Expert Systems With Applications 207 (2022) 117867

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A comparative study of common steps in video-based remote heart rate detection methods

Lakmini Malasinghe [a,b,1], Stamos Katsigiannis [c,*,1], Keshav Dahal [b,1], Naeem Ramzan [b,1]

[a] Sri Lanka Institute of Information Technology, New Kandy Road, Malabe, Sri Lanka
[b] University of the West of Scotland, High St., Paisley, PA1 2BE, United Kingdom
[c] Durham University, Stockton Road, Durham, DH1 3LE, United Kingdom

ARTICLE INFO

Keywords:
Remote heart rate detection
Video-based heart rate detection

ABSTRACT

Video-based remote heart rate detection is a promising technology that can offer convenient and low-cost heart rate monitoring within, but not limited to, the clinical environment, especially when attaching electrodes or pulse oximeters on a person is not possible or convenient. In this work, we examined common steps used in video-based remote heart rate detection algorithms, in order to evaluate their effect on the overall performance of the remote heart rate detection pipeline. Various parameters of the examined methods were evaluated on three public and one proprietary dataset in order to establish a video-based remote heart rate detection pipeline that provides the most balanced performance across various diverse datasets. The experimental evaluation demonstrated the effect and contribution of each step and parameter set on the estimation of the heart rate, resulting in an optimal configuration that achieved a best RMSE value of 9.51.

1. Introduction

Heart rate is a simple physiological cue that can reveal many things about a person's health. Heart rate is the rate at which the heart pumps blood inward and outward itself to supply blood to the whole body and support life. Heart rate detection has been performed for centuries using manual methods, such as manual pulse checking (Pickering, 2013). Then came the stethoscope, which made manual checking easier and is still widely used today. Due to the importance of the accuracy of heart rate measurement (Davidovic et al., 2013; Hori & Okamoto, 2012), human errors in manual pulse checking should be minimised, if not eliminated (Kobayashi, 2013). Therefore, it has been of interest for many decades to find better methods for heart rate detection. The invention of pulse oximeters was an important step towards minimising such human errors, with pulse oximeters being widely used today (Sinex, 1999). One limitation of pulse oximeters is that they must be attached to the human body.

Remote heart rate detection using imaging-based methods is a fast developing research area, as its non-invasive nature overcomes the inconveniences of with-contact or invasive procedures, and can become the solution for cases where patients become unsuitable candidates for with-contact approaches because of the difficulty in attaching electrodes or sensors to those who have, for example, sensitive skin due to age, burns, injuries, or delicate skin (newborns and neonates). The principle behind image-based remote heart rate monitoring is to use a camera to capture a physiological phenomenon that results from the beating of the heart and derive the heart rate. With every outward pump from the heart, the blood circulates through the veins and the colour of the surrounding skin changes when oxygen-rich blood is in the veins. When it is pumped back into the heart, this skin colour changes again. This recurrent activity causes minute instantaneous colour changes in skin which cannot be detected by the human eye. As such phenomena are not visible to the naked eye, technology is used to obtain such signals using a camera over a brief period and then have the acquired data (video/images) processed in a suitable manner to extract the heart rate (Sun & Thakor, 2016). Despite the amount of research in this field, only a limited number of studies (Kranjec et al., 2017; Tarassenko et al., 2014) have been conducted in clinical environments. Factors that help or hinder the process of remote heart rate detection must therefore be investigated. Some such factors have been identified and tested in previous works in this field, e.g., illumination rectification (Li et al., 2014), and different colour spaces (McDuff et al., 2014b). However, the literature still has a void for a broad study that tests many of these factors at once.

This work attempts to fill this void. Firstly, a thorough literature review of some notable works in the recent literature was conducted

* Corresponding author.
E-mail addresses: lakmini.m@sliit.lk (L. Malasinghe), stamos.katsigiannis@durham.ac.uk (S. Katsigiannis), Keshav.Dahal@uws.ac.uk (K. Dahal), Naeem.Ramzan@uws.ac.uk (N. Ramzan).
[1] All authors contributed equally to this work.

https://doi.org/10.1016/j.eswa.2022.117867
Received 10 June 2020; Received in revised form 7 April 2022; Accepted 11 June 2022
Available online 22 June 2022

in order to determine which methods are most commonly used as steps within video-based heart rate detection pipelines. Then, a system was developed to test these different steps in order to examine their contribution towards the final outcome. Three public datasets and a dataset created by the authors were used to evaluate the performance. The system was designed in such a way that each step could be included or excluded from the main pipeline (enabled or disabled) and various parameter values could be tested when applicable. Finally, through an exhaustive performance evaluation, this work provides an overview of how the examined methods affect the final results and proposes the best combination for a video-based remote heart rate detection pipeline. To the best of our knowledge, this is the only work that evaluates such a system with four diverse datasets.

The rest of this work is organised as follows: Section 2 provides an overview of previously proposed video-based remote heart rate detection approaches. Section 3 describes the experimental procedure followed, whereas results are presented and discussed in Section 4. Finally, conclusions are drawn in Section 5.

2. Background

Remote patient monitoring is one of the most widely-studied fields in biomedical engineering. This work is focused on one area within this broad topic, i.e. heart rate detection. For a more general overview on the field of remote patient monitoring, we would like to direct the reader to our previous work (Malasinghe et al., 2017) and the article by Sathyanarayana et al. (2015). As explained before, since heart rate can hold the key to identifying many underlying cardiovascular system-related illnesses, it is in the spotlight within the field of remote patient monitoring. Traditional methods such as manual pulse checking and counting, which were carried out for centuries are still in widespread use. However, as these require a trained doctor or a medical professional, they have very little use in continuous monitoring in non-hospital environments.

To eliminate the chances of human errors, hospitals started using devices employing contact photoplethysmography (PPG) techniques (Moraes et al., 2018). With every cycle of blood circulation (heart beat), blood circulates through the body. When blood is pumped out of the heart, blood volume in veins increases and when blood travels back to the heart, the volume decreases. If light is directed upon the vein and its absorption or reflection is captured by suitable equipment, these blood volume changes can be observed, and heart's systole and diastole can be "seen". Observing over a period will give sufficient data to derive heart rate. This is the basic principle of PPG. PPG technologies can be divided into categories such as contact PPG and non-contact PPG (Sun & Thakor, 2016). The non-contact PPG is the basis of imaging PPG (iPPG) where capturing data is done by capturing images, an approach used in most contactless heart rate extraction studies. Allen (2007) compiled a widely cited review that compared many research works on traditional PPG up to 2000, while the reviews by Hassan et al. (2017) and McDuff et al. (2015) are two of the most cited recent reviews on remote PPG. Moço et al. (2018) provided a deeper mathematical explanation of the origin of PPG signals in visible and infrared wavelengths, explaining the properties of skin and other factors that impact on the overall remote PPG measurement. Taking into consideration these physiological properties of skin has been shown as useful when implementing a system, as opposed to heuristic methods used in other studies to remove artefacts.

The rest of this section describes many important works on some of the main tasks required for image-based remote heart rate detection pipelines.

2.1. Colour space and colour channels

The first step in most image-based remote heart rate detection algorithms is the acquisition of raw video data and, depending on the method, the conversion of the raw video to a suitable colour space for further analysis. When using typical video, one of the most common approaches is the use of the three channels of the RGB colour space, as proposed by Poh et al. (2010, 2011) and Verkruysse et al. (2008), while Cheng et al. (2017) suggested that the use of only the Green channel out of the three is sufficient. The green channel is usually considered to be better than the red and blue channels, although some studies reveal surprising findings, such as the red channel containing more information related to heart rate (Alzahrani & Whitehead, 2015; Bosi et al., 2016; Pal et al., 2013). The use of the LAB colour space (Fernandes et al., 2017), of the cyan, green, and orange (CGO) bands (McDuff et al., 2014a), and of a Near Infrared (NIR) channel (Kado et al., 2018) have also been proposed.

2.2. Regions of Interest (ROIs)

The captured data must be carefully selected so that the required heart rate signal is contained in them, which means that one or more suitable Regions of Interest (ROI) within each video frame must be chosen. In a notable early experiment on contactless heart rate monitoring (Costa, 1995), the selected ROIs were in the forearm near the wrist and elbow. Cennini et al. (2010) and Yang et al. (2015) proposed the use of the palm as the ROI, while others proposed the combination of the palm and face regions (Fan & Li, 2018; Wei et al., 2017). Some studies suggest areas like neck (Bosi et al., 2017), lips (Procházka et al., 2016), or pupils (Parnandi & Gutierrez-Osuna, 2013). Verkruysse et al. (2008) tested many different ROIs, such as a rectangular area on the forehead, a minute area on the forehead, a region covering portion of hair and a section from the background. Selecting suitable ROIs is a challenging task that greatly affects the performance of the proposed methods. The most commonly used ROIs are located within the region of the face. As a result, ROI computation consists of first detecting the face of the individual and then selecting the appropriate ROIs within the face area. A good ROI should include an area on the skin that does not have non-skin properties like cosmetics and hair, or areas that are comparatively less illuminated properly like eyelids. Common selections are either or both cheeks (Malasinghe et al., 2018), the forehead (Haque et al., 2016; Wiede et al., 2016), full face, or a portion of the face (Haque et al., 2016; Li et al., 2014; Poh et al., 2010, 2011). Some studies use more than one potential ROI to improve the results (Datcu et al., 2013; Kado et al., 2018), while others detect the face and then divide the face into many small segments (Gupta et al., 2017). In Procházka et al. (2016), four regions are considered, which include the whole lips area and the lips divided into three segments, in order to compare results from each region. The shape of ROI has been rectangular in most studies, although different shapes can be seen in Bobbia et al. (2017), Malasinghe et al. (2018) and Procházka et al. (2016). Most modern approaches include choosing multiple areas on the face and then choosing a unique ROI for each subject (Bobbia et al., 2017; Kado et al., 2018). If face detection is not performed, then the ROIs are calculated all over the image (Bobbia et al., 2016, 2017).

2.3. Artefact removal

After selecting suitable ROIs, the pixel intensity within these ROIs from all the video frames is used in order to create time series that depict the variation in pixel intensity across time. However, apart from heart rate-related information, these time series contain artefacts originating from various sources, such as illumination, shadows, movement, camera noise, external objects, etc. To mitigate their effects, researchers have applied artefact removal and denoising methods. Noise, light flicker, and spontaneous peaks in the signal can be generally removed

with a moving average filter (Smith, 2003). Normalisation and detrending the input signals are also very important steps that remove artefacts to a good extent (Tanabe et al., 2002).

One of the main sources of artefacts is illumination. Incident light on the subject can significantly alter the image recorded by a camera (Basri & Jacobs, 2003). A significant first attempt to rectify lighting artefacts was to define a threshold for the change rate and if the change is higher than the threshold, then the current reading is discarded and historical readings (previous frame) are considered (Poh et al., 2011), by using for example the NC-VT algorithm (Malik et al., 1989; Vila et al., 1997). Many studies show that the collection of source signals from more than one region is an important step towards removing illumination and motion artefacts (Bobbia et al., 2017; Villarroel et al., 2017; Wei et al., 2017). By using multiple regions, a common portion of the signal can be deducted, which can be assumed to contain noise and illumination artefacts. Li et al. (2014) followed this approach by using background light removal and filtering. Cheng et al. (2017) proposed an illumination variation-resistant method using joint blind source separation (JBSS) and ensemble empirical mode decomposition (EEMD). The use of LAB colour space seemed to improve performance for Fernandes et al. (2017), who reported that LAB colour space output is not affected by lighting conditions as much as the RGB colour space.

The second most significant source of artefacts is motion. Motion can be divided as rigid and non-rigid; the former referring to voluntary large movements, such as moving of limbs, posture changes or head rotations, while the latter to minute motions such as muscle vibrations, bobbing of head due to breathing and other physiological phenomena. For a stationary subject, however, the latter is more relevant as in most experiments, the subjects are requested to sit and be as still as possible. This work also considers subjects without voluntary movements, therefore, large/rigid motion effects removal is outside the scope of this study. Most approaches use face detection and face tracking throughout the recording duration, which leads to the removal of some motion effects. Effect removal can also be done using filtering techniques. Non-rigid motion filtering was applied in Li et al. (2014), resulting to improved performance. Band-pass filtering for extracting the frequencies that correspond to the human heart rate range is a common filtering method (Bosi et al., 2017), with the frequency range varying between studies, e.g. 0.7–3 Hz (Cheng et al., 2017), 0.75–3.5 Hz (Cennini et al., 2010), 0.5–3 Hz (Fernandes et al., 2017), etc. (van Gastel et al., 2015) proposed a remote heart signal extraction method with very good motion robustness using the near infra-red spectrum instead of visible light, showing that successful motion artefact compensation is possible even in the presence of severe motion artefacts. Various other denoising and artefact removal methods have been proposed in the literature, utilising a combination of filters and other methods (Elfaramawy et al., 2017; Smilkstein et al., 2014; Wu et al., 2012).

2.4. Blood Volume Pulse (BVP) signal extraction

After converting the raw data to the selected colour space, selecting the ROIs, and applying artefact removal, the next step is the extraction of the BVP signal. Independent Component Analysis (ICA) has been a popular choice for this task (Bakhtiyari et al., 2017). Poh et al. (2010, 2011) have shown promising results using ICA with Joint Approximation Diagonalisation of Eigen-matrices (JADE). McDuff et al. (2014b) used ICA and FFT to extract systolic and diastolic peaks, while Zhang et al. (2017) approached the detection of heart rate using an ICA with second order blind identification. Macwan et al. (2018) demonstrated the performance of constrained ICA in remote PPG, using periodicity and chrominance as constraints, while Zhao et al. (2013) detected both heart and respiration rates using delay-coordinate transformation and ICA-based deconstruction of single channel images. Wiede et al. (2016) applied ICA and Principal Component Analysis (PCA) on the intensity and motion signals respectively, and used their fusion to determine the heart rate.

PCA was also used along with empirical mode decomposition (Bogdan et al., 2015), singular value decomposition (SVD) (Janssen et al., 2016), variational mode decomposition (Sharma, 2019) and along FFT (Bosi et al., 2016) for remote heart rate extraction. Cennini et al. (2010) also used the FFT, while EEMD and FFT have been applied in combination in Cheng et al. (2017). A new method using stochastically obtained PPG signal was proposed in Chwyl et al. (2016) that estimates a PPG using Bayesian minimisation with a Monte Carlo sampling approach to yield a valid heart rate. Other proposed methods for PPG estimation included transmittance PPGI (Amelard et al., 2015), deep learning (Niu et al., 2018), Eulerian video magnification (Alzahrani & Whitehead, 2015), joint blind source separation (Qi et al., 2017), Independent Vector Analysis (IVA) (Qi et al., 2017), and auto-regressive modelling and pole cancellation (Tarassenko et al., 2014), among others.

2.5. Heart rate estimation

By this step, all artefacts have been filtered out to a good extent and the heart rate frequency should be the available prominent frequency. To this end, the final heart rate estimation step starts with the obtained pulse signal being converted to the frequency domain (e.g., using Welch's method) and the frequency with the highest power response is selected as the heart rate frequency. This frequency is then multiplied by 60 to compute the heart rate in beats per minute (bpm) (Kado et al., 2018; Li et al., 2014; Poh et al., 2010).

3. Methodology

It is evident from the previous section that although numerous methods for video-based remote heart rate detection have been proposed, there is a lot of overlap regarding the approaches followed. In this work, we examined the effect of some commonly used steps within a video-based remote heart rate detection pipeline, evaluated their performance on four different datasets, and proposed the best combination of methods that offers the most balanced performance across all the examined datasets. To achieve this, a software pipeline was developed, so that methods could be enabled or disabled within the pipeline, allowing the testing of the many parameters to be more systematic. It must be noted that all the examined methods in this work, as well as the experimental pipeline, were implemented by the authors using MATLAB 2018a. In addition, the Augsburg Biosignal ToolBox (AuBT) (Wagner et al., 2005) was used in order to compute the heart rate ground truth signals from the ECG recordings of the examined datasets.

3.1. Datasets

The first step was the selection of suitable datasets for the evaluation of the examined methods. Three publicly available datasets, originally created for emotion recognition studies, that contained video and ECG, BVP or heart rate information were used. In addition, a dataset created by the authors in a previous study (Malasinghe et al., 2018) was also used. An overview of the datasets used is provided in Table 1.

(i) **DEAP** (Koelstra et al., 2012) contains physiological signal recordings and videos from 22 subjects, with 40 trials per subject. The videos were recorded at 50 fps with a resolution of 1280 × 1024 and were 60 s long. During the dataset creation, the subjects sat in front of a computer screen and watched several video clips which were selected to elicit specific emotions. Their facial expressions, EEG signals, and PPG signals (BVP) were recorded at 512 Hz. The heart rate ground truth for the DEAP dataset was created using the available BVP signals and by estimating the time interval between the heartbeats (Peper et al., 2007), i.e. the high peaks within the BVP signal. It must be noted that

Table 1
Datasets.

Dataset	DEAP	MAHNOB-HCI	UBFC-RPPG	KINECT
Video recording device	Sony DCR-HC27E	Allied Vision Stingray F-046	Logitech C920 HD Pro	MS Kinect v2.0
Heart rate recording device	Biosemi Active II	Biosemi Active II	CMS50E pulse oximeter	SHIMMER v2.0
Video characteristics	1280 × 1024 @50 Hz	780 × 580 @60 Hz	640 × 480 @30 Hz	1920 × 1080 @30 Hz
Videos	40 × 22 subjects	20 × 26 subjects	1 × 42 subjects	1 × 15 subjects
Exceptions	Subjects 3,5,14 (39 videos) Subject 11 (37 videos)	Subject 3,5 (17 videos) Subject 14 (16 videos) Subject 9 not used	n/a	n/a

only 39 videos were available for subjects #3, #5, and #14, whereas only 37 videos were available for subject #11.

(ii) **MAHNOB-HCI** (Soleymani et al., 2012) contains physiological signal recordings and videos from 27 subjects, with 20 trials per subject. Videos were captured at 60 fps with a resolution of 780 × 580. Similar to DEAP, videos were captured while subjects sat and watched videos that contained material selected to elicit different types of emotions. The heart rate associated with each video recording was computed from the ECG signals contained in the dataset using the Augsburg Biosignal ToolBox (Wagner et al., 2005). It must be noted that only 17 videos were available for subject #3 and only 16 videos for subject #14. Subject #9 was omitted since the computed heart rate values were outside the human range, similar to the first three videos for subject #5.

(iii) **UBFC-RPPG** (Bobbia et al., 2017) contains PPG and video recordings from 42 subjects (1 trial each). The videos were captured at 30 fps with a resolution of 640 × 480 in 8-bit RGB format using a low cost webcam (Logitech C920 HD Pro), while the subjects sat 1 m away from the camera and played a mathematical game that aimed at varying their heart rate to mimic a real-life human–computer interaction scenario. Furthermore, the acquisition environment had varying sunlight and indoor light. Ground truth heart rates were computed from the recorded PPG signals using code provided by the UBFC-RPPG dataset creators (Bobbia et al., 2017).

(iv) **KINECT** dataset was prepared in our previous study (Malasinghe et al., 2018) and contains ECG and video recordings from 15 subjects (1 trial per subject). Kinect for Windows v2.0 sensor/camera was used to record videos at 30 fps with a resolution of 1920 × 1080. The subjects sat in front of the camera about 1 m away and were asked to sit as still as possible (avoiding large voluntary movements) for 60 s. Frames were captured as uncompressed images, while Kinect's embedded face detection and tracking mechanisms were utilised to record the coordinates of the face, eyes, nose and mouth. It must be noted that due to the experimental design in Malasinghe et al. (2018), the video recordings in the KINECT dataset include only the face regions and not the whole frames. ECG signals were captured using a SHIMMER v2.0 wireless ECG sensor (Burns et al., 2010) at a sampling rate of 512 Hz. For the computation of the ground truth, heart rates were computed from the ECG recordings using the Augsburg Biosignal ToolBox (Wagner et al., 2005).

3.2. Examined methods

3.2.1. ROI selection

The following three ROIs where examined, based on their popularity across the literature:

(i) *Full face*: The face region is detected using Haar cascades (Viola & Jones, 2001) for the DEAP, MAHNOB-HCI, and UBFC-RPPG datasets, and using the Kinect face detection mechanism for the KINECT dataset. The whole face region with a width W_{face} and a height H_{face} is selected as the ROI, as shown in Fig. 1(a).

(ii) *60% of face region width*: From the detected face region, the central 60% across the horizontal axis is selected as the ROI, with the height remaining equal to the whole face's area, as shown in Fig. 1(b). The starting and ending horizontal indexes Idx_{start} and Idx_{end} are computed as:

$$Idx_{start} = \lceil 0.2 \cdot W_{face} \rceil \tag{1}$$

$$Idx_{end} = W_{face} - Idx_{start} \tag{2}$$

By subtracting 20% of the face region's width from each side along the horizontal axis, the central 60% of the face area is obtained.

(iii) *Rectangular area on cheek*: From the detected face region, a rectangular area of width and height equal to $0.2 \cdot W_{face}$ within the cheek area is selected as the ROI, as shown in Fig. 1(c). The upper left point of the rectangle is located at (ROI_x, ROI_y), computed as follows:

$$(ROI_x, ROI_y) = (Mid_{row}, Mid_{column} + Offset) \tag{3}$$

where $Mid_{row} = \lceil \frac{W_{face}}{2} \rceil$, $Mid_{column} = \lceil \frac{H_{face}}{2} \rceil$, and $Offset = \lceil \frac{W_{face}}{2} \rceil$.

Any further processing within the remote heart rate detection pipeline is applied similarly to each ROI.

3.2.2. Source signal creation

The most common approach in the literature for converting the source image ROIs to a 1-D source signal is by computing the arithmetic mean of pixel intensities of the ROIs of each video frame for each colour channel, thus creating three (in case of RGB) time series depicting the arithmetic mean of pixel intensities across time (Bosi et al., 2016; Poh et al., 2011). Let $I_{c,t}(i,j)$ be the intensity of pixel (i,j) for colour channel c at frame t for a ROI of size $M \times N$ pixels, K the number of available video frames, $S_{c,t}$ the value of the time series for channel c at frame t, and $S_c = \{S_{c,1}, S_{c,2}, \ldots, S_{c,K}\}$. $S_{c,t}$ is computed as:

$$S_{c,t} = \frac{1}{M \cdot N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I_{c,t}(i,j) \tag{4}$$

In this work, the RGB colour space was used, thus three time series were computed for each ROI, i.e. S_R, S_G, and S_B.

The second approach for source signal creation examined in this work is the creation of time series based on ratios between different RGB channels' pixel intensities. The benefits of this approach were studied in de Haan and Jeanne (2013), showing that it leads to reduced specular reflection when the light is incident upon the skin. Following this approach, three time series were computed and used as the source signals S_1, S_2, and S_3, where $S_{x,t}$ is the value of the time series for source signal x at frame t and $S_x = \{S_{x,1}, S_{x,2}, \ldots, S_{x,K}\}$.

$$S_{1,t} = \frac{S_{G,t}}{S_{R,t}} - 1 \tag{5}$$

$$S_{2,t} = \frac{S_{G,t}}{S_{B,t}} - 1 \tag{6}$$

$$S_{3,t} = \frac{S_{B,t}}{S_{R,t}} - 1 \tag{7}$$

It is worth noting that a similar concept is used in Bobbia et al. (2017) with a slightly different equation for computing the input signals.

Fig. 1. ROI types. (a) Full face. (b) 60% of the face area. (c) Rectangular area on cheek.

3.2.3. Background light reduction

When light is incident upon the skin, some of it gets reflected, some refracted, and some absorbed. When a person is video recorded in any environment with light, the recorded video and images contain artefacts originating from the light within the environment. These artefacts affect the brightness values of the pixels, thus introducing noise and hindering the performance of processing methods that rely on pixel brightness values. Consequently, these artefacts should be reduced as much as possible. An illumination rectification approach proposed by Li et al. (2014) that uses a background area as reference to reduce the effects of illumination interference is examined in this work. In this method, the values of the time series S_x are considered to be affected by two factors, the cardiac pulse and environmental illumination variations, and these variations are assumed to be additive:

$$S_x = s_x + y_x \tag{8}$$

where s_x are the variations caused in channel x due to cardiac pulse and y_x the variations caused in channel x due to illumination interference. A rectangular region within the background, i.e. not including the examined ROI, is then used as a reference and the time series $S_x^{(Background)}$ for each channel x are computed similar to the time series S_x that refer to the examined ROIs. Then, the illumination rectified time series S'_x is computed as:

$$S'_x = S_x - h \cdot S_x^{(Background)} \tag{9}$$

where h is a constant. In this work, various values of the constant h were tested, determined through preliminary experiments.

3.2.4. Normalisation

Normalisation of the source signals is performed in the vast majority of image-based remote heart rate detection methods, as it is known to reduce noise and make the final signal independent of the source. The normalised version S''_x of a time series S_x was computed as:

$$S''_x = \frac{S_x - \mu_{S_x}}{\sigma_{S_x}} \tag{10}$$

where x denotes the colour channel or the ratio-based input, μ_{S_x} the arithmetic mean of S_x and σ_{S_x} the standard deviation of S_x.

3.2.5. De-trending

It has been found that heart rate variability (HRV) usually consists of non-stationarities due to respiratory rate oscillations and other cardiac phenomena (Berntson et al., 1997). To reduce these, the use of a smoothness priors de-trending approach that operates like a time-varying finite-impulse response (FIR) high-pass filter was proposed by Tarvainen et al. (2002). This method uses regularised least squares solution on the identified RR interval series and uses a regularisation parameter λ to control it. In this work, this de-trending method was tested for various values of λ, determined through preliminary experiments.

3.2.6. Moving average filter

The use of a moving average filter is also very common. Windows of size 5, 7, and 9 samples were examined after conducting preliminary experiments with various values and concluding that these are the most appropriate for testing.

3.2.7. Non-rigid motion filtering

The purpose of non-rigid motion elimination is to remove sudden noisy segments from the signal which are the result of motion. When sudden noisy segments are present in the signal, they can be misinterpreted as signals corresponding to heart rate, thus interfering with its correct estimation. A filtering approach proposed by Li et al. (2014) was examined in this work in order to remove segments from the signal corresponding to non-rigid motion. To this end, the time series S_x is first divided into m segments of length $\frac{K}{m}$, with K being the number of frames in the video. Then, the standard deviation $\sigma_{S_x^{(k)}}$ of each segment $S_x^{(k)}$, $k = 1, 2, \ldots, m$ is computed and the 5% of the segments with the highest $\sigma_{S_x^{(k)}}$ are discarded. The filtered time series is finally constructed by concatenating the remaining segments. In this work, m was set to 30 for all datasets apart from UBFC-RPPG ($m_{\text{UBFC-RPPG}} = 22$).

3.2.8. Independent Component Analysis (ICA)

ICA is a blind source separation technique. The purpose of blind source separation is to recover useful signals from a mixture of signals. In this work, there are three available source signals (RGB or RGB ratios) and ICA is used in order to decompose them to three independent signals in order to remove inter-dependencies. ICA can be performed using many different algorithms, such as FastICA, constrained-ICA, JADE (Cardoso & Souloumiac, 1993), RobustICA, etc. In this work, the ICA with JADE (Cardoso, 1997) was used. JADE is an algorithm that uses fourth order moments to find the independence between the various source signals present in the scenario and then uses contrast functions and matrix computations to obtain the set of source signals.

3.2.9. Selection of channel with highest power peak

In cases that multiple source signals exist, e.g. three RGB channels, the heart rate estimation pipeline is applied to each source signal, resulting in three separate heart rate estimates. The processed source signal containing the highest power peak within the frequency range of the human heart rate can be selected as the most suitable candidate for heart rate estimation, as suggested in Poh et al. (2010). It must be noted that, in this work, when ICA is applied, the independent component containing the highest power peak is always selected for the final heart rate estimation. To select the channel (or independent component) that contains the highest power peak, the periodogram power spectral density (PSD) estimate of each source signal is first computed, as shown in Fig. 2. Then, then maximum PSD amplitude within the frequency range of the human heart rate of each source signal is measured, and the source signal with the highest maximum PSD amplitude among the source signals is selected for the final heart rate estimation.

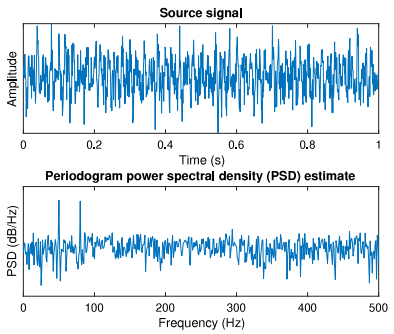

Fig. 2. Example of power spectral density estimate of a random source signal for the selection of the source signal with the highest power peak within the frequency range of the human heart rate.

3.2.10. Bandpass filtering

Since the frequency range of human heart rate is known, bandpass filtering is used in order to remove the frequency content outside this range. Furthermore, since the heart rates available in the examined datasets are lower than 120 bpm, a Hamming window-based FIR bandpass filter with a cut-off frequency of $[0.8, 2]$ Hz was applied, corresponding to heart rates between 48 and 120 bpm.

3.2.11. Power Spectral Density (PSD) estimation

To extract the heart rate from the examined signal, the frequency with the maximum power must be detected. To this end, Welch's method is used to estimate the PSD distribution, and the frequency exhibiting the maximum power response is assumed to be the heart rate frequency f_{HR}.

3.2.12. Heart rate computation

After detecting the heart rate frequency f_{HR}, the heart rate (HR), in bpm, corresponding to the examined source signal is computed as:

$$HR = 60 \cdot f_{HR} \tag{11}$$

3.3. Experimental parameters

The above methods were implemented within a software pipeline that allowed enabling and disabling steps and setting specific parameters when applicable. To provide a fair and consistent comparison across the four datasets used, all combinations of methods and parameters where tested on all four datasets. Since the duration of the video recordings varied, the first 60 s from each video were taken into consideration and the heart rates were computed for the two 30 s segments, as proposed in Li et al. (2014). The method parameters examined (when applicable) were selected through preliminary experiments that helped to determine parameter ranges that provided acceptable performance. The examined parameters are summarised in Table 2. Using just these parameters, 2304 combinations[2] of methods and parameters were evaluated for each dataset, leading to a total of $4 \times 2304 = 9216$ experiments. Experiments took 5–7 days to complete for DEAP and MAHNOB-HCI and 1–2 days for UBFC-RPPG and KINECT on a computer with an Intel® Core™ i5-4590 CPU with 4 cores at 3.30 GHz, 8 GB of DDR3 RAM, using the 64-bit Windows 8 OS. Testing a wider range of parameters would be too time consuming due to the computational time needed. It must also be noted that, as shown in Table 2, the normalisation, bandpass filtering, and PSD estimation steps were applied for all the settings examined.

[2] 3 ROIs × 2 input signals × 4 background light reduction settings × 4 de-trending settings × 3 moving average filter settings × 2 non-rigid motion estimation settings × 2 ICA settings × 2 highest power peak selection settings = 2304 settings.

Table 2
The methods examined in this work and the associated experimental parameters.

Method	Options
ROI	Full face, 60% of face, Rectangular area on cheek
Input signals	RGB channels, RGB channel ratios
Background light reduction	OFF ($h = 0$), ON ($h = 0.5, 0.75, 1$)
Normalisation	ON
De-trending	OFF, ON ($\lambda = 100, 110, 120$)
Moving average filter	$Window = 5, 7, 9$ samples
Non-rigid motion filtering	OFF, ON ($m = 30$)
ICA	OFF, ON
Highest power peak selection	OFF, ON
Bandpass filtering	$[0.8, 2]$ Hz
PSD estimation	ON

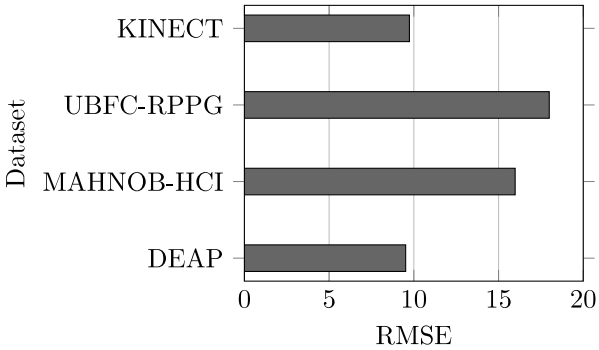

Fig. 3. Best performance per dataset in terms of RMSE.

4. Results

The developed software pipeline was used in order to evaluate the performance of all the examined combinations of methods and parameters on the four previously described datasets. The heart rate for the two 30 s segments of each video sequence and each setting was computed and compared against the ground truth. Performance was measured in terms of the Mean Error (ME), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Pearson's Correlation Coefficient (ρ) between the computed heart rate and the respective ground truth heart rate.

4.1. Overall results

The best performing combination of methods and parameters for each examined dataset is shown in Table 3 for each of the metrics examined. When RMSE is considered as the benchmark metric (Fig. 3), the best RMSE values reached 9.5191 for DEAP, 15.9801 for MAHNOB-HCI, 17.9926 for UBFC-RPPG, and 9.7389 for KINECT, while the mean error was significantly low for all the datasets, ranging from −1.15 to 2.0737 bpm, and the highest ρ exhibited very high variation ($\sigma_{\rho_{best}} = 0.21$) spanning from 0.2426 for MAHNOB-HCI to 0.7642 for the KINECT dataset.

When MAE is considered as the benchmark metric, the MAE values of the best performing configurations reached 6.6533 for DEAP, 7.2833 for KINECT, 11.8661 for MAHNOB-HCI, and 12.2696 for UBFC-RPPG, while the RMSE, ME, and ρ remained identical to when RMSE was considered as the benchmark metric, except for the MAHNOB-HCI dataset where RMSE increased to 15.9932, ρ decreased to 0.2047, and ME decreased to 1.3629.

When ρ is considered as the benchmark metric, the ρ values of the best performing configurations reached 0.7642 for KINECT, 0.5055 for DEAP, 0.4767 for UBFC-RPPG, and 0.2664 for MAHNOB-HCI, while the RMSE, MAE, and ME remained identical to when RMSE was considered as the benchmark, except for the MAHNOB-HCI dataset where RMSE

Table 3
Best performance per dataset in terms of each of the examined metrics.

Metric	Dataset	ROI	Bg light (h)	Mov Avg Filt	De-trending (λ)	Source	Non-rigid motion	ICA	Highest power	ME	MAE	RMSE	ρ
RMSE	DEAP	60% face	–	9	120	RGB	–	✓	✓	−0.89	6.6533	9.5191	0.5055
	MAHNOB-HCI	Cheek	–	9	120	RGB	–	✓	✓	2.0737	11.9049	15.9801	0.2426
	UBFC-RPPG	60% face	–	5	100	RGB	–	✓	✓	−0.0923	12.2696	17.9926	0.4767
	KINECT	60% face	–	5	120	RGB	–	✓	✓	−1.15	7.2833	9.7389	0.7642
MAE	DEAP	60% face	–	9	120	RGB	–	✓	✓	−0.89	6.6533	9.5191	0.5055
	MAHNOB-HCI	Cheek	–	9	120	RGB	–	✓	✓	1.3629	11.8661	15.9932	0.2047
	UBFC-RPPG	60% face	–	5	100	RGB	–	✓	✓	−0.0923	12.2696	17.9926	0.4767
	KINECT	60% face	–	5	120	RGB	–	✓	✓	−1.15	7.2833	9.7389	0.7642
ρ	DEAP	60% face	–	9	120	RGB	–	✓	✓	−0.89	6.6533	9.5191	0.5055
	MAHNOB-HCI	Cheek	1	9	–	RGB (G)	–	–	–	−10.2346	13.0095	16.938	0.2664
	UBFC-RPPG	60% face	–	5	100	RGB	–	✓	✓	−0.0923	12.2696	17.9926	0.4767
	KINECT	60% face	–	5	120	RGB	–	✓	✓	−1.15	7.2833	9.7389	0.7642
ME	DEAP	Cheek	0.5	5	–	RGB	✓	–	✓	−0.0123	11.4453	15.238	0.1635
	MAHNOB-HCI	60% face	–	7	110	Ratio	–	–	✓	0.0617	19.9325	24.1038	−0.0568
	UBFC-RPPG	60% face	0.75	7	100	RGB	–	✓	✓	−0.0565	13.2696	18.9444	0.4226
	KINECT	60% face	–	9	–	Ratio	✓	✓	✓	0.0042	22.1458	25.8554	−0.0253

increased to 16.938, MAE increased to 13.0095, and ME increased to −10.2346.

It is evident that the best performing configuration for the DEAP, UBFC-RPPG, and KINECT datasets is the same when considering RMSE, MAE, or ρ as the benchmark metric. Interestingly, the order of the best performing datasets varies according to the metric used. When MAE and RMSE are considered, the best results are achieved for DEAP, followed by KINECT, MAHNOB-HCI, and finally UBFC-RPPG. When ρ is considered, the best results are achieved for KINECT (0.7642), followed by DEAP (0.5055), UBFC-RPPG (0.4767), and finally MAHNOB-HCI (0.2426) which exhibited very low correlation to the ground truth heart rate.

When ME is used as the benchmark metric, the achieved MEs for all datasets are extremely low, ranging from −0.0565 to 0.0617 bpm. However, considering that for the same configurations, MAE and RMSE values are much higher than their respective values for the best configurations when other metrics are considered as benchmark, it is evident that ME is not a suitable performance metric due to the very high positive or negative errors in the predicted heart rates. As a result, ME was not considered important for the rest of this work. Considering that the positive or negative errors in the predicted heart rate should be as small as possible, the RMSE was selected as the benchmark metric for the rest of this work since it penalises large errors by design.

It is evident that the use of 60% of the face area as the ROI provided the best results for all datasets apart from MAHNOB-HCI for which the cheek ROI provided the best results. The best performing settings for all the datasets used the raw RGB channels instead of the RGB channel ratios, in combination with ICA and the selection of the channel containing the highest power peak for determining the heart rate. De-trending was also used in all the best performing settings, with $\lambda = 120$ for DEAP, MAHNOB-HCI, and KINECT, and $\lambda = 100$ for UBFC-RPPG. Regarding the size of the moving average filter's window, a window of size 9 samples performed better for DEAP and MAHNOB-HCI, while a window of size 5 samples performed better for UBFC-RPPG and KINECT.

4.2. Results per component

The effects of each method or parameter (when applicable) on the performance for each dataset was reported in Tables 4–10. The configurations reported in these tables refer to those that provided the best results in terms of RMSE for each dataset when the examined method/parameter is adjusted.

ROI: The best performance per examined ROI for each dataset is shown in Table 4 and Fig. 4(a). It is evident that the use of 60% of the face area as the ROI provided the best performance for all datasets, except for MAHNOB-HCI, for which the use of the cheek area provided a −0.8412 lower RMSE value.

Source signal: Regarding the source signal used, it is evident from Table 5 that the use of the RGB channels as the source provided the best results, with RMSE values being significantly lower than the best RMSE achieved when using the ratios of the RGB channels as the source signals. Furthermore, when ICA and the selection of the channel with the highest power peak were not used, individual RGB channels provided considerably worse performance, with the green channel providing the best performance among them for all datasets.

De-trending: From Table 6 and Fig. 4(b), it is evident that the use of de-trending leads to better RMSE values for all the examined datasets. For DEAP, MAHNOB-HCI, and KINECT, a de-trending λ equal to 120 provided the best performance. In the case of UBFC-RPPG, the best RMSE (17.9926) was achieved for $\lambda = 100$, although the difference in the RMSE achieved for $\lambda = 110$ and $\lambda = 120$ is significantly small, +0.0476 and +0.0379 respectively, thus it can be considered as insignificant.

Background light reduction: It is evident from Table 7 and Fig. 4(c) that lower values of parameter h lead to better RMSE, with the best RMSE for all datasets achieved when no background light reduction is used ($h = 0$). It must be noted that for MAHNOB-HCI, using an $h = 0.5$ leads to a 0.0131 lower RMSE than to not using the background light reduction method. However, this difference in RMSE can be considered as insignificant, thus performance is similar to when background light reduction is not applied. Furthermore, it must be noted that since only the face area of the video frames is available for the KINECT dataset, background light reduction could not be examined for it.

Moving Average Filter: From Table 8 and Fig. 4(d), it is evident that the optimal window size for the moving average filter differs among the examined datasets. For DEAP and MAHNOB-HCI, bigger window sizes led to better RMSE values, while the opposite was observed for UBFC-RPPG and KINECT. This effect can be attributed to the number of samples in the signal. The videos in the UBFC-RPPG and KINECT were recorded at 30 Hz, while videos in DEAP were recorded at 50 Hz and in MAHNOB-HCI at 60 Hz. Consequently, the optimal window size of 5 samples for UBFC-RPPG and KINECT corresponded to a window size of $\frac{5\,\text{samples}}{30\,\text{Hz}} \approx 0.167\,\text{s}$, while for DEAP the optimal window size of 9 samples corresponded to $\frac{9\,\text{samples}}{50\,\text{Hz}} = 0.18\,\text{s}$ and for MAHNOB-HCI to $\frac{9\,\text{samples}}{60\,\text{Hz}} = 0.15\,\text{s}$. Fig. 5 depicts the best RMSE achieved for each dataset in relation to the corresponding duration of the examined moving average filter window sizes. The best RMSE for all datasets was achieved for a window of size between 0.150 s and 0.180 s, corresponding to 5 samples for UBFC-RPPG and KINECT, and 9 samples for DEAP and MAHNOB-HCI.

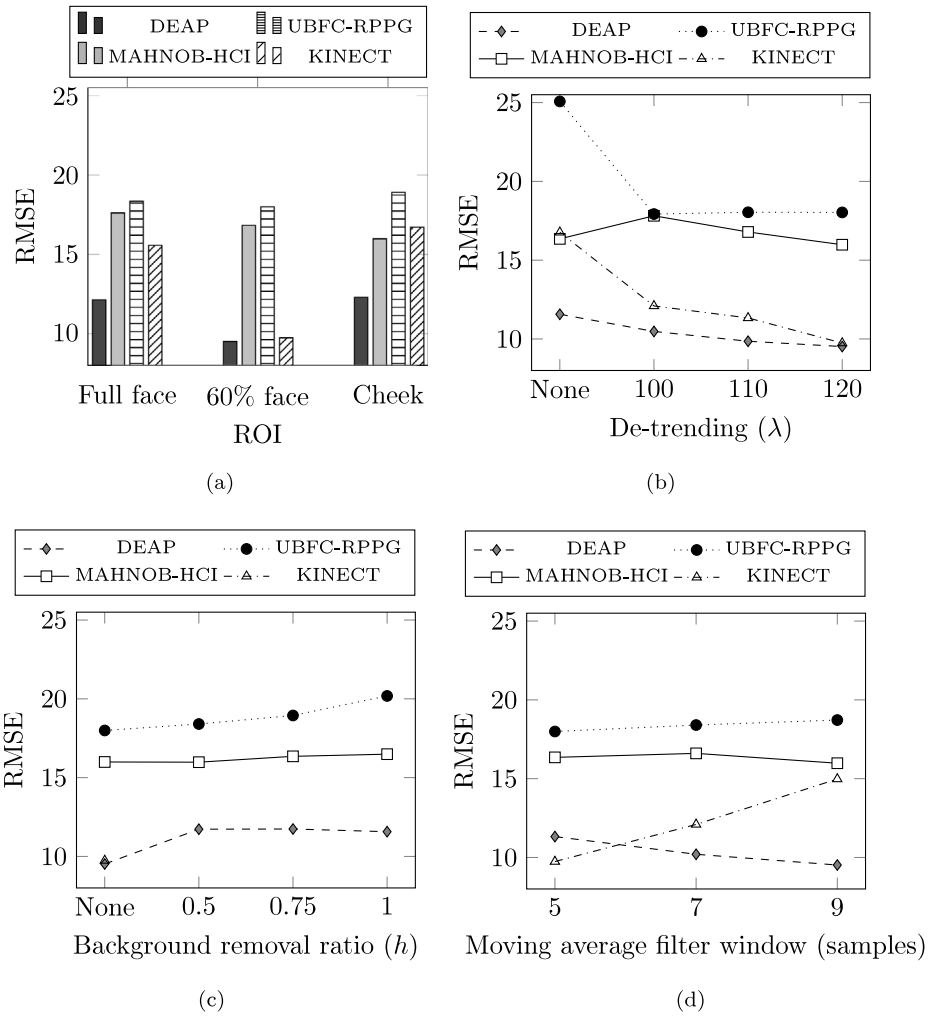

Fig. 4. The effects of different parameter values for the best performing configurations for each dataset for (a) ROI, (b) De-trending, (c) Background light removal, and (d) Moving average filter window. Note: For the background light removal plot and the KINECT dataset, results are only available when background light removal is not used.

Table 4
Best results per dataset in terms of RMSE depending on the ROI.

ROI	Dataset	Bg light (h)	Mov Avg Filt	De-trending (λ)	Source	Non-rigid motion	ICA	Highest power	ME	MAE	RMSE	ρ
Full face	DEAP	0.5	9	120	RGB	–	✓	✓	1.0618	8.4348	12.1346	0.3297
	MAHNOB-HCI	–	9	120	RGB	–	✓	✓	3.2158	13.3366	17.6052	0.1341
	UBFC-RPPG	0.75	5	100	RGB	–	✓	✓	−0.9125	13.3946	18.3438	0.4270
	KINECT	–	5	110	RGB	–	✓	✓	−8.0375	11.3708	15.5677	0.5040
60% face	DEAP	–	9	120	RGB	–	✓	✓	−0.89	6.6533	9.5191	0.5055
	MAHNOB-HCI	–	9	120	RGB	–	✓	✓	1.8881	12.4083	16.8213	0.1635
	UBFC-RPPG	–	5	100	RGB	–	✓	✓	−0.0923	12.2696	17.9926	0.4767
	KINECT	–	5	120	RGB	–	✓	✓	−1.15	7.2833	9.7389	0.7642
Cheek	DEAP	0.5	9	120	RGB	–	✓	✓	2.7503	8.3007	12.2826	0.2901
	MAHNOB-HCI	–	9	120	RGB	–	✓	✓	2.0737	11.9049	15.9801	0.2426
	UBFC-RPPG	1	5	100	RGB	–	✓	✓	−1.9125	13.2804	18.8946	0.4587
	KINECT	–	5	100	RGB	–	✓	✓	−11.2125	14.0792	16.7068	0.5489

Non-rigid motion filtering: The use of the non-rigid motion filtering led to worse RMSE values for all datasets, as shown in Table 9. Consequently, the best results achieved for each dataset, when non-rigid motion filtering is not used, correspond to the overall best results for the examined datasets.

ICA: The use of ICA led to better RMSE values for all datasets, as shown in Table 10. Consequently, the best results achieved for each dataset, when ICA is used, correspond to the overall best results for the examined datasets.

4.3. Best performing method combination

It is evident from Table 3 that when RMSE is used as the benchmark metric, the best performing configurations for all the examined datasets do not use the background light reduction technique, use de-trending, use the RGB channels as the source, do not use non-rigid motion filtering, use ICA, and use the independent component containing the highest power peak for the extraction of the final heart rate. However, some parameters of the methods used are not similar for all the examined datasets. The selected ROI was 60% of the face area for all

Table 5
Best results per dataset in terms of RMSE depending on the source signal.

Source	Dataset	ROI	Bg light (h)	Mov Avg Filt	De-trending (λ)	Non-rigid motion	ICA	Highest power	ME	MAE	RMSE	ρ
RGB	DEAP	60% face	–	9	120	–	✓	✓	−0.89	6.6533	9.5191	0.5055
	MAHNOB-HCI	Cheek	–	9	120	–	✓	✓	2.0737	11.9049	15.9801	0.2426
	UBFC-RPPG	60% face	–	5	100	–	✓	✓	−0.0923	12.2696	17.9926	0.4767
	KINECT	60% face	–	5	120	–	✓	✓	−1.15	7.2833	9.7389	0.7642
Ratio	DEAP	Full face/Cheek	–	9	–	✓	✓	✓	8.6489	17.347	21.6632	0.0437
	MAHNOB-HCI	Cheek	0.75	7	110	✓	✓	✓	5.1872	17.5987	21.7492	0.0654
	UBFC-RPPG	60% Face	0.5	7	110	–	✓	✓	−7.9054	18.1351	23.7117	0.2022
	KINECT	Cheek	–	9	–	–	✓	✓	−3.275	16	19.3712	0.4096
RGB (R)	DEAP	Cheek	0.5	9	–	✓	–	–	−2.1272	10.9628	13.6398	0.1335
	MAHNOB-HCI	Cheek	0.5	5	–	✓	–	–	−7.8203	13.3401	17.1324	0.1890
	UBFC-RPPG	60% face	0.75	7	120	✓	–	–	−8.178	18.1863	23.9274	0.0698
	KINECT	Cheek	–	5	–	–	–	–	−4.1833	17.7917	20.0984	0.4232
RGB (G)	DEAP	Cheek	1	9	–	✓	–	–	−3.5088	8.6863	11.5693	0.2858
	MAHNOB-HCI	Cheek	0.75	5	–	✓	–	–	−8.2659	12.5371	16.3518	0.2528
	UBFC-RPPG	60% face	1	5	110	✓	–	–	−7.6744	15.7685	21.6047	0.3148
	KINECT	Cheek	–	5	120	✓	–	–	6.2375	16.0625	18.3778	0.5392
RGB (B)	DEAP	Cheek	1	9	–	–	–	–	−3.8118	10.3052	13.3693	0.1942
	MAHNOB-HCI	Cheek	0.75	7	–	✓	–	–	−8.5748	13.7615	17.4265	0.1665
	UBFC-RPPG	60% face	1	7	120	–	–	–	−10.8351	19.9149	25.6766	0.1968
	KINECT	Cheek	–	7	100	–	–	–	−12.2333	16.325	18.8953	0.4187

Table 6
Best results per dataset in terms of RMSE depending on de-trending.

De-trending (λ)	Dataset	ROI	Bg light (h)	Mov Avg Filt	Source	Non-rigid motion	ICA	Highest power	ME	MAE	RMSE	ρ
–	DEAP	Cheek	1	9	RGB (G)	✓	–	–	−3.5088	8.6863	11.5693	0.2858
	MAHNOB-HCI	Cheek	0.75	5	RGB (G)	✓	–	–	−8.2659	12.5371	16.3518	0.2528
	UBFC-RPPG	60% face	0.75	5	Ratio (S_2)	✓	–	–	−9.6161	20.397	25.073	0.2007
	KINECT	Full face	–	5	RGB	–	✓	✓	−9.925	13.7	16.7669	0.4891
100	DEAP	60% face	–	9	RGB	–	✓	✓	1.1537	7.0875	10.4786	0.4745
	MAHNOB-HCI	Cheek	0.5	9	RGB (G)	✓	–	–	4.9129	13.378	17.817	0.2105
	UBFC-RPPG	60% face	–	5	RGB	–	✓	✓	−0.0923	12.2696	17.9926	0.4767
	KINECT	60% face	–	7	RGB	–	✓	✓	−3.125	8.3	12.0959	0.6547
110	DEAP	60% face	–	9	RGB	–	✓	✓	0.0631	6.7831	9.8552	0.4972
	MAHNOB-HCI	Cheek	–	9	RGB	–	✓	✓	3.3003	12.503	16.7898	0.1804
	UBFC-RPPG	60% face	–	5	RGB	–	✓	✓	−0.1208	12.2982	18.0402	0.4757
	KINECT	60% face	–	5	RGB	–	✓	✓	−2.7	8.2083	11.3345	0.6819
120	DEAP	60% face	–	9	RGB	–	✓	✓	−0.89	6.6533	9.5191	0.5055
	MAHNOB-HCI	Cheek	0.5	9	RGB	–	✓	✓	2.0737	11.9049	15.9801	0.2426
	UBFC-RPPG	60% face	–	5	RGB	–	✓	✓	0.2839	12.2768	18.0305	0.4680
	KINECT	60% face	–	5	RGB	–	✓	✓	−1.15	7.2833	9.7389	0.7642

Table 7
Best results per dataset in terms of RMSE depending on background light reduction.

Bg light (h)	Dataset	ROI	Mov Avg Filt	De-trending (λ)	Source	Non-rigid motion	ICA	Highest power	ME	MAE	RMSE	ρ
–	DEAP	60% face	9	120	RGB	–	✓	✓	−0.89	6.6533	9.5191	0.5055
	MAHNOB-HCI	Cheek	9	120	RGB	–	✓	✓	1.3629	11.8661	15.9932	0.2047
	UBFC-RPPG	60% face	5	100	RGB	–	✓	✓	−0.0923	12.2696	17.9926	0.4767
	KINECT	60% face	5	120	RGB	–	✓	✓	−1.15	7.2833	9.7389	0.7642
0.5	DEAP	Cheek	9	–	RGB (G)	–	–	–	−4.6424	8.5553	11.5879	0.2745
	MAHNOB-HCI	Cheek	9	120	RGB	–	✓	✓	2.0737	11.9049	15.9801	0.2426
	UBFC-RPPG	60% face	7	120	RGB	–	✓	✓	−0.7768	12.7423	18.4022	0.4669
	KINECT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
0.75	DEAP	Cheek	9	–	RGB (G)	–	–	–	−4.5575	8.7257	11.7397	0.2867
	MAHNOB-HCI	Cheek	5	–	RGB (G)	✓	–	–	−8.2659	12.5371	16.3518	0.2528
	UBFC-RPPG	60%	7	100	RGB	–	✓	✓	−0.0565	13.2696	18.9444	0.4226
	KINECT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
1	DEAP	Cheek	9	–	RGB (G)	✓	–	–	−3.5088	8.6863	11.5693	0.2858
	MAHNOB-HCI	Cheek	9	120	RGB (G)	✓	–	–	2.5125	12.3098	16.4885	0.2637
	UBFC-RPPG	Cheek	5	110	RGB	–	✓	✓	−6.5292	11.8288	20.1839	0.449
	KINECT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

datasets, except for MAHNOB-HCI where the cheek area was used, the de-trending λ was equal to 120 for all datasets except for UBFC-RPPG, where it was equal to 100, and the size of the moving average filter window was 9 samples for DEAP and MAHNOB-HCI, while it was 5 samples for UBFC-RPPG and KINECT. To determine a configuration that would provide the most balanced performance across all datasets,

Table 8
Best results per dataset in terms of RMSE depending on the moving average filter window.

Mov Avg Filt	Dataset	ROI	Bg light (h)	De-trending (λ)	Source	Non-rigid motion	ICA	Highest power	ME	MAE	RMSE	ρ
5	DEAP	60% face	–	120	RGB	–	✓	✓	0.709	7.6002	11.324	0.4336
	MAHNOB-HCI	Cheek	0.75	–	RGB (G)	✓	–	–	−8.2659	12.5371	16.3518	0.2528
	UBFC-RPPG	60% face	–	100	RGB	–	✓	✓	−0.0923	12.2696	17.9926	0.4767
	KINECT	60% face	–	120	RGB	–	✓	✓	−1.15	7.2833	9.7389	0.7642
7	DEAP	60% face	–	120	RGB	–	✓	✓	−0.1086	6.9875	10.2012	0.4888
	MAHNOB-HCI	Cheek	1	–	RGB (G)	✓	–	–	−8.5718	12.7949	16.6050	0.2411
	UBFC-RPPG	60% face	0.5	120	RGB	–	✓	✓	−0.7768	12.7423	18.4022	0.4669
	KINECT	60% face	–	100	RGB	–	✓	✓	−3.125	8.3000	12.0959	0.6547
9	DEAP	60% face	–	120	RGB	–	✓	✓	−0.89	6.6533	9.5191	0.5055
	MAHNOB-HCI	Cheek	0.5	120	RGB	–	✓	✓	2.0737	11.9049	15.9801	0.2426
	UBFC-RPPG	60% face	0.5	100	RGB	–	✓	✓	−1.0506	13.0327	18.7194	0.4318
	KINECT	60% face	–	120	RGB	–	✓	✓	−6.475	10.4917	14.9773	0.4985

Table 9
Best results per dataset in terms of RMSE depending on non-rigid motion filtering.

Non-rigid motion	Dataset	ROI	Bg light (h)	Mov Avg Filt	De-trending (λ)	Source	ICA	Highest power	ME	MAE	RMSE	ρ
–	DEAP	60% face	–	9	120	RGB	✓	✓	−0.89	6.6533	9.5191	0.5055
	MAHNOB-HCI	Cheek	0.5	9	120	RGB	✓	✓	2.0737	11.9049	15.9801	0.2426
	UBFC-RPPG	60% face	–	5	100	RGB	✓	✓	−0.0923	12.2696	17.9926	0.4767
	KINECT	60% face	–	5	120	Ratio	✓	✓	−1.15	7.2833	9.7389	0.7642
✓	DEAP	Cheek	1	9	–	RGB (G)	–	–	−3.5088	8.6863	11.5693	0.2858
	MAHNOB-HCI	Cheek	0.5	7	–	RGB	–	✓	−8.491	12.9494	16.6872	0.2321
	UBFC-RPPG	60% face	0.75	5	120	RGB	✓	✓	−3.8244	14.1958	19.4477	0.4100
	KINECT	60% face	–	5	100	RGB	✓	✓	−7.4208	12.8708	16.2005	0.4560

Table 10
Best results per dataset in terms of RMSE depending on ICA.

ICA	Dataset	ROI	Bg light (h)	Mov Avg Filt	De-trending (λ)	Source	Non-rigid motion	Highest power	ME	MAE	RMSE	ρ
–	DEAP	Cheek	1	9	–	RGB (G)	✓	–	−3.5088	8.6863	11.5693	0.2858
	MAHNOB-HCI	Cheek	0.8	5	–	RGB (G)	✓	–	−8.2659	12.5371	16.3518	0.2528
	UBFC-RPPG	60% face	1	5	110	RGB (G)	✓	–	−7.6744	15.7685	21.6047	0.3148
	KINECT	Cheek	–	5	120	Ratio (S_2)	✓	–	6.2375	16.0625	18.3778	0.5392
✓	DEAP	60% face	–	9	120	RGB	–	✓	−0.89	6.6533	9.5191	0.5055
	MAHNOB-HCI	Cheek	0.5	9	120	RGB	–	✓	2.0737	11.9049	15.9801	0.2426
	UBFC-RPPG	60% face	–	5	100	RGB	–	✓	−0.0923	12.2696	17.9926	0.4767
	KINECT	60% face	–	5	120	RGB	–	✓	−1.15	7.2833	9.7389	0.7642

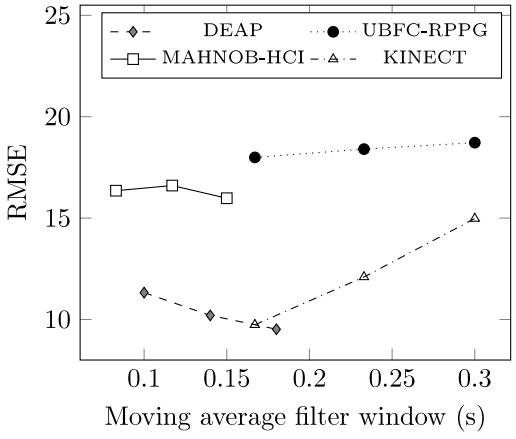

Fig. 5. The effects of the moving average filter window duration (s) for the best performing configuration across all datasets.

the performance using the parameter values that provided the best performance for the majority of datasets was examined for the different best configurations:

- If the de-trending λ is set to 120 for the best performing configuration for the UBFC-RPPG dataset, the acquired RMSE reaches 18.0305, as also shown in Table 6. The difference in RMSE from the best performing configuration for UBFC-RPPG is only 0.0379, which can be considered as insignificant. Consequently, a de-trending $\lambda = 120$ can be considered as the overall best λ for all datasets.
- If the 60% of the face area is set as the ROI for the best performing configuration for the MAHNOB-HCI dataset, the acquired RMSE reaches 16.8213, as also shown in Table 4. The difference in RMSE from the best performing configuration for MAHNOB-HCI (15.9932) is 0.8281, which can be considered acceptable for determining the configuration that provides balanced performance across all datasets. Consequently, the 60% of the face area can be considered as the overall best ROI for all datasets.
- Establishing the moving average filter window size that would provide the most balanced results is more complex since a size of 9 works best for two datasets and a size of 5 for the remaining two. However, as explained in Section 4.2, the optimal moving average filter window size in terms of samples varied because of the different sampling rates of the signals from each dataset. When the window size is examined in terms of duration (s), the best RMSE was achieved for a window of size between 0.150 s and 0.180 s, which corresponds to 9 samples for DEAP and MAHNOB-HCI, and 5 samples for UBFC-RPPG and KINECT.

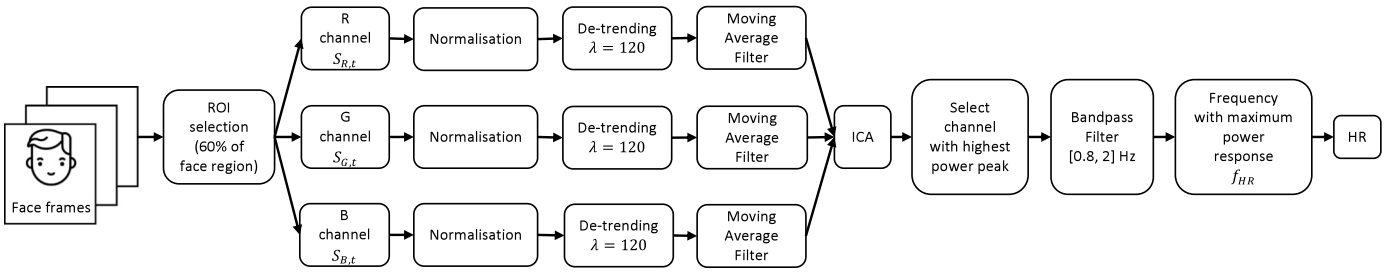

Fig. 6. Flowchart of the best performing method combination.

Table 11
Configuration that provided the most balanced performance across all datasets.

Method	Options
ROI	60% of face
Input signals	RGB channels
Background light reduction	OFF ($h = 0$)
Normalisation	ON
De-trending	ON ($\lambda = 120$)
Moving average filter	$[0.15 - 0.18]$ s
Non-rigid motion correction	OFF
ICA	ON
Highest power peak selection	ON
Bandpass filtering	$[0.8, 2]$ Hz

The configuration that provided the most balanced performance across all four datasets is summarised in Table 11, while its flowchart is provided in Fig. 6.

4.4. Computational complexity

One important aspect of a heart rate detection algorithm is its computational complexity. Considering its potential applications in health monitoring, a video-based remote heart rate detection algorithm would ideally be computed in real-time or in near real-time speeds. The video processing part of the proposed optimal configuration is limited to performing face detection and computing the arithmetic mean of the pixel intensity within the region of interest for each frame. Both tasks are highly parallelisable, and modern multi-core CPUs can perform them in real-time at very high frame rates. Any subsequent processing is then performed on one-dimensional time series containing the mean pixel intensity for each frame. Given that only signals of short duration will be examined at each given heart rate estimation iteration, that simple arithmetic operations on 1D signals can be computed extremely fast, that the use of FFT makes signal filtering very computationally efficient on multi-core CPUs, and that ICA can be efficiently performed on multi-core CPUs using available highly optimised linear algebra libraries, the computational complexity of the proposed optimal configuration is suitable for real-time computations on modern computers and mobile devices.

When compared to other video-based remote heart detection algorithms like (Li et al., 2014; Poh et al., 2010, 2011; van Gastel et al., 2015), the proposed optimal configuration benefits from the absence of a background light reduction step and a non-rigid/voluntary motion filtering step, while the rest of the steps are similar in complexity. The Kado et al. (2018) approach is similar to the proposed optimal configuration, lacking the de-trending step, but uses ROIs from multiple channels and fuses their histogram information after a histogram voting step, thus being more complex than the proposed optimal configuration. The Bernacchia et al. (2014) approach is more simplistic, lacking the normalisation, de-trending, moving average filtering and bandpass filtering steps of the proposed optimal configuration, thus being less robust to noisy signals. Other methods, like the (Tarassenko et al., 2014) approach, are considerably more complex, employing steps like image registration for facial landmark detection, image segmentation, auto-regressive modelling, and pole cancellation and selection. Consequently, the proposed optimal configuration exhibits a balanced performance across various datasets, while keeping computational complexity sufficiently low.

