Remote heart rate measurement using low-cost RGB face video: A technical literature review

Article *in* Frontiers of Computer Science (electronic) · September 2018

DOI: 10.1007/s11704-016-6243-6

5 authors, including:

Philipp V. Rouast
University of Newcastle

Marc T. P. Adam
University of Newcastle

Raymond Chiong
University of Newcastle

David Cornforth
University of Newcastle

Front. Comput. Sci. ACCEPTED VERSION
DOI 10.1007/s11704-016-6243-6
Philipp V. ROUAST, Marc T.P. ADAM, Raymond CHIONG, David CORNFORTH, Ewa LUX: Remote heart rate measurement using low-cost RGB face video: a technical literature review
REVIEW ARTICLE

Remote heart rate measurement using low-cost RGB face video: a technical literature review

Philipp V. ROUAST[1], Marc T.P. ADAM[2], Raymond CHIONG (✉) [2], David CORNFORTH[2], Ewa LUX[1]
1 Institute of Information Systems and Marketing, Karlsruhe Institute of Technology, 76133 Karlsruhe, Germany
2 School of Design, Communication and Information Technology, The University of Newcastle, Callaghan, NSW 2308, Australia

Abstract Remote photoplethysmography (rPPG) allows remote measurement of the heart rate using low-cost RGB imaging equipment. In this study, we review the development of the field of rPPG since its emergence in 2008. We also classify existing rPPG approaches and derive a framework that provides an overview of modular steps. Based on this framework, practitioners can use our classification to design algorithms for an rPPG approach that suits their specific needs. Researchers can use the reviewed and classified algorithms as a starting point to improve particular features of an rPPG algorithm.

Keywords Affective computing, Heart rate measurement, Remote, Non-contact, Camera-based, Photoplethysmography

1 Introduction

As a source of information about a subject's physical and affective state, heart rate measurement (HRM) is of interest to researchers, medical practitioners, and retail users alike. A classical application of HRM is for monitoring in a hospital environment. However, recently, access to HR data has been necessary for applications related to personal fitness [1], electronic commerce [2,3], financial trading [4], and corporate technostress [5].

A measured HR is derived from a volumetric measurement (plethysmogram) of the heart as the number of contractions per minute. Typically, HRM is conducted using methods that require skin contact. In the case of electrocardiograms (ECG), this contact is necessary to measure electrical changes on the skin. The type of photoplethysmogram (PPG) available on some smart watches uses skin contact to obtain a plethysmogram optically. Although they are noninvasive, these techniques are obtrusive in that they require contact with the human skin, which can be detrimental to subjects with sensitive skin (e.g., neonates). It can also be irritating (e.g., for subjects having to wear a fitness tracker) or distracting (e.g., when worn in a professional environment). In these example scenarios, using a less obtrusive, contactless means of measurement would be beneficial.

During the last decade, considerable research has been published on HRMs that do not require skin contact. The developed techniques use a color model

Received Month dd, yyyy; accepted August 23, 2016

E-mail: raymond.chiong@newcastle.edu.au

based on red, green, and blue (RGB) imaging to acquire a signal from a distance of up to several meters. These techniques are thus commonly referred to as remote photoplethysmography (rPPG) because of their similarity to traditional PPG. Research has shown that reliable HRM can be achieved using low-cost, consumer-grade digital cameras and ambient light sources. The proposed methods capture the subject's head on video (e.g., using a webcam), from which the plethysmographic signal is recovered using several image processing techniques and transformations.

Two main approaches have emerged from existing studies on rPPG: 1) HRM based on periodic variation of the subject's skin color, and 2) HRM based on periodic head movement. Both of those observable phenomena are caused by the human cardiac cycle and thus allow researchers to infer an HRM from an estimated plethysmographic signal.

Since rPPG was first proposed in 2008 [6], the focus has shifted from demonstrating feasibility in optimal, lab-like conditions to a variety of more complex algorithms for realistic scenarios. The existing review studies in this field, such as [7–9], provide a theoretical background and overview of the field. However, none of them focuses entirely on low-cost cameras nor provides a structured classification of existing approaches. Therefore, our contributions are as follows:

1) to provide an overview of research conducted in this field;
2) to present a technical account of the typical components of rPPG algorithms and identify the main challenges; and
3) to classify published studies by their choice of algorithm and contributions to the field.

Finally, we also provide suggestions for future research in the field of rPPG.

2 Research methodology

A clear consensus concerning the name of the field that we discuss in this study has yet to emerge. While researching, we came across 15 different terms used by different authors. These typically employ lexical combinations that begin with such words as "remote", "non-contact", "camera-based", "video-based", "contactless", "contact-free", "imaging" and end with such terms as "photoplethysmography", "heart rate measurement", "heart rate estimation", "heart rate monitoring", as well as various abbreviations thereof. We chose to use the term "remote photoplethysmography" (rPPG) because it is by far the most frequently used (ca. 50%) and is an original name [6] for this class of algorithms.

In the process of identifying a wide range of relevant published studies, we used previously listed terms to conduct searches in Google Scholar. To this search field we added studies that have cited the two seminal studies on rPPG [6,10] by reverse-searching citations.

Because we review studies on rPPG that used low-cost face video, we include only studies whose goal was to obtain HRM using videos of subject faces. Recording equipment must be of commercial grade. Therefore, those studies that used professional equipment such as high-speed cameras were excluded. As of this writing, we found 35 publications that match our criteria. Fig. 1 provides an overview of the publication count by year.

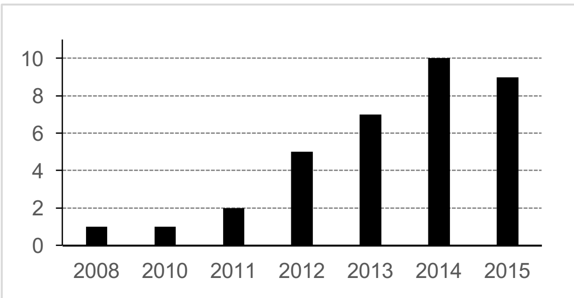

Fig. 1 Number of studies by year

3 Background

Phenomena exploited in rPPG are closely related to the cardiac cycle. During each cycle, blood is moved from the heart to the head through the carotid arteries. We will see that this periodic inflow of blood affects both the optical properties of facial skin and the mechanical movement of the head, enabling researchers to measure HR remotely.

The interplay of light and living tissue is complex, as many processes such as scattering, absorption, and reflection are at play. Research has shown that reflection of light is dependent on, among other factors, blood volume change and blood vessel wall movement [11,12]. Given suitable illumination, changes in light reflected from facial skin are thus observable, as the blood flow and variation of blood volume follow the cardiac cycle. Traditionally, dedicated light sources with red or near-infrared wavelengths [11] have been used to obtain a (contact) photoplethysmogram. However, recent research has shown that ambient light can be sufficient to obtain a plethysmographic signal [6] (as illustrated in Fig. 2a).

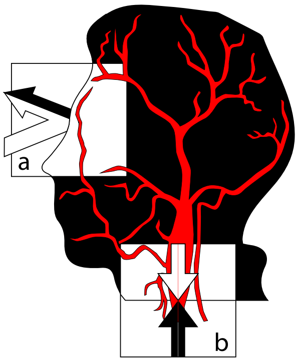

Fig. 2 Illustration of phenomena used in rPPG

More recently, some research has focused on remotely capturing the mechanical impact of blood flowing in through the carotid arteries at either side of the head [13]. The idea of exploiting the Newtonian reaction of the human body to the displacement of blood dates back to the 1930s [14]. This approach [13] considers the head-neck system and the trunk as a sequence of stacked inverted pendulums and surmises that the opposite reaction to blood inflow causes a displacement of the head by approximately 5 mm (illustrated in Fig. 2b). Of the two approaches, that based on skin color variation, being the original, has been discussed in many more studies.

4 Early work and recent development

Hertzman and Spealman first noted in 1937 that the variation in light transmission of a finger could be detected by a photoelectric cell [15]. The formative period of rPPG research began in 2008 with Verkruysse and colleagues first showing that video recordings of a subject's face under ambient light contain a signal sufficiently rich to measure the HR [6]. They asked volunteers to sit motionless while their faces were recorded using inexpensive consumer cameras from a distance of 1-2 m. Fig. 3 illustrates the typical setup of such studies.

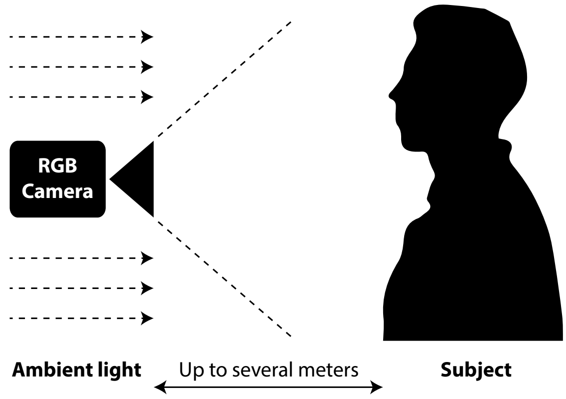

Fig. 3 Typical setup of an rPPG application

Verkruysse at al. used color recordings of different quality. For example, a resolution of 640 x 480, which is a standard graphic mode of the video graphics array (VGA), and a frame rate of 30 frames per second (fps) were used. In these recordings, the region of interest (ROI) was manually selected. From the pixels contained in the ROI, the raw signal was computed per frame as the mean value of each of the RGB color channels. To determine power spectral density of the signal, Verkruysse et al. used the fast Fourier transform (FFT) algorithm. They showed that the signal for the green channel contains the strongest plethysmographic signal, clearly indicating the fundamental HR frequency, up to its fourth harmonic. This is consistent with the fact that hemoglobin absorbs green light better than it does red and blue.

Paving the way for future research was the first study with the explicit goal of measuring HR using video recorded with a standard laptop webcam [10]. This study by Poh et al. used a face detector to track a subject's face frame by frame, with a box containing the subject's face as the ROI and a moving window of 30 s to achieve a continuous measurement. Improving on this approach in [6], all three channels of RGB information were used. Blind source separation (BSS) estimated the plethysmographic signal as a linear combination of all three raw signals. The parameters for this combination were estimated using independent component analysis (ICA). However, Poh et al. always chose the second component produced by ICA as the plethysmographic signal, a shortcoming they later addressed in an improved version of their algorithm [16]. The HR was estimated as the frequency with the highest response after an FFT.

With the general feasibility of rPPG being established, an increasing number of publications have been produced in subsequent years, as shown in Fig. 1[1)]. Initial contributions include comparing alternative methods for BSS and different selections regarding color channels [17,18], as well as adding temporal filters before BSS is performed [16,17]. An approach using an ROI and neural-network-based skin detection proposed by [19] allows for more accurate measurement. Another study [20] compared various linear and nonlinear techniques for BSS and found that Laplacian eigenmap produces the best results.

The plethysmographic signal in a subject's face can be visualized by decomposing the video sequence into different spatial frequency bands and then magnifying a desired frequency band using bandpass filtering [21]. When this process is applied to facial videos, slight temporal changes are detectable. This shows that HR and individual heart beats can be extracted from the amplified signal [22].

A fundamentally different approach used to obtain a raw signal was presented by Balakrishnan et al. in [13]. Instead of relying on color change, this study demonstrated the possibility of extracting a plethysmographic signal from the periodic motion of the subject's head, which occurs because of the influx of blood to the head. Balakrishnan et al. tracked an array of feature points in the subject's face frame by frame, recording the longitudinal trajectories. After performing temporal filtering to remove unwanted frequencies, they used BSS to obtain a sufficiently strong plethysmographic signal to estimate the HR. One weakness of this approach is the fact of signal loss during bigger motions. Two additional studies explored this approach. One [23] showed that a single tracking point can provide sufficient information for HRM. The other [24] achieved an improved performance by replacing the FFT with discrete cosine transform (DCT) in the estimation step.

[1)] The lower number of publications in 2015 may be attributed to publication and indexing lag.

Until then, the research on rPPG remained in an early stage. Although accurate measurements were shown to be possible using two signal sources, this was accomplished under mostly controlled conditions using stationary subjects. In more recent research, the focus has been on more realistic settings containing naturally moving or exercising subjects and more challenging illumination. The two recently explored problems are reducing noise from subject motion and addressing low signal strength (e.g., resulting from illumination and dark skin tone).

A group at Philips Research [25] addressed the problem of moving subjects with respect to the light source. They argued that an optimal fixed combination of bandpassed RGB channel signals can be found based on a ratio of normalized color signals when assuming "standardized" skin, thus eliminating noise derived from specular reflection. A deficiency of this approach was that it excluded BSS from the algorithm's design. The researchers then further formalized and improved their approach [26] by proposing a combination with BSS techniques.

As other researchers [27] have found, the choice of ROI has a major influence on the quality of the plethysmographic signal, as not all areas in the face exhibit the same signal quality. The most recent studies have focused on more intelligent ROI selection and tracking to achieve motion robustness. Detection of facial landmark points is typically the basis for a more detailed ROI (e.g., to define [28–32] and track [29,31] custom ROIs). An approach by Feng et al. [33] found an array of points in the subject's face that can be subsequently tracked in order to update the ROI on the subject's forehead. Consistent with the findings of [27], Feng et al. later improved their algorithm to use the area of the cheeks [34]. Further reductions in noise were made possible by the so-called adaptive bandpass filter adopted by some authors [32–35], the cut-off frequencies for which were based on past HR estimates. Custom additional filtering steps introduced by some authors also aimed at reducing noise (e.g., [29] used an adaptive filter to reduce noise from illumination changes using background illumination as a reference).

Further recent developments include variations in the number of used raw signals, such as the inclusion of cyan and orange frequencies [30,36]. In a different approach [31], the facial region was divided into many small ROIs that yielded an array of signals from the green channel, each of which was later combined using a weighted average based on a goodness metric. Similarly, the researchers in [37] stochastically selected an array of points and combined them using an importance-weighted Monte Carlo approach. The use of BSS, followed by component selection, have recently been optimized using machine learning techniques [38,39].

Despite the fact that these recent improvements allow rPPG algorithms to be applied to more realistic situations, virtually all studies have continued to focus on proof of concept using pre-recorded videos. Although one study [40] presented concepts for real-time applications, only one other work reported data from a real-time rPPG application [41]. Signal-to-noise ratios and error rates have typically been reported, but comparing different approaches is difficult. This is because most authors have tended to create their own test scenarios using a variety of cameras and often have not specified the algorithms used for compression, thus making reproduction difficult. An exception to this is a study that benchmarked rPPG algorithms using videos from a publicly available database [29]. However, no consistent practice has yet been adopted.

5 rPPG algorithm classification

We give a general classification of existing rPPG approaches based on the type of signal (color or motion). We then propose a general algorithm framework (see Fig. 4) and classify the chosen approaches accordingly. An overview of the corresponding classifications is given in Table 1.

This framework is based on the biological measuring chain [42]. We subdivide a typical rPPG algorithm into three key steps: (i) extraction of the raw signal from several video frames, (ii) estimation of the plethysmographic signal, and (iii) HR estimation. Each of these steps has several components that may be subject to various approaches or can be skipped as in existing studies.

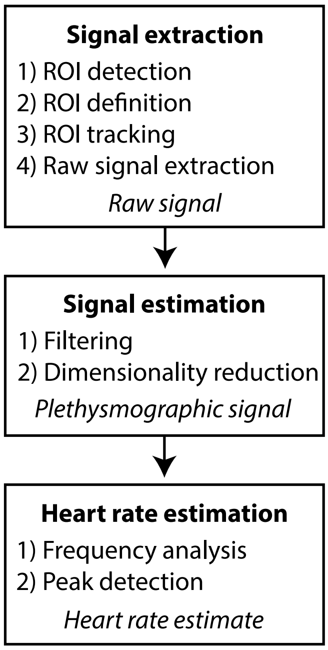

Fig. 4 Generalized rPPG algorithm framework

The majority of studies (91%) used facial color variation as the raw signal for rPPG. This periodic color variation occurs as the skin's light absorption changes in accordance with the cardiac cycle. Through the use of an RGB camera, these slight color variations can be remotely registered. The remaining 9% of studies were based on periodic head movements, which can likewise be monitored using remote imaging. These head movements represent an equal and opposite reaction to blood being pumped to the head through the aorta with each cardiac cycle. In the following subsection, we describe how we use our proposed general framework to classify all studies, while highlighting those methods that are based on color variation.

5.1 Signal extraction

ROI detection. Because the rPPG algorithms we consider are based on the human face, ROI detection is necessary to determine the bounds of the face in a video frame. This information is typically an intermediate step from which a more accurate ROI is later defined. In some of the literature, especially in earlier studies in which the subject was asked to sit motionless (e.g., [6,13,17]), the bounds of the face were selected manually from one of the first frames.

The most frequently used method is the algorithm of

Viola and Jones [43], which is a machine-learning-based approach that uses a cascade of simple features to classify faces. The popularity of this approach is partially due to its availability in the OpenCV computer vision library, which many authors have used to implement their rPPG algorithms. A bounding box of the face is returned when using the Viola-Jones algorithm.

As an alternative to face detection, Lee et al. [19] proposed using an algorithm to detect skin regions. Skin-like pixels were selected using a neural-network-based classifier. Additional areas such as the neck and arms may be included in the ROI in this manner. The drawback of using only skin detection is the presence of additional noise when objects are present that have color similar to the skin, as Lee et al. themselves acknowledged. Therefore, some studies (e.g., [25]) used skin selection within the bounds of the face given by the Viola-Jones algorithm.

Recent approaches dealing with subject motion require more detailed information about face location. Facial landmark points provide the basis for more detailed ROI definitions as well as ROI tracking. Using active appearance models (AAM) [44], which is a statistical model of the human face in which appearance is matched to the given video frame, results in a set of coordinates of known facial landmarks. After face detection using the Viola-Jones algorithm, [28] and [32] included this step in their rPPG algorithm. Three other algorithms for facial landmark detection have been used: [29] applied discriminative response map fitting (DRMF) [45] after face detection; [31] used an algorithm for deformable model fitting; and [46] and [30] used an algorithm that combines a regression-based approach with a probabilistic face-shape model [47]. The last two studies directly used facial landmark detection without prior face detection.

ROI definition. The ROI is the area within a video frame that contains pixels providing the raw signal for the algorithm. Utilizing information from Viola-Jones or manual face detection, researchers have the option of simply using the given bounding box of the face as the ROI [6,17,18,39,48,49]. Some authors also selected an experimentally determined fixed subset of the bounding box. As the bounding box from the Viola-Jones algorithm typically includes background pixels on either side, a common method is to include 60% of its width [10,16,20,38]. Other studies [13,24,50] used different experimentally obtained subsets of the bounding box as the ROI. Two notable subsets of the bounding box that may be determined solely from the bounding box or additionally by the coordinates of the eyes are the forehead [6,17,33] and cheeks [34], having been identified as promising regions [27].

Researchers working with facial landmark points used these to define more exact and robust ROIs. Using nine landmark points, [29] defined a region that includes the cheeks and no background pixels, similar to [30] and [36], which defined a region that includes the forehead and the area below the eyes.

Another recent approach involves defining multiple ROIs and generating one RGB signal each for subsequent analysis. For example, [28] and [32] used landmark points to define several ROIs representing regions of the face. The approaches of [27,31,35,37,51] are more rigorous, each using a large array of small ROIs. These studies select a subset of available ROIs using a criterion of signal quality, thus yielding a dynamic ROI.

ROI tracking. Noise caused by subject motion may render the signal useless for rPPG. Thus, the goal of ROI tracking is to ensure that the pixels contained in the ROI belong to a skin region invariant to subject motion. Some earlier studies that assumed the subject was stationary did not use tracking, particularly when manual ROI detection was involved [6,17].

A straightforward method to achieve ROI tracking is to simply re-detect the ROI for every frame. Two-thirds of the authors achieved ROI tracking using this method (Table 1). However, drawbacks exist with this method. Because the bounding box returned by the frequently used Viola-Jones object detector is not very exact, ROIs based on its fluctuating output may in turn cause unwanted noise. Considering computational complexity, it is obviously suboptimal to re-run ROI detection for every frame, especially if real-time applications are intended.

Through use of a set of tracking points or objects and a tracking algorithm, the location of the ROI can be updated frame by frame without having to re-detect the ROI. The good-features-to-track algorithm [52], used by the authors of [29] and [31] as tracking points, returns the most prominent corners within the ROI. Using the Kanade-Lucas-Tomasi (KLT) feature tracker based on [53], the authors estimated an affine transform to update the ROI based on subject motion. Similarly, [33] and [34] used the KLT tracking algorithm based on the points identified using the speeded-up robust features (SURF) [54] algorithm. The authors in [19] used kernel-based object tracking [55] to update the location of the skin regions included in their ROI. For each ROI corresponding to a single pixel, [35] used tracking-by-detection with kernels (CSK) [56] to compensate for rigid motion and an optical flow algorithm proposed by Farnebäck [57] to compensate for non-rigid motion.

Raw signal extraction. The raw signal is extracted from a video frame by frame according to the ROI position. For color-based methods, this yields series $I_i(t)$ for the color channels $i \in \{R, G, B\}$. Values are calculated by averaging the respective color channel of all pixels contained in the ROI of the frame at time t. This is known as spatial pooling and has the purpose of averaging out camera noise contained in single pixels. The number of ROIs and selection of channels for which this step is performed vary across studies. In the case of very small ROIs, the image can be downsampled to avoid noise [35]. To visualize temporal changes, [21] first decomposed an image into different spatial frequency bands without explicitly extracting single values per frame. They referred to this approach as localized spatial pooling.

Extraction of the raw signal for methods based on head motion requires selecting tracking points within the ROI. All three author teams that worked with this type of method used the good-features-to-track algorithm [52]. Although [13] and [24] used an array of tracking points, [23] used only the best identified point. Using the KLT tracking algorithm, the authors computed the trajectory of each point. The raw signal then consisted of series $T_{i,a}(t)$ for tracking point i and axis a. Whereas [13] and [24] used just the vertical axis, [23] used both the vertical and horizontal axes.

Table 1 gives the number of series and the signal (in brackets) to which they correspond (e.g., 3 (RGB) for the three channels of red, green, and blue). The table also gives the number of ROIs (e.g., n x 1 (y) denoting n tracking points for the y axis).

5.2 Signal estimation

Filtering. Despite ROI tracking, the raw signal may still contain unwanted noise, which depends on subject motion, illumination changes, and other factors. Using information about the frequencies of these expected noise sources and the range of feasible HR frequencies, researchers typically apply one or more digital filters to the raw signal. The goal is to increase the signal-to–noise ratio and thus improve the quality of the estimated plethysmographic signal. Given a raw signal consisting of multiple series, the filters are normally applied to each series before dimensionality reduction. However, some authors apply filters after dimensionality reduction, whereas some do so both before and after. In Table 1, "(1)" indicates filtering before and "(2)" filtering after dimensionality reduction.

Because the level of raw signals (e.g., color space value or pixel trajectory) has no meaning when assessing periodicity, a common first step is to centralize or normalize the raw signals. Centralizing is a process in which the mean μ_S is subtracted from a signal S. Normalization additionally divides the signal by its standard deviation σ_S.

Both unwanted high and low frequency noise can be eliminated using bandpass filtering. This requires an assumption regarding the band of frequencies that is feasible for human HR. A common choice of band is [0.7 Hz, 4 Hz] [10,16,29,34], which corresponds to an HR between 42 and 240 beats per minute (bpm).

Additional methods that remove unwanted high and low frequency noise include the moving average filter and the detrending method. The moving average filter is a rolling window that averages a given number of values, thus representing a low-pass equivalent. The detrending method [58] is based on a smoothness priors approach and represents a simple and efficient means of removing the long-running trend from a signal. It can be seen as a high-pass equivalent. In Fig. 5, we give a simple example of the removal of low- and high-frequency noise from the green channel obtained from a subject's forehead.

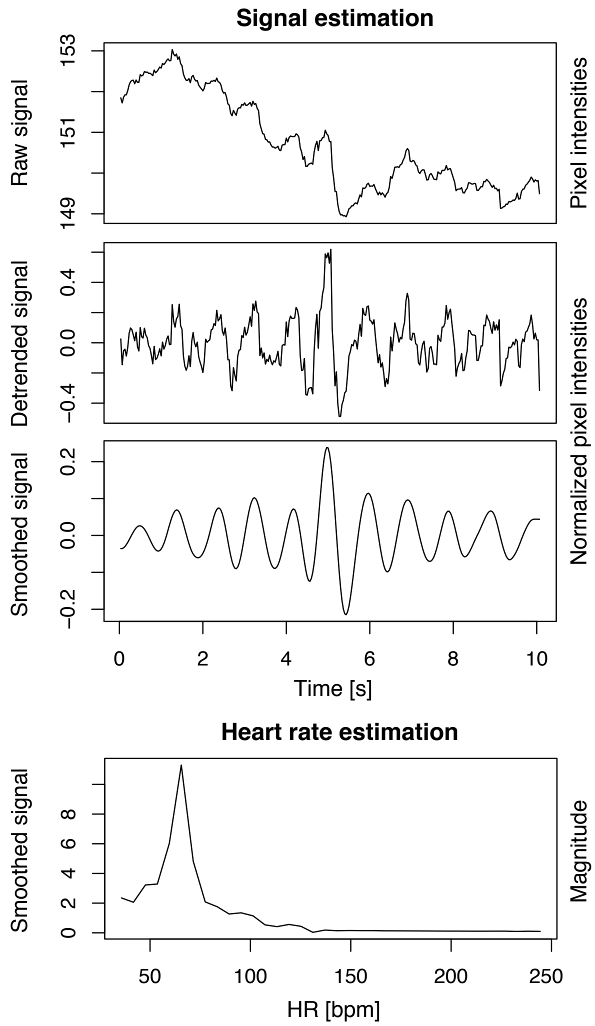

Fig. 5 Exemplary values from a simple rPPG application using only the green channel

A novel component in recent publications is an adaptive bandpass [33–35] that dynamically changes the cutoff frequencies based on previously estimated HR, thus guiding the algorithm to produce consistent HR estimates. Some authors have experimented with additional noise reduction by eliminating outliers in the signals. For example, [29] eliminated the noisiest segments in a considered signal as measured by standard deviation. Similar approaches were followed by Wei et al. [20], who eliminated outliers in the signal, and Wang et al. [35], who pruned spatially by excluding non-skin pixels and outliers with respect to the color space. To address noise caused by illumination changes (e.g., playing a movie on a screen), [29] used the background illumination as reference and applied an adaptive filter to remove illumination noise from the signal.

Dimensionality reduction. Most authors used a raw signal that consists of more than one single series (e.g., signals corresponding to the RGB channels). It is assumed that the raw signals contain a one dimensional plethysmographic signal $p(t)$, which can be represented as a linear combination of these raw signals using a weighted sum. Estimating the weights for this combination has proven difficult and is one of the most debated issues in the literature on rPPG.

The first approach proposed by Poh et al. [10] used an algorithm for BSS to determine the optimal combination of raw signals. They chose ICA, which separates the raw signals into independent, non-Gaussian signals. In their original rPPG algorithm, Poh et al. determined that the second component produced by ICA is typically the most periodic one, which seems to correspond to the plethysmographic signal $p(t)$. Several other authors adopted this method [18,28,48]. Theoretically, however, the order that ICA components appear in is random, which is why Poh et al. later introduced a selection criterion in the improved version of their rPPG algorithm [16]. This criterion chooses the component with the highest peak in the frequency power spectrum (i.e., a component with a high periodicity [16,30,36,50]). Another related criterion chooses the highest periodicity according to the percentage of spectral power accounted for by the first harmonic [13,23,24]. The authors in [33] used correlation with the reference sine function to determine the best component. A second popular algorithm for BSS, first used by [17] and later by [13,24,35], is the principal component analysis (PCA), which separates raw signals into linearly uncorrelated components and orders them based on variance. Criteria used for component selection in ICA can be equally applied to components produced by PCA. Machine learning was also used by [38] to select the most appropriate component produced by ICA.

Table 1 Classification of published rPPG algorithms

	Signal extraction						Signal estimation		Heart rate estimation	Comment
Paper (Year)	**Signal type**	**ROI detect.**	**ROI definit.**	**ROI track.**	**Raw signal extr.**	**Raw signal dim.**	**Filtering**	**Dim. red.**		**Contribution/deficiencies**
[6] (2008)	Color	Manual	BB, FH		Spatial pooling	3 (RGB)	Centralize, Bandpass			First showing feasibility of rPPG/Data processed manually
[10] (2010)	Color	VJ	SBB	RE	Spatial pooling	3 (RGB)	(1) Normalize	ICA	FFT	Using face detection, BSS with ICA, HR estimation/Fixed component selection after ICA
[17] (2011)	Color	Manual	BB, FH		Spatial pooling	2 (RG/ GB/RB)	(1) Bandpass	ICA, PCA	FFT	Comparing ROI types and BSS techniques/No automatic process
[16] (2011)	Color	VJ	SBB		Spatial pooling	3 (RGB)	(1) Detrend+Normalize (2) MA + Bandpass	ICA	Peak detection	Improvement over [10] with temporal filtering and intelligent component selection
[18] (2012)	Color	VJ	BB	RE	Spatial pooling	3 (RGB), 1 (G)	(1) Normalize	ICA	FFT	Feasibility using smartphones as video source and for computation/No real-time measurement in mobile apps
[21] (2012)	Color				LSP		Bandpass			Visualize and amplify small temporal changes/Only visualization of the pulse
[22] (2012)	Color		Many ROIs		LSP	n x 3 (RGB)	(1) Bandpass	Importance metric	FFT, Peak detection	Use signal amplification proposed in [21] to measure the HR
[19] (2012)	Color	SK	Skin regions	KBOT	Spatial pooling	3 (RGB)		Fixed linear	FFT	Propose skin detection and tracking/Possibly additional noise from areas similar to skin
[20] (2013)	Color	VJ	SBB	RE	Spatial pooling	3 (RGB)	(2) Eliminate outliers + MA + Bandpass	Laplacian Eigenmap	Peak detection	Comparing BSS methods/No filter before BSS
[13] (2013)	Motion	VJ, Manual	SBB		GFTT + KLT	n x 1 (y)	(1) Bandpass	PCA	FFT, Peak detection	First proposing an approach based on head motion/Prone to noise from larger motions
[23] (2013)	Motion	Manual			GFTT + KLT	2 (x, y)	(1) Normalize + Bandpass	ICA	FFT	Based on horizontal and vertical trajectory of one pt./Prone to noise from larger motions
[25]	Color	VJ +	Skin	RE	Spatial	3 (RGB)	(1) Normalize +	Fixed	FFT	Fixed signal combination based on normalized

(2013)		SK	regions		pooling		Bandpass	linear		skin/Not taking advantage of BSS
[50] (2013)	Color	Face detect.	SBB	RE	Spatial pooling	3 (RGB)		ICA	STFT	Using the STFT for HR estimation/No filtering
[28] (2013)	Color	VJ + AAM	10 ROIs	RE	Spatial pooling	3 (RGB)	(1) Detrending + Normalize	ICA	FFT	Integration of AAM
[48] (2013)	Color	Manual	BB	RE	Spatial pooling	3 (RGB)	(1) Lowpass + Normalize + Detrending + MA	ICA, PCA	FFT	Compare BSS techniques, find ICA to be most consistent/Manual face detection
[27] (2013)	Color		Many ROIs		Spatial pooling	n x 1 (G)			FFT	Determining the optimal ROI selection/No actual HRM conducted
[24] (2014)	Motion	VJ	SBB		GFTT + KLT	n x 1 (y)	(1) Moving average + Bandpass	PCA	DCT	Using DCT instead of FFT for HR estimation/Prone to noise from larger motions
[29] (2014)	Color	VJ + DRMF	FLM based	GFTT + KLT	Spatial pooling	1 (G)	IR + NRME + Detrending + MA + Bandpass		FFT	Robustness against motion and illumination changes/Only using the green channel
[38] (2014)	Color	VJ	SBB	RE	Spatial pooling	3 (RGB)	(1) Detrending + Normalize	ICA	FFT + ML	Using different machine learning methods to extract HR from features
[39] (2014)	Color	Face detec.	BB	RE	Spatial pooling	3 (RGB)	(1) Bandpass	ICA + fixed linear + SVR	FFT + SVR	Using SVR to extract the HR from frequency domain features/No detailed ROI
[33] (2014)	Color	VJ	FH	SURF + KLT	Spatial pooling	3 (RGB)	(1) Adaptive bandpass	ICA	FFT	Motion compensation using tracking and adaptive bandpass
[32] (2014)	Color	VJ + AAM	FLM based	RE	Spatial pooling	1 (G)	Outlier removal + Centralize + Detrending + Lowpass		FFT, Peak detection	Using AAM and custom filtering/Relies on a commercial facial analysis framework, only using the green channel
[49] (2014)	Color	manual	BB		Spatial pooling	2 (RG)	(2) Bandpass	Fixed linear	FFT	Using fixed signal combination instead of BSS/Manual face detection, no sliding window
[26] (2014)	Color	VJ + SK	Skin regions	RE	Spatial pooling	3 (RGB)	(1) Normalize + Bandpass	ICA, PCA + fixed lin.	FFT	Improvement over [25] by combining the fixed dimensionality reduction approach with BSS methods

[30] (2014)	Color	FLM	FLM based	RE	Spatial pooling	5 RGBCO	(1) Detrend + Normalize (2) Bandpass	ICA	Peak detection	Extract BVP waveform and systolic and diastolic peaks
[34] (2015)	Color	VJ	Cheeks	SURF + KLT	Spatial pooling	2 (RG)	(1) Bandpass (2) Adaptive bandpass	Adaptive GRD	FFT	Improvement over [33] using the cheeks and adaptive red-green difference
[31] (2015)	Color	FLM	Many ROIs	GFTT + KLT	Spatial pooling	n x 1 (G)	(1) Bandpass	Goodness metric	FFT, peak detection	Increase robustness by tracking an array of small ROIs
[35] (2015)	Color	Manual + FLM	Many ROIs	CSK + Farneb äck	Spatial pooling	n x 3 (RGB)	(1) Spatial pruning + Exclude least periodic + Adaptive bandpass	PCA	FFT	Improve the signal-to-noise ratio by exploiting spatial redundancy of the image sensor
[36] (2014)	Color	FLM	FLM based	RE	Spatial pooling	5 RGBCO	(1) Detrend, Normalize (2) Bandpass	ICA	Peak detection	Show that a five band camera leads to better performance
[40] (2015)	Color	VJ	Skin regions	KLT	Spatial pooling	3 (RGB)	(1) Normalize (2) Bandpass	LDA	FFT	Propose a real-time approach, use of LDA
[37] (2015)	Color		Many ROIs	RE		2 (RG)	Erythema transform	Bayesian minim.	FFT	Stochastically selected points and Bayesian estimation
[41] (2015)	Color	VJ	Nose	KLT	Spatial pooling	1 (G)	Bandpass, Kalman filter		Peak detection	Real-time application/Only using the green channel
[51] (2015)	Color	VJ	Many ROIs	Dynam ic	Spatial pooling	n x 1 (G)	Bandpass	Overlap add	FFT	Dynamic ROI automatically adjusting to signal quality/Only using the green channel
[59] (2015)	Color	[60]	BB	[60]	Spatial pooling	3 (RGB)	(1) Normalize (2) Bandpass + MA	ICA	FFT	Assess optimal camera distance from subject/Non-automated ICA component selection
[61] (2015)	Color	VJ	FH	[62]	Spatial pooling	3 (RGB)	(1) MA + Normalize (2) Bandpass	ICA	Peak detection	Combine HRM with other physiological information

Note: VJ: Viola-Jones algorithm; SK: skin detection; AAM: active appearance model; DRMF: discriminative response map fitting; FLM: facial landmark detection; BB: bounding box; SBB: subset of bounding box; FH: forehead; RE: re-detection; KBOT: kernel-based object tracking; GFTT: good-features-to-track; KLT: Kanade-Lucas-Tomasi tracking algorithm; SURF: speeded-up robust features; CSK: tracking-by-detection with kernels; IR: illumination rectification; MA: moving average; NRME: non-rigid motion elimination; LSP: localized spatial pooling; GRD: green-red difference; STFT: short time Fourier transform; ML: machine learning; SVM: support vector machine; SURF: speeded-up robust features

Similarly, [39] used support vector regression (SVR) to extract the plethysmographic signal $p(t)$ from a set of features in the frequency domain. Linear discriminant analysis (LDA) was used by [40] to reduce dimensionality. They constructed class values from the red channel and built the data from the other two channels.

In contrast to the algorithmically determined choice of weights through BSS, using fixed weights has been proposed by several authors. Although [19] determined fixed weights using a brute force technique, others derived weights from models of skin illumination. Under the assumptions of a standardized skin color, [25] proposed a theoretically motion robust method that uses all three RGB color channels to build two orthogonal color difference signals. These are then combined to yield the estimate of $p(t)$. The authors in [25] later acknowledged several limitations of their method and proposed combining it with BSS to support component selection [26]. Similarly, [34] derived an adaptive green-red difference (GRD) from a model of the skin and its relationship to the plethysmographic signal. This GRD is their estimate of $p(t)$.

A model of light interaction with the human skin, which involves a temporal quotient of raw signal values, was used by [49] to derive a different estimator for the plethysmographic signal. In [37], raw color series from single pixels were transformed using a custom erythema transform and used to estimate PPG by Bayesian estimation. Performance of nonlinear BSS techniques were compared by [20]. They found that Laplacian eigenmap performed best based on their data.

5.3 Heart rate estimation

Frequency analysis. Given an estimate $\hat{p}(t)$ of the plethysmographic signal, the HR frequency can be estimated using frequency analysis. For this purpose, this signal, which contains a distinct periodicity, is converted to the frequency domain using a discrete Fourier transform. The preferred algorithm by most authors (Table 1) is the FFT. Exceptions are [24], which used the DCT; [29], which used Welch's method for density estimation; and [50], which used the short-time Fourier transform (STFT). In the frequency domain, the frequency corresponding to the index with the highest spectral power is chosen as an estimate for the HR frequency. The intuition for this step is given in Fig. 5.

Peak detection. Using individual peaks, extracting more information such as HR variability from the inter-beat intervals is possible. To refine the signal for peak detection, the signal is usually interpolated using a cubic spline function [16,30,36]. The peaks can then be easily identified using a moving window, as they are the maxima within the signal.

6 Applications

Many promising application areas of rPPG algorithms (such as medicine and personal fitness) are frequently referred to in the reviewed literature. To date, existing applications range from simple experiments to assessing algorithm accuracy in controlled conditions. Researchers typically collect their own data (i.e., a video recording of the subject's face and corresponding ground truth measurement) using an established HRM method such as PPG or ECG. Thus, virtually all published work on rPPG is based on offline computations. An online application of an rPPG algorithm in an economic scenario was used in a lab experiment in [63].

Most studies employed between 10 and 20 subjects, optimally of both genders, various ages, and skin colors. Subjects typically sat at a desk and were given instructions to remain motionless (e.g., [10,16,17,28,48]) or move in a natural manner (e.g., [20,29,38,39]) while performing a task on a computer. Recent studies (e.g., [25,32,38,50,64]) also tested accuracy using exercising subjects. The camera was mounted on a location 0.5–3 m from the subject, who was illuminated with a mix of ambient light. Cameras used in the experiments varied from built-in and external webcams, smartphones, and point-and-shoot cameras to digital single-lens reflex cameras and mirrorless models. Recorded at between 10 and 30 fps, videos were saved in compressed or uncompressed form in various resolutions (from VGA to 720p resolution). Experiment durations varied from 20 s to over 10 min.

Table 2 Algorithm applications and reported accuracies

Paper	Subjects	Baseline method	Motion	RMSE [bpm]
[10]	12	PPG	S	2.29
			N	4.63
[16]	12	PPG	S	1.24
[22]	11	ECG	S	3.92
[25]	117	PPG	S	0.40
[50]	1	PPG	S	2.19
			E	2.26
[28]	6	ECG	S	1.47
[48]	18	PPG	S	7.73
[29]	10	ECG	S	1.27
[38]	10	ECG	N	3.64
			E	4.33
[39]	4	ECG	N	7.28
[65]	15	ECG	N	3.10
[40]	10	PPG	S	1.53
			N	5.72
[66]	10	PPG	N	0.11

Note: S: still; N: natural movement; E: exercising

As argued in Section 4, because of the lack of a widely used database of face videos and corresponding ground truth data, a comparison of reported results is rather difficult, especially given the multitude of parameters, with different choices potentially biasing the experimental results. Nevertheless, Table 2 provides a summary of application data from reviewed studies in which the number of subjects, baseline methods, motion instructions, and root mean square errors (RMSE) have been reported. If no standard case or average is given, we report the average of reported values. Again, these should be interpreted with care as a direct comparison between studies is not possible. As expected, when considering studies that tested different motion scenarios, errors increase when subject activities increase. Given a normal HR of 70 bpm, we can see that the reported RMSE is significantly less than 10% in most cases. On average, the reported errors are higher than those reported for some contact PPG [67]. However, the recent progress, especially considering motion and illumination robustness, is encouraging.

7 Conclusion and research implications

In this study, we addressed the growing phenomenon of rPPG. We provided a systematic literature review describing the research conducted in this field up to the present. We discussed the seminal work that the current rPPG literature is largely based on and gave an overview of the field's development over the last decade. We showed that a body of literature has increased over time as research has progressed and interest grown in rPPG for HRM. We described the more recent advances in rPPG based on skin color and head movement. We also included a technical description of the different physical attributes and software algorithms for rPPG. For the first time, published literature was tabulated based on choice of algorithm and contribution to the field. We also identified the main challenges in this field of research and provided suggestions on future research.

The two main challenges currently investigated in rPPG were identified as: increasing algorithm robustness with respect to subject noise and addressing low signal strength due to illumination and skin types. Many of these challenges have been effectively addressed, but the explored use cases are mostly remote from realistic real-world scenarios. Specifically, future rPPG algorithms must focus on a trade-off between the amount of processed information and algorithm complexity, because real-time applications will place a constraint on computation time.

Clearly, conducting remote HRM using low cost video equipment is possible, and previous studies show the increasing sophistication of rPPG. rPPG has a wide range of applications and the rise in publications over the period of our survey indicates an increasing interest in reliable rPPG algorithms. This is attributable to the demand for contactless HRM solutions in the medical, professional, and consumer sectors.

Acknowledgements This work was supported by a fellowship within the FITweltweit programme of the German Academic Exchange Service (DAAD).

References

1. Zhang Z, Pi Z, Liu B. Troika: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. IEEE Transactions on Biomedical Engineering, 2015, 62(2): 522–531
2. Adam MTP, Krämer J, Weinhardt C. Excitement up! Price down! Measuring emotions in dutch auctions. International Journal of Electronic Commerce, 2012, 13(2): 7–39
3. Adam MTP, Krämer J, Müller MB. Auction fever! How time pressure and social competition affect bidders' arousal and bids in retail auctions. Journal of Retailing, 2015, 91(3): 468–485
4. Astor PJ, Adam MTP, Jerčić P, Schaaff K, Weinhardt C. Integrating biosignals into information systems: A neurois tool for improving emotion regulation. Journal of Management Information Systems, 2013, 30(3): 247–278
5. Riedl R. On the biology of technostress: Literature review and research agenda. ACM SIGMIS Database, 2013, 44(1): 18–55
6. Verkruysse W, Svaasand LO, Nelson JS. Remote plethysmographic imaging using ambient light. Optics Express, 2008, 16(26): 21434–21445
7. McDuff DJ, Estepp JR, Piasecki AM, Blackford EB. A survey of remote optical photoplethysmographic imaging methods. In: Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2015, 6398–6404
8. Liu H, Wang Y, Wang L. A review of non-contact, low-cost physiological information measurement based on photoplethysmographic imaging. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. 2012, 2088–2091
9. Kranjec J, Begus S, Gersak G, Drnovsek J. Non-contact heart rate and heart rate variability measurements: A review. Biomedical Signal Processing and Control, 2014, 13(1): 102–112
10. Poh M-Z, McDuff DJ, Picard RW. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Optics Express, 2010, 18(10): 10762–10774
11. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiological measurement, 2007, 28(3): R1–R39
12. Lindberg L-G, Öberg PA. Optical properties of blood in motion. Optical Engineering, 1993, 32(2): 253–257
13. Balakrishnan G, Durand F, Guttag J. Detecting pulse from head motions in video. In: Proceedings of the 2013 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2013, 3430–3437
14. Starr I, Rawson AJ, Schroeder HA, Joseph NR. Studies on the estimation of cardiac ouptut in man, and of abnormalities in cardiac function, from the heart's recoil and the blood's impacts; the ballistocardiogram. American Journal of Physiology, 1939, 127(1939): 1–28
15. Hertzman AB, Spealman CR. Observations on the finger volume pulse recorded photoelectrically. American Journal of Physiology, 1937, 119(2): 334–335
16. Poh M-Z, McDuff DJ, Picard RW. Advancements in noncontact, multiparameter physiological measurements using a webcam. IEEE Transactions on Biomedical Engineering, 2011, 58(1): 1–4
17. Lewandowska M, Ruminski J, Kocejko T. Measuring pulse rate with a webcam - A non-contact method for evaluating cardiac activity. In: Proceedings of the 2011 Federated Conference on Computer Science and Information Systems (FedCSIS). 2011, 405–410
18. Kwon S, Kim H, Park KS. Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. In: Proceedings of the 2012 IEEE Annual International Conference of the Engineering in Medicine and Biology Society. 2012, 2174–2177
19. Lee K-Z, Hung P-C, Tsai L-W. Contact-free heart rate measurement using a camera. In: Proceedings of the 2012 Ninth Conference on Computer and Robot Vision (CRV). 2012, 147–152
20. Wei L, Tian Y, Wang Y, Ebrahimi T. Automatic webcam-based human heart rate measurements using laplacian eigenmap. In: Lecture Notes in Computer Science. 2013, 281–292
21. Wu H-Y, Rubinstein M, Shih E, Guttag J V,

Durand F, Freeman WT. Eulerian video magnification for revealing subtle changes in the world. ACM Transactions on Graphics, 2012, 31(4): 1–8
22. Wu H-Y. Eulerian video processing and medical applications. Master's thesis. Massachusetts Institute of Technology, 2012
23. Shan L, Yu M. Video-based heart rate measurement using head motion tracking and ica. In: Proceedings of the 2013 6th International Congress on Image and Signal Processing. 2013, 160–164
24. Irani R, Nasrollahi K, Moeslund TB. Improved pulse detection from head motions using dct. In: Proceedings of the 9th International Conference on Computer Vision Theory and Applications. 2014, 118–124
25. De Haan G, Jeanne V. Robust pulse rate from chrominance-based rPPG. IEEE Transactions on Biomedical Engineering, 2013, 60(10): 2878–2886
26. De Haan G, Van Leest A. Improved motion robustness of remote-PPG by using the blood volume pulse signature. Physiological measurement, 2014, 35(9): 1913–1926
27. Lempe G, Zaunseder S, Wirthgen T, Zipser S, Malberg H. Roi selection for remote photoplethysmography. In: Meinzer H-P, Deserno MT, Handels H, Tolxdorff T, eds. Informatik aktuell. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, 99–103
28. Datcu D, Cidota M, Lukosch S, Rothkrantz L. Noncontact automatic heart rate analysis in visible spectrum by specific face regions. In: Proceedings of the 14th International Conference on Computer Systems and Technologies. 2013, 120–127
29. Li X, Chen J, Zhao G, Pietikäinen M. Remote heart rate measurement from face videos under realistic situations. In: Proceedings of the 2014 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2014, 4264–4271
30. McDuff D, Gontarek S, Picard RW. Remote detection of photoplethysmographic systolic and diastolic peaks using a digital camera. IEEE Transactions on Biomedical Engineering, 2014, 61(12): 2948–2954
31. Kumar M, Veeraraghavan A, Sabharwal A. DistancePPG: Robust non-contact vital signs monitoring using a camera. Biomedical Optics Express, 2015, 6(5): 1565
32. Tasli HE, Gudi A, Den Uyl M. Remote ppg based vital sign measurement using adaptive facial regions. In: Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP). 2014, 1410–1414
33. Feng L, Po L-M, Xu X, Li Y. Motion artifacts suppression for remote imaging photoplethysmography. In: Proceedings of the 19th International Conference on Digital Signal Processing (DSP). 2014, 18–23
34. Feng L, Po LM, Xu X, Li Y, Ma R. Motion-resistant remote imaging photoplethysmography based on the optical properties of skin. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 25(5): 879–891
35. Wang W, Stuijk S, De Haan G. Exploiting spatial redundancy of image sensor for motion robust rppg. IEEE Transactions on Biomedical Engineering, 2015, 62(2): 415–425
36. McDuff D, Gontarek S, Picard RW. Improvements in remote cardiopulmonary measurement using a five band digital camera. IEEE Transactions on Biomedical Engineering, 2014, 61(10): 2593–2601
37. Chwyl B, Chung AG, Deglint J, Wong A, Clausi DA. Remote heart rate measurement through broadband video via stochastic bayesian estimation. Vision Letters, 2015, 1(1): 5
38. Monkaresi H, Calvo RA, Yan H. A machine learning approach to improve contactless heart rate monitoring using a webcam. IEEE Journal of Biomedical and Health Informatics, 2014, 18(4): 1153–1160
39. Hsu Y, Lin YL, Hsu W. Learning-based heart rate detection from remote photoplethysmography features. In: Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2014, 4433–4437
40. Tran DN, Lee H, Kim C. A robust real time system for remote heart rate measurement via a camera. In: Proceedings of the 2015 IEEE International Conference on Multimedia and Expo (ICME). 2015, 1–6
41. Li M-C, Lin Y-H. A real-time non-contact pulse rate detector based on smartphone. In: 2015 International Symposium on Next-Generation Electronics (ISNE). 2015, 1–3
42. Hoffmann K-P. Biosignale erfassen und

verarbeiten. In: Kramme R, ed. Medizintechnik. Springer, 2011, 667–688
43. Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2001, 511–518
44. Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. In: Proc. European Conference on Computer Vision (ICCV). 1998, 484–498
45. Asthana A, Zafeiriou S, Cheng S, Pantic M. Robust discriminative response map fitting with constrained local models. In: Proceedings of the 2013 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2013, 3444–3451
46. Saragih JM, Lucey S, Cohn JF. Deformable model fitting by regularized landmark mean-shift. International Journal of Computer Vision, 2011, 91(2): 200–215
47. Martinez B, Valstar MF, Binefa X, Pantic M. Local evidence aggregation for regression-based facial point detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(5): 1149–1163
48. Holton BD, Mannapperuma K, Lesniewski PJ, Thomas JC. Signal recovery in imaging photoplethysmography. Physiological measurement, 2013, 34(11): 1499–1511
49. Xu S, Sun L, Rohde GK. Robust efficient estimation of heart rate pulse from video. Biomedical Optics Express, 2014, 5(4): 1124–1135
50. Yu Y-P, Kwan B-H, Lim C-L, Wong S-L, Raveendran P. Video-based heart rate measurement using short-time fourier transform. In: Proceedings of the 2013 International Symposium on Intelligent Signal Processing and Communication Systems. 2013, 704–707
51. Feng L, Po LM, Xu X, Li Y, Cheung C-H, Cheung K-W, Yuan F. Dynamic ROI based on K-means for remote photoplethysmography. In: Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2015, 1310–1314
52. Shi J, Tomasi C. Good features to track. In: Proceedings of the 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1994, 593–600
53. Lucas BD, Kanade T. An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI). 1981, 674–679
54. Bay H, Tuytelaars T, Van Gool L. SURF: Speeded up robust features. In: Computer Vision – ECCV 2006. 2006, 404–417
55. Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564–577
56. Henriques JF, Caseiro R, Martins P, Batista J. Exploiting the circulant structure of tracking-by-detection with kernels. In: Lecture Notes in Computer Science. 2012, 702–715
57. Farnebäck G. Two-frame motion estimation based on polynomial expansion. In: Lecture Notes in Computer Science. 2003, 363–370
58. Tarvainen MP, Ranta-Aho PO, Karjalainen PA. An advanced detrending method with application to hrv analysis. IEEE Transactions on Biomedical Engineering, 2002, 49(2): 172–175
59. Han B, Ivanov K, Wang L, Yan Y. Exploration of the optimal skin-camera distance for facial photoplethysmographic imaging measurement using cameras of different types. In: Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare. 2015, 186–189
60. Zhang K, Zhang L, Yang M-H. Real-time Compressive Tracking. In: Computer Vision – ECCV 2012. 2012, 864–877
61. Fernández A, Carúz JL, Usamentiaga R, Alvarez E, Casado R. Unobtrusive health monitoring system using video-based physiological information and activity measurements. In: Proceedings of the 2015 International Conference on Computer, Information and Telecommunication Systems (CITS). 2012, 1–5
62. Danelljan M, Häger G, Felsberg M. Accurate scale estimation for robust visual tracking. In: Proceedings of the British Machine Vision Conference 2014. 2014, 1–10
63. Rouast P V, Adam MTP, Cornforth DJ, Lux E, Weinhardt C. Using contactless heart rate measurements for real-time assessment of affective

states. In: Davis FD, Riedl R, Vom Brocke J, Léger P-M, Randolph AB, eds. Information Systems and Neuroscience. 2016,

64. Monkaresi H, Hussain MS, Calvo RA. Using remote heart rate measurement for affect detection. In: Proceedings of the Twenty-Seventh International Florida Artificial Intelligence Research Society Conference. 2014, 118–123

65. Zhao F, Li M, Qian Y, Tsien JZ. Remote measurements of heart and respiration rates for telemedicine. PLoS ONE, 2013, 8(10): e71384

66. McDuff D, Gontarek S, Picard R. Remote measurement of cognitive stress via heart rate variability. In: Proceedings of the 2014 36th IEEE Annual International Conference of Engineering in Medicine and Biology Society (EMBC). 2014, 2957–2960

67. Rahman MA, Barai A, Islam MA, Hashem MMA. Development of a device for remote monitoring of heart rate and body temperature. In: Proceedings of the 2012 15th International Conference on Computer and Information Technology (ICCIT). 2012, 411–416

Philipp V. Rouast holds a BSc degree in Industrial Engineering from Karlsruhe Institute of Technology, Germany. As part of his MSc, he is working on a real-time application of Remote Photoplethysmography at the University of Newcastle, Australia. His research interests include machine learning, computer vision and data analytics.

Marc T. P. Adam is a senior lecturer in information technology at the University of Newcastle, Australia. He received a Diploma in Computer Science from the University of Applied Sciences Würzburg, Germany, and a PhD in Economics of Information Systems from Karlsruhe Institute of Technology, Germany. In his research, he investigates the interplay of cognitive and affective processes of human users in electronic commerce.

Raymond Chiong is a senior lecturer at the University of Newcastle, Australia. He is also a guest research professor with the Center for Modern Information Management at Huazhong University of Science and Technology, China, and a visiting scholar with the Department of Automation, Tsinghua University, China. He obtained his PhD degree from the University of Melbourne, Australia, and an MSc degree from the University of Birmingham, UK. His research interests include optimization, machine learning and data analytics. He was the Editor-in-Chief of the Interdisciplinary Journal of Information, Knowledge, and Management. Currently, he is an Editor of Engineering Applications of Artificial Intelligence and an Associate Editor of the IEEE Computational Intelligence Magazine. To date, he has produced over 120 refereed publications.

David Cornforth received the BSc degree in Electrical and Electronic Engineering from Nottingham Trent University, UK, in 1982, and the PhD degree in Computer Science from the University of Nottingham, UK, in 1994. He has been an educator and researcher at Charles Sturt University, the University of New South Wales, and now at the University of Newcastle, Australia. He has also been a research scientist at the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Newcastle, Australia. His research interests are in health information systems, pattern recognition, artificial intelligence, multi-agent simulation, and optimisation.

Ewa Lux is a PhD candidate at the Institute of Information Systems and Marketing, Karlsruhe Institute of Technology in Germany. She received a Bachelor degree in Business Information Systems from Baden-Wuerttemberg Cooperative State University in cooperation with the German Central Bank, and a Master degree in Information Engineering and Management from Karlsruhe Institute of Technology, Germany. Her research interest focuses on emotional processes of economic decision making in electronic markets.