n) \tag{15b}$$

$$\bar{B}(n) = \frac{1}{L}\sum_{l=1}^{L} B_l(n) \tag{15c}$$

R_l, G_l and B_l denote the l^{th} selected skin pixel in the frame and L is the number of all selected skin pixels in this frame. $\bar{R}(n)$, $\bar{G}(n)$ and $\bar{B}(n)$ represent the mean value of the facial skin colour for a certain frame n. As a result, we obtained a time varying signal for the skin colour.

3.4 Time Signal Processing

In order to remove remaining noise sources, the colour time varying signal has to be further processed to increase the SNR and to obtain a robust heart rate signal.

The first step of the time signal processing was to normalise the signal to attain a zero mean and a standard deviation of one:

$$\hat{R}(n) = \frac{1}{\sigma_R}(\bar{R}(n) - \mu_R)\,, \tag{16a}$$

$$\hat{G}(n) = \frac{1}{\sigma_G}(\bar{G}(n) - \mu_G)\,, \tag{16b}$$

$$\hat{B}(n) = \frac{1}{\sigma_B}(\bar{B}(n) - \mu_B)\,, \tag{16c}$$

where $\hat{R}$, $\hat{G}$ and $\hat{B}$ refer to the normalised colour channels. μ_C is the mean value and σ_C is the standard deviation of the corresponding colour channel $C \in \{R,G,B\}$:

$$\mu_C = \frac{1}{N}\sum_{n=1}^{N} \bar{C}(n)\,, \tag{17}$$

$$\sigma_C = \sqrt{\frac{1}{N}\sum_{n=1}^{N} (\bar{C}(n) - \mu_C)^2}\,, \tag{18}$$

where $\bar{C}(n)$ represents the original colour channels and N is the sequence length of the colour signal for a single channel.

This was followed by a bandpass filter BP, which excludes implausible frequencies, see Equation 19. The frequencies lower than 0.7 Hz and higher than 4 Hz were cut off. For this implementation, an FIR filter with an order of 128 was chosen to ensure a constant group delay. The filtered colur channels are then denoted as R_{BP}, G_{BP} and B_{BP}.

$$R_{\text{BP}}(n) = \text{BP}(n) * \hat{R}(n) \tag{19a}$$

$$G_{\text{BP}}(n) = \text{BP}(n) * \hat{G}(n) \tag{19b}$$

$$B_{\text{BP}}(n) = \text{BP}(n) * \hat{B}(n) \tag{19c}$$

Even now, the three filtered colour channels can still contain noise sources. In order to separate the wanted pulse signal from the noise sources, a decomposition of the colour channels by an ICA was applied. The goal is to determine three new independent components IC_1, IC_2 and IC_3:

$$\begin{bmatrix} R_{\text{BP}}(n) \\ G_{\text{BP}}(n) \\ B_{\text{BP}}(n) \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} \cdot \begin{bmatrix} \text{IC}_1(n) \\ \text{IC}_2(n) \\ \text{IC}_3(n) \end{bmatrix}. \tag{20}$$

In our implementation, we used the FastICA approach of Hyvrinen (Hyvärinen, 1999).

In the next step, the independent component that contains the wanted signal should be selected for the further processing. We assume that the independent component with the highest periodicity p is most likely the one that contains the pulse signal. The periodicity p of a signal is defined as as the ratio between

the accumulated coefficients in a range of 0.05 Hz around the dominant frequency f_d and the accumulated coefficients of the total power spectrum, see Equation 21. f_s is the sampling frequency.

$$p = \frac{\sum_{f_d-0.025}^{f_d+0.025} \hat{S}_{xx}^{av}(k)}{\sum_{0}^{f_s} \hat{S}_{xx}^{av}(k)} \tag{21}$$

In order to calculate p, the spectrum of each independent component has to be obtained. One possibility to do this is by the Welch's estimate of the power spectrum density (PSD) $\hat{S}_{xx}^{av}(k)$:

$$\hat{S}_{xx}^{av}(k) = \frac{1}{N} \sum_{n=1}^{N} \hat{P}_n(k). \tag{22}$$

Thereby, $\hat{P}_n(k)$ denotes the periodogram and k is the discrete iterator in the frequency domain instead of the continuous variable f. The periodogram uses a hamming window for each segment.

After having selecting the best independent component IC_i, this component was split into segments of 10 s with an overlap of 90 % of the segments. This small segment size guarantees a flexibility when the heart rate changes rapidly, for example during a training exercise. The dominant frequency f_{FFT} for each segment k was determined by calculating the FFT for this segment and by determining the maximum in the spectrum:

$$f_{\mathrm{FFT}}(k) = \max(|\mathrm{FFT}(\mathrm{IC}_i)|). \tag{23}$$

In presence of strong motion artefacts, other high peaks can appear in the spectrum. They can be misinterpreted as the real heart rate signal. In order to avoid this, an adaptive filtering is introduced. We assume that the heart rate does not change by more than 15 BPM (0.25 Hz) between two adjacent frames. The mean value of the estimated heart rates in the two previous segments was defined as the guide frequency f_{gui}. As shown in Figure 5, only the part of the spectrum for which applies $f_{\mathrm{gui}} \pm 0.25\,\mathrm{Hz}$ was taken into consideration for the final heart rate HR. To obtain the final heart rate in beats per minute (BPM), the frequency f_{FFT} has to be multiplied by 60:

$$\mathrm{HR}(k) = f_{\mathrm{FFT}}(k) \cdot 60\ [\mathrm{BPM}]. \tag{24}$$

4 RESULTS AND DISCUSSION

4.1 Setting

As a basis for our evaluation, we created a database of eleven probands with 117 different videos in total. The probands are of different gender, age and skin colour to guarantee a high variability. In total, twelve different scenarios were considered: Starting from a control scenario without any noise sources, we recorded scenarios with illumination artefacts caused by a lighting source placed above the face or placed at one side of the face, which results in different kinds of shadows. Moreover, the probands had to perform different motions to obtain scenarios with translations and rotations of the head (pitch, yaw and roll), scaling as well as non-rigid movements to represent motion artefacts. Furthermore, we combined motion artefacts and intensity artefacts in one scenario. In order to generate videos with a varying heart rate, which is natural in the context of rehabilitation exercises, videos after sport and during cycling exercises were made.

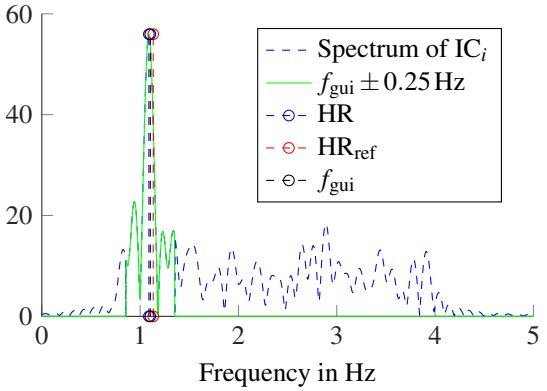

Figure 5: For the adaptive filtering, only frequency peaks that are in the range of $f_{\mathrm{gui}} \pm 0.25\,\mathrm{Hz}$ were considered. f_{gui} represents the guide frequency, HR the selected heart rate and $\mathrm{HR}_{\mathrm{ref}}$ the corresponding reference heart rate.

For all recordings, an industrial camera, i. e. an Allied Manta G201c, was chosen. The automatic exposure time control and the automatic white balancing were disabled in order not to influence the measurements. The video sequences had a length of 1,000 frames and were recorded with a fixed frame rate of 10 FPS.

A Polar FT1 heart rate monitor was used as a reference system. This system measures the heart rate by means of a chest strap and displays it. This display was visible in all recorded videos, so that a reference value for the heart rate could be obtained for every frame.

4.2 Accuracy

The evaluation criterion that we have chosen for the accuracy analysis is the root-mean-square error (RMSE) for a sequence m, see Equation 25.

$$\mathrm{RMSE_m} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} |\mathrm{HR}(n) - \mathrm{HR_{ref}}(n)|^2} \tag{25}$$

In this equation, HR is the estimated heart rate and HR_{ref} the reference heart rate. For the single scenarios, we calculated the mean value $\overline{RMSE}$ of all sequences M.

$$\overline{RMSE} = \frac{1}{M}\sum_{m=1}^{M} RMSE_m \tag{26}$$

Every video consist of 91 segments, which results in 10,647 evaluated segments in total for 117 videos. This outlines the extent of the data base and its statistical relevance.

In Table 1, the results for the single scenarios are presented. As expected, the control scenario without any challenges shows the best $\overline{RMSE}$ with 1.19 BPM. Since the error of the reference system can be quantified with ± 1 BPM, this result proves to be of high quality.

The scenarios with illumination artefacts show shadows and reflection. This causes the $\overline{RMSE}$ to increase up to 1.38 BPM for the side illumination and 1.48 BPM for the upper illumination, which is still accurate. Solely occuring illumination artefacts do not show a large impact on the proposed algorithm.

While the determined heart rate for translation can be rated as accurate as well, the error is increasing for the scaling and rotation scenarios. This can be explained by a more challenging tracking and therefore larger changes in the size of the bounding box. The non-rigid movements show the largest $\overline{RMSE}$ for the motion scenarios with 2.46 BPM. This is logical: due to the change of the shape of the face, which is a result of speaking and facial expressions, the size and the location of the bounding box is influenced.

When motion and illumination artefacts are combined in one scenario, the $\overline{RMSE}$ increases up to 2.92 BPM. The scenarios after the sport and during the cycling showed an increased $\overline{RMSE}$ of 1.53 BPM and 2.11 BPM. Especially the heart rate determination during the cycling is very challenging because of its periodically motions. However, all scenarios showed an $\overline{RMSE}$ below 3 BPM, which seems to be accurate for the use case e-rehabilitation.

4.3 Robustness

For the evaluation, not only mean values of a complete sequence, such as the $\overline{RMSE}$, are relevant. It is also of high importance that the differences between the estimated heart rate and the reference heart rate are not too high for single segments. This criterion is referred to as robustness. In Figure 6, for example, the reference heart rate and the estimated heart rate are shown in one plot for a video after the sport. It can be seen that the estimated heart rate is very close

Table 1: $\overline{RMSE}$ for all scenarios in BPM.

Evaluated Scenario	$\overline{RMSE}$
Control	1.19
Upper illumination	1.49
Side illumination	1.38
Translation	1.36
Yaw	1.70
Pitch	1.93
Roll	1.86
Scaling	1.81
Non-rigid motion	2.46
Motion and illumination	2.93
After sport	1.53
During cycling	2.11

to the reference heart rate for the majority of the segments. For some segments, however, this difference is slightly higher.

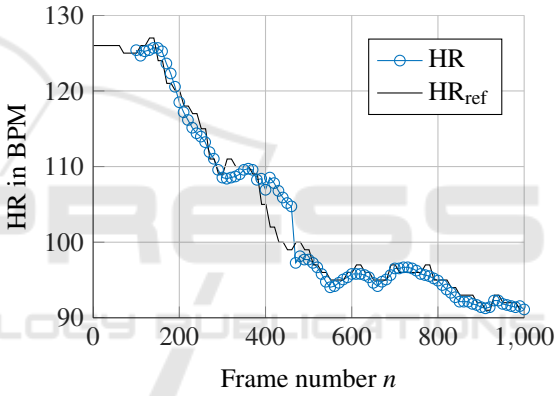

Figure 6: Comparison of the computed heart rate (blue dots) and the reference heart rate (black curve) after a sport exercise.

In order to perform a more detailed analysis of the single differences, the amount of segments Φ that have a difference d below a certain value is plotted over the difference using all scenarios, as shown in Figure 7. The difference d is calculated as follows:

$$d(\phi) = HR(\phi) - HR_{ref}(\phi). \tag{27}$$

phi denotes the segment number.

In Figure 7, it can be seen that 98.3 % of the segments in the control sequences have a difference below 4 BPM, for example. For the upper illumination 97 % and the side illumination 97.5 % of the segments have a difference smaller than 4 BPM. For the rigid motion 93.3 % and for the non-rigid motions 88.8 % of the segments show a maximum difference of 4 BPM. In the case where strong motions and intensity artefacts occur together, this rate drops

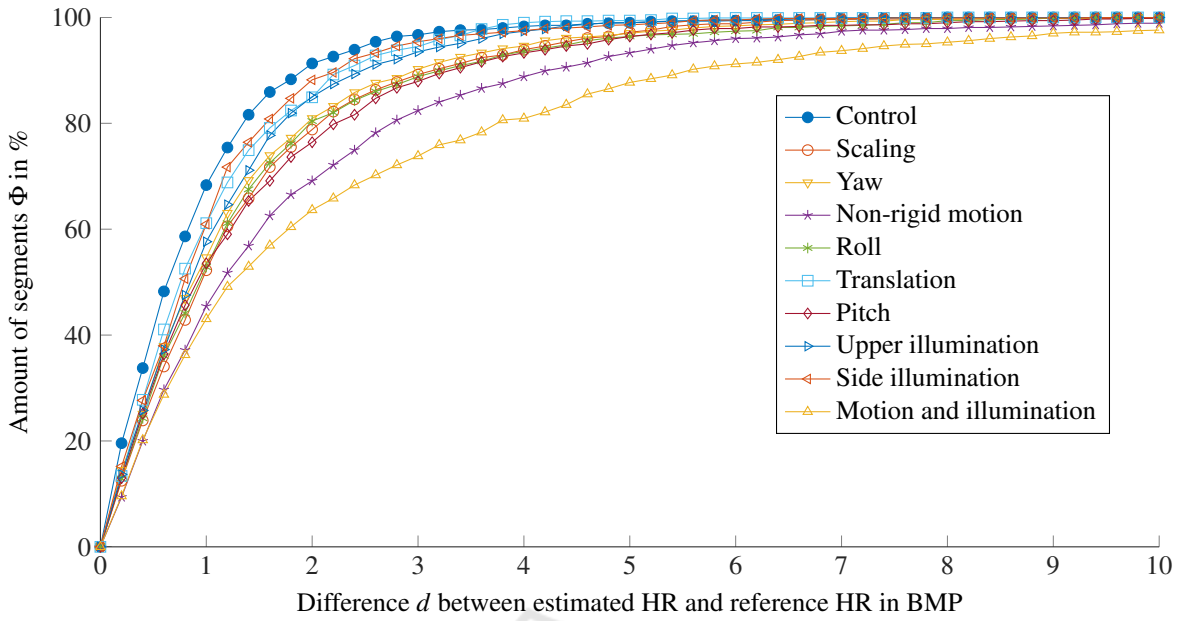

Figure 7: The overall results are visualised. The y-axis indicates how many percent of the measured data points show a better performance than certain difference in BPM.

to 80.9 %. If all scenarios are considered, we determined that 90 % of the segments have a difference that is smaller than 4 BPM. This robustness is regarded as sufficient for the field of e-rehabilitation. In this application, it is not of high importance whether the heart rate at a certain time is exactly 120 BPM or 122 BPM, for example. The detection of relative changes or the velocity of heart rate changes within or after an exercise is more important.

5 CONCLUSIONS

In this study, we presented a new method for remote heart rate determination, which is robust against intensity and motion artefacts. This method consist of an accurate tracking and an individual, situation depending skin colour determination. That is accompanied by a bandpass filtering, an ICA and a frequency determination.

For the evaluation, the accuracy was calculated by means of a reference system. With an $\overline{\text{RMSE}}$ below 3 BPM, this method provides a good basis for an application in e-rehabilitation. Even in the scenarios during sport activities, this method demonstrated robustness.

In future, we plan to evaluate this method in a field study in rehabilitation facilities. Furthermore, we intend to extent the algorithms for the use of thermal cameras. Finally, it is planned to evaluate this method in other application fields, such as AAL or driver's monitoring.

ACKNOWLEDGEMENTS

This project is funded by the European Social Fund (ESF). We thank all volunteers who took part in the recordings.

REFERENCES

Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3):R1–R39.

Balakrishnan, G., Durand, F., and Guttag, J. (2013). Detecting Pulse from Head Motions in Video. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3430–3437.

Gal, N., Andrei, D., Neme, D. I., Ndan, E., and Stoicu-Tivadar, V. (2015). A Kinect based intelligent e-rehabilitation system in physical therapy. *Digital Healthcare Empowering Europeans*, pages 489–493.

Garbey, M., Sun, N., Merla, A., and Pavlidis, I. (2007). Contact-Free Measurement of Cardiac Pulse Based on the Analysis of Thermal Imagery. *Biomedical Engineering, IEEE Transactions on*, 54(8):1418–1426.

Garud, H., Ray, A. K., Mahadevappa, M., Chatterjee, J., and Mandal, S. (2014). A fast auto white balance scheme for digital pathology. In *2014 IEEE-EMBS Internati-*

onal Conference on Biomedical and Health Informatics, BHI 2014, pages 153–156.

Hertzman, A. B. and Spealman, C. R. (1937). Observations on the finger volume pulse recorded photoelectrically. *American Journal of Physiology*, 119:334–335.

Humphreys, K., Markham, C., and Ward, T. (2005). A CMOS camera-based system for clinical photoplethysmographic applications. In *Proceedings of SPIE*, volume 5823, pages 88–95.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *Neural Networks, IEEE Transactions on*, 10(3):626–634.

Lewandowska, M., Ruminski, J., Kocejko, T., and Nowak, J. (2011). Measuring pulse rate with a webcam - a non-contact method for evaluating cardiac activity. In *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on*, pages 405–410.

Liao, S., Jain, A. K., and Li, S. Z. (2016). A Fast and Accurate Unconstrained Face Detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):211–223.

Poh, M.-Z., McDuff, D., and Picard, R. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, 18(10):10762–10774.

Poh, M.-Z., McDuff, D., and Picard, R. (2011). Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam. *Biomedical Engineering, IEEE Transactions on*, 58(1):7–11.

Shi, J. and Tomasi, C. (1993). Good Features to Track. Technical report, Cornell University, Ithaca, NY, USA.

Smith, A. R. (1978). Color gamut transform pairs. *ACM SIGGRAPH Computer Graphics*, 12(3):12–19.

Su, C.-J., Chiang, C.-Y., and Huang, J.-Y. (2014). Kinect-enabled home-based rehabilitation system using Dynamic Time Warping and fuzzy logic. *Applied Soft Computing*, 22:652–666.

Tarassenko, L., Villarroel, M., Guazzi, A., Jorge, J., Clifton, D. A., and Pugh, C. (2014). Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiological Measurement*, 35(5):807–831.

Tomasi, C. and Kanade, T. (1991). Detection and Tracking of Point Features. Technical report, Carnegie Mellon University.

van Gastel, M., Zinger, S., Kemps, H., and de With, P. (2014). e-health video system for performance analysis in heart revalidation cycling. In *Consumer Electronics Berlin (ICCE-Berlin), 2014 IEEE Fourth International Conference on*, pages 31–35.

Verkruysse, W., Svaasand, L. O., and Nelson, J. S. (2008). Remote plethysmographic imaging using ambient light. *Optics Express*, 16(26):21434–21445.

Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154.

Wiede, C., Richter, J., Apitzsch, A., KhairAldin, F., and Hirtz, G. (2016a). Remote Heart Rate Determination in RGB Data. In *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods*, pages 240–246, Rome.

Wiede, C., Richter, J., and Hirtz, G. (2016b). Signal fusion based on intensity and motion variations for remote heart rate determination. In *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*, pages 526–531.

Zhu, X. and Ramanan, D. (2012). Face detection, pose estimation, and landmark localization in the wild. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2879–2886.