5. Conclusion

In this work, common methods used in video-based remote heart rate detection algorithms were examined in order to evaluate their effect on the overall performance of the remote heart rate detection pipeline. Various parameters of the examined methods were evaluated on three public and one proprietary dataset in order to establish a video-based remote heart rate detection pipeline that provides the most balanced performance across various diverse datasets, contrary to most works in the literature that rely on only one dataset for their results and fine-tune the proposed methods for the used dataset. Through the experimental evaluation, it was shown that the use of the 60% of the face area as the ROI, the use of the RGB channels as the source signal, the use of de-trending with $\lambda = 120$, the use of a moving average filter with a window size between 0.15 s and 0.18 s, and the use of ICA, provided the most balanced performance in terms of RMSE across the four examined datasets.

The use of 60% of the face area as the ROI provided better results compared to using the full face or a rectangular area at the cheek. Using the RGB channels as the source signal provided significantly better results compared to the ratios of the RGB signals. Furthermore, in cases where ICA and the selection of the channel with the highest power peak is not used, the Green channel provided the best performance, on par with the well-established fact that the green channel is a better candidate for remote PPG than other channels. The use of de-trending provided better results for all datasets, with higher λ leading to lower RMSE in most cases. Contrary to this, the use of background light reduction led to higher RMSE values for all datasets, with higher h leading to worse RMSE in most cases. Similarly, the use of non-rigid motion filtering always resulted in worse performance, while the use of ICA led to significantly better performance.

An interesting observation from the experimental results is that several inter-dependencies exist between the examined parameters and methods. This work attempted to identify how individual parameters affect the performance of the remote heart rate detection pipeline, as well as identify some performance trends when various parameters and methods are used together. Future work will focus on identifying these potential inter-dependencies between the various methods and parameters, as well as examine other methods that can be integrated in the video-based remote heart rate detection pipeline.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially funded by the Erasmus Mundus - Action 2: SmartLink Project.

References

Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, *28*(3), R1–R39. http://dx.doi.org/10.1088/0967-3334/28/3/R01.

Alzahrani, A., & Whitehead, A. (2015). Preprocessing realistic video for contactless heart rate monitoring using video magnification. In *2015 12th conference on computer and robot vision (CRV)* (pp. 261–268). http://dx.doi.org/10.1109/CRV.2015.41.

Amelard, R., Scharfenberger, C., Kazemzadeh, F., Pfisterer, K. J., Lin, B. S., Clausi, D. A., & Wong, A. (2015). Feasibility of long-distance heart rate monitoring using transmittance photoplethysmographic imaging (PPGI). *Scientific Reports*, *5*, 14637. http://dx.doi.org/10.1038/srep14637.

Bakhtiyari, K., Beckmann, N., & Ziegler, J. (2017). Contactless heart rate variability measurement by IR and 3D depth sensors with respiratory sinus arrhythmia. *Procedia Computer Science*, *109*(Supplement C), 498–505. http://dx.doi.org/10.1016/j.procs.2017.05.319.

Basri, R., & Jacobs, D. W. (2003). Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(2), 218–233. http://dx.doi.org/10.1109/TPAMI.2003.1177153.

Bernacchia, N., Scalise, L., Casacanditella, L., Ercoli, I., Marchionni, P., & Tomasini, E. P. (2014). Non contact measurement of heart and respiration rates based on kinect. In *2014 IEEE international symposium on medical measurements and applications (MeMeA)* (pp. 1–5). http://dx.doi.org/10.1109/MeMeA.2014.6860065.

Berntson, G. G., Bigger, J. T., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., Nagaraja, H. N., Porges, S. W., Saul, J. P., Stone, P. H., & Molen, M. W. D. E. R. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology*, *34*(6), 623–648. http://dx.doi.org/10.1111/j.1469-8986.1997.tb02140.x.

Bobbia, S., Benezeth, Y., & Dubois, J. (2016). Remote photoplethysmography based on implicit living skin tissue segmentation. In *2016 23rd international conference on pattern recognition (ICPR)* (pp. 361–365). http://dx.doi.org/10.1109/ICPR.2016.7899660.

Bobbia, S., Macwan, R., Benezeth, Y., Mansouri, A., & Dubois, J. (2017). Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, *124*, 82–90. http://dx.doi.org/10.1016/j.patrec.2017.10.017.

Bogdan, G., Radu, V., Octavian, F., Alin, B., Constantin, M., & Cristian, C. (2015). Remote assessment of heart rate by skin color processing. In *2015 IEEE international black sea conference on communications and networking (BlackSeaCom)* (pp. 112–116). http://dx.doi.org/10.1109/BlackSeaCom.2015.7185097.

Bosi, I., Cogerino, C., & Bazzani, M. (2016). Real-time monitoring of heart rate by processing of microsoft kinect v2.0 generated streams. In *2016 international multidisciplinary conference on computer and energy science (SpliTech)* (pp. 1–6). http://dx.doi.org/10.1109/SpliTech.2016.7555944.

Bosi, I., Cogerino, C., & Bazzani, M. (2017). Real-time monitoring of heart rate by processing of near infrared generated streams. In *SMART2017: Sixth international conference on smart cities, systems, devices and technologies* (pp. 19–24).

Burns, A., Greene, B. R., McGrath, M. J., O'Shea, T. J., Kuris, B., Ayer, S. M., Stroiescu, F., & Cionca, V. (2010). SHIMMER - A wireless sensor platform for noninvasive biomedical research. *IEEE Sensors Journal*, *10*(9), 1527–1534. http://dx.doi.org/10.1109/JSEN.2010.2045498.

Cardoso, J.-F. (1997). Blind separation of real signals using JADE. URL: http://www.indiana.edu/{~}pcl/busey/temp/eeglabtutorial4.301/allfunctions/jader.m.

Cardoso, J., & Souloumiac, A. (1993). Blind beamforming for non-Gaussian signals. *IEE Proceedings F - Radar and Signal Processing*, *140*(6), 362–370. http://dx.doi.org/10.1049/ip-f-2.1993.0054.

Cennini, G., Arguel, J., Akşit, K., & van Leest, A. (2010). Heart rate monitoring via remote photoplethysmography with motion artifacts reduction. *Optical Express*, *18*(5), 4867–4875. http://dx.doi.org/10.1364/OE.18.004867.

Cheng, J., Chen, X., Xu, L., & Wang, Z. J. (2017). Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition. *IEEE Journal of Biomedical and Health Informatics*, *21*(5), 1422–1433. http://dx.doi.org/10.1109/JBHI.2016.2615472.

Chwyl, B., Chung, A. G., Amelard, R., Deglint, J., Clausi, D. A., & Wong, A. (2016). SAPPHIRE: STochastically acquired photoplethysmogram for heart rate inference in realistic environments. In *2016 IEEE international conference on image processing (ICIP)* (pp. 1230–1234). http://dx.doi.org/10.1109/ICIP.2016.7532554.

Costa, G. D. (1995). Optical remote sensing of heartbeats. *Optics Communications*, *117*(5), 395–398. http://dx.doi.org/10.1016/0030-4018(95)00181-7.

Datcu, D., Cidota, M., Lukosch, S., & Rothkrantz, L. (2013). Noncontact automatic heart rate analysis in visible spectrum by Specific Face Regions. In *CompSysTech '13, 14th international conference on computer systems and technologies* (pp. 120–127). New York, NY, USA: ACM, http://dx.doi.org/10.1145/2516775.2516805.

Davidovic, G., Iric-Cupic, V., Milanov, S., Dimitijevic, A., & Petrovic-Janicijevic, M. (2013). When heart goes "boom" to fast. Heart rate greater than 80 as mortality predictor in acute myocardial infarction. *American Journal of Cardiovascular Diseases*, *3*(3), 120–128.

de Haan, G., & Jeanne, V. (2013). Robust pulse rate from chrominance-based rPPG. *IEEE Transactions on Biomedical Engineering*, *60*(10), 2878–2886. http://dx.doi.org/10.1109/TBME.2013.2266196.

Elfaramawy, T., Fall, C. L., Morisette, M., Lellouche, F., & Gosselin, B. (2017). Wireless respiratory monitoring and coughing detection using a wearable patch sensor network. In *2017 15th IEEE international new circuits and systems conference (NEWCAS)* (pp. 197–200). http://dx.doi.org/10.1109/NEWCAS.2017.8010139.

Fan, Q., & Li, K. (2018). Non-contact remote estimation of cardiovascular parameters. *Biomedical Signal Processing and Control*, *40*, 192–203. http://dx.doi.org/10.1016/j.bspc.2017.09.022.

Fernandes, S. L., Gurupur, V. P., Sunder, N. R., Arunkumar, N., & Kadry, S. (2017). A novel nonintrusive decision support approach for heart rate measurement. *Pattern Recognition Letters*, http://dx.doi.org/10.1016/j.patrec.2017.07.002.

Gupta, P., Bhowmick, B., & Pal, A. (2017). Serial fusion of Eulerian and Lagrangian approaches for accurate heart-rate estimation using face videos. In *2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 2834–2837). http://dx.doi.org/10.1109/EMBC.2017.8037447.

Haque, M. A., Irani, R., Nasrollahi, K., & Moeslund, T. B. (2016). Heartbeat rate measurement from facial video. *IEEE Intelligence Systems*, *31*(3), 40–48. http://dx.doi.org/10.1109/MIS.2016.20.

Hassan, M. A., Malik, A. S., Fofi, D., Saad, N., Karasfi, B., Ali, Y. S., & Meriaudeau, F. (2017). Heart rate estimation using facial video: A review. *Biomedical Signal Processing and Control*, *38*(Supplement C), 346–360. http://dx.doi.org/10.1016/j.bspc.2017.07.004.

Hori, M., & Okamoto, H. (2012). Heart rate as a target of treatment of chronic heart failure. *Journal of Cardiology*, *60*(2), 86–90. http://dx.doi.org/10.1016/j.jjcc.2012.06.013.

Janssen, R., Wang, W., Moço, A., & de Haan, G. (2016). Video-based respiration monitoring with automatic region of interest detection. *Physiological Measurement*, *37*(1), 100–114. http://dx.doi.org/10.1088/0967-3334/37/1/100.

Kado, S., Monno, Y., Moriwaki, K., Yoshizaki, K., Tanaka, M., & Okutomi, M. (2018). Remote heart rate measurement from RGB-nir video based on spatial and spectral face patch selection. In *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 5676–5680). http://dx.doi.org/10.1109/EMBC.2018.8513464.

Kobayashi, H. (2013). Effect of measurement duration on accuracy of pulse-counting. (2013/10/11). *Ergonomics*, *56*(12), 1940–1944. http://dx.doi.org/10.1080/00140139.2013.840743.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patra, I. (2012). DEAP: A Database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing*, *3*(1), 18–31. http://dx.doi.org/10.1109/T-AFFC.2011.15.

Kranjec, J., Beguš, S., Geršak, G., šinkovec, M., Drnovšek, J., & Hudoklin, D. (2017). Design and clinical evaluation of a non-contact heart rate variability measuring device. *Sensors*, *17*(11), 2637. http://dx.doi.org/10.3390/s17112637.

Li, X., Chen, J., & Pietikäinen, M. (2014). Remote heart rate measurement from face videos under realistic situations. In *2014 IEEE conference on computer vision and pattern recognition* (pp. 4264–4271). http://dx.doi.org/10.1109/CVPR.2014.543.

Macwan, R., Benezeth, Y., & Mansouri, A. (2018). Remote photoplethysmography with constrained ICA using periodicity and chrominance constraints. *Biomedical Engineering Online*, *17*(1), 22. http://dx.doi.org/10.1186/s12938-018-0450-3.

Malasinghe, L., Katsigiannis, S., Ramzan, N., & Dahal, K. (2018). Remote heart rate extraction using microsoft kinecttm V2.0. In *10th international conference on bioinformatics and biomedical technology (ICBBT)* (pp. 1–6). http://dx.doi.org/10.1145/3232059.3232060.

Malasinghe, L. P., Ramzan, N., & Dahal, K. (2017). Remote patient monitoring: a comprehensive study. *Journal of Ambient Intelligence and Humanized Computing*, *10*, 57–76. http://dx.doi.org/10.1007/s12652-017-0598-x.

Malik, M., Cripps, T., Farrell, T., & Camm, A. (1989). Prognostic value of heart rate variability after myocardial infarction. a comparison of different data-processing methods. *Medical and Biological Engineering and Computing*, *27*, 603–611. http://dx.doi.org/10.1007/BF02441642.

McDuff, D., Estepp, J., Piasecki, A., & Blackford, E. (2015). A survey of remote optical photoplethysmographic imaging methods. In *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 6398–6404). http://dx.doi.org/10.1109/EMBC.2015.7319857.

McDuff, D., Gontarek, S., & Picard, R. W. (2014a). Improvements in remote cardiopulmonary measurement using a five band digital camera. *IEEE Transactions Biomedical Engineering*, *61*(10), 2593–2601. http://dx.doi.org/10.1109/TBME.2014.2323695.

McDuff, D., Gontarek, S., & Picard, R. W. (2014b). Remote detection of photoplethysmographic systolic and diastolic peaks using a digital camera. *IEEE Transactions on Biomedical Engineering*, *61*(12), 2948–2954. http://dx.doi.org/10.1109/TBME.2014.2340991.

Moço, A. V., Stuijk, S., & de Haan, G. (2018). New insights into the origin of remote PPG signals in visible light and infrared. *Scientific Reports*, *8*(1), 8501. http://dx.doi.org/10.1038/s41598-018-26068-2.

Moraes, J., Rocha, M., Vasconcelos, G., Vasconcelos Filho, J., de Albuquerque, V., & Alexandria, A. (2018). Advances in photopletysmography signal analysis for biomedical applications. *Sensors*, *18*(6), http://dx.doi.org/10.3390/s18061894.

Niu, X., Han, H., Shan, S., & Chen, X. (2018). SynRhythm: LEarning a deep heart rate estimator from general to specific. In *2018 24th international conference on pattern recognition (ICPR)* (pp. 3580–3585). http://dx.doi.org/10.1109/ICPR.2018.8546321.

Pal, A., Sinha, A., Dutta Choudhury, A., Chattopadyay, T., & Visvanathan, A. (2013). A robust heart rate detection using smart-phone video. In *MobileHealth '13*, *3rd ACM MobiHoc workshop on pervasive wireless healthcare* (pp. 43–48). New York, NY, USA: ACM, http://dx.doi.org/10.1145/2491148.2491156.

Parnandi, A., & Gutierrez-Osuna, R. (2013). Contactless measurement of heart rate variability from pupillary fluctuations. In *2013 humaine association conference on affective computing and intelligent interaction (ACII)* (pp. 191–196). http://dx.doi.org/10.1109/ACII.2013.38.

Peper, E., Harvey, R., Lin, I.-M., Tylova, H., & Moss, D. W. (2007). Is there more to blood volume pulse than heart rate variability, respiratory sinus arrhythmia, and cardiorespiratory synchrony? *Biofeedback*, *35*(2), 54–61.

Pickering, D. (2013). How to measure the pulse. *Community Eye Health*, *26*(82), 37.

Poh, M.-Z., McDuff, D. J., & Picard, R. W. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation.. *Optical Express*, *18*(10), 10762–10774. http://dx.doi.org/10.1364/OE.18.010762.

Poh, M.-Z., McDuff, D., & Picard, R. W. (2011). Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, *58*(1), 7–11. http://dx.doi.org/10.1109/TBME.2010.2086456.

Procházka, A., Schätz, M., Vyšata, O. r., & Vališ, M. (2016). Microsoft kinect visual and depth sensors for breathing and heart rate analysis. *Sensors*, *16*(7), 996. http://dx.doi.org/10.3390/s16070996.

Qi, H., Guo, Z., Chen, X., Shen, Z., & Jane Wang, Z. (2017). Video-based human heart rate measurement using joint blind source separation. *Biomedical Signal Processing and Control*, *31*, 309–320. http://dx.doi.org/10.1016/j.bspc.2016.08.020.

Sathyanarayana, S., Satzoda, R. K., Sathyanarayana, S., & Thambipillai, S. (2015). Vision-based patient monitoring: a comprehensive review of algorithms and technologies. *Journal of Ambient Intelligence and Humanized Computing*, *9*, 225–251. http://dx.doi.org/10.1007/s12652-015-0328-1.

Sharma, H. (2019). Heart rate extraction from PPG signals using variational mode decomposition. *Biocybernetics and Biomedical Engineering*, *39*(1), 75–86. http://dx.doi.org/10.1016/j.bbe.2018.11.001.

Sinex, J. E. (1999). Pulse oximetry: Principles and limitations. *The American Journal of Emergency Medicine*, *17*(1), 59–66. http://dx.doi.org/10.1016/S0735-6757(99)90019-0.

Smilkstein, T., Buenrostro, M., Kenyon, A., Lienemann, M., & Larson, G. (2014). Heart rate monitoring using kinect and color amplification. In *2014 IEEE healthcare innovation conference (HIC)* (pp. 60–62). http://dx.doi.org/10.1109/HIC.2014.7038874.

Smith, S. W. (2003). Moving average filters. In *Digital signal processing: A practical guide for engineers and scientists* (pp. 277–284). Elsevier, http://dx.doi.org/10.1016/B978-0-7506-7444-7/50052-2.

Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, *3*(1), 42–55. http://dx.doi.org/10.1109/T-AFFC.2011.25.

Sun, Y., & Thakor, N. (2016). Photoplethysmography revisited: From contact to noncontact, from point to imaging. (2015/09/15). *IEEE Transactions on Biomedical Engineering*, *63*(3), 463–477. http://dx.doi.org/10.1109/TBME.2015.2476337.

Tanabe, J., Miller, D., Tregellas, J., Freedman, R., & G.Meyer, F. (2002). Comparison of detrending methods for optimal fMRI preprocessing. *NeuroImage*, *15*(4), 902–907. http://dx.doi.org/10.1006/nimg.2002.1053.

Tarassenko, L., Villarroel, M., Guazzi, A., Jorge, J., Clifton, D. A., & Pugh, C. (2014). Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiological Measurement*, *35*(5), 807–831. http://dx.doi.org/10.1088/0967-3334/35/5/807.

Tarvainen, M. P., Ranta-aho, P. O., & Karjalainen, P. A. (2002). An advanced detrending method with application to HRV analysis. *IEEE Transactions on Biomedical Engineering*, *49*(2), 172–175. http://dx.doi.org/10.1109/10.979357.

van Gastel, M., Stuijk, S., & de Haan, G. (2015). Motion robust remote-PPG in infrared. *IEEE Transactions on Biomedical Engineering*, *62*(5), 1425–1433. http://dx.doi.org/10.1109/TBME.2015.2390261.

Verkruysse, W., Svaasand, L. O., & Nelson, J. S. (2008). Remote plethysmographic imaging using ambient light. *Optical Express*, *16*(26), 21434–21445. http://dx.doi.org/10.1364/oe.16.021434.

Vila, J., Palacios, F., Presedo, J., Fernandez-Delgado, M., Felix, P., & Barro, S. (1997). Time-frequency analysis of heart-rate variability. *IEEE Engineering in Medicine and Biology Magazine*, *16*(5), 119–126. http://dx.doi.org/10.1109/51.620503.

Villarroel, M., Jorge, J., Pugh, C., & Tarassenko, L. (2017). Non-contact vital sign monitoring in the clinic. In *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)* (pp. 278–285). http://dx.doi.org/10.1109/FG.2017.43.

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *2001 IEEE computer society conference on computer vision and pattern recognition (CVPR)*. http://dx.doi.org/10.1109/CVPR.2001.990517.

Wagner, J., Kim, J., & Andre, E. (2005). From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In *2005 IEEE international conference on multimedia and expo* (pp. 940–943). http://dx.doi.org/10.1109/ICME.2005.1521579.

Wei, B., He, X., Zhang, C., & Wu, X. (2017). Non-contact, synchronous dynamic measurement of respiratory rate and heart rate based on dual sensitive regions. *Biomedical Engineering Online*, *16*(17), http://dx.doi.org/10.1186/s12938-016-0300-0.

Wiede, C., Richter, J., & Hirtz, G. (2016). Signal fusion based on intensity and motion variations for remote heart rate determination. In *2016 IEEE international conference on imaging systems and techniques (IST)* (pp. 526–531). http://dx.doi.org/10.1109/IST.2016.7738282.

Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., & Freeman, W. (2012). EulerIan video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics*, *31*(4), 65:1–65:8. http://dx.doi.org/10.1145/2185520.2185561.

Yang, L., Liu, M., Dong, L., Zhao, Y., & Liu, X. (2015). Motion-compensated non-contact detection of heart rate. *Optics Communications*, *357*, 161–168. http://dx.doi.org/10.1016/j.optcom.2015.08.017.

Zhang, C., Wu, X., Zhang, L., He, X., & Lv, Z. (2017). Simultaneous detection of blink and heart rate using multi-channel ica from smart phone videos. *Biomedical Signal Processing and Control*, *33*, 189–200. http://dx.doi.org/10.1016/j.bspc.2016.11.022.

Zhao, F., Li, M., Qian, Y., & Tsien, J. Z. (2013). Remote measurements of heart and respiration rates for telemedicine. *PLoS One*, *8*(10), Article e71384. http://dx.doi.org/10.1371/journal.pone.0071384.