Smart Health 9–10 (2018) 2–11

Contents lists available at ScienceDirect

Smart Health

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

A low-cost, camera-based continuous PPG monitoring system using Laplacian pyramid

Omkar R. Patil[a], Wei Wang[b], Yang Gao[b], Wenyao Xu[b], Zhanpeng Jin[b,*]

[a] Department of Biomedical Engineering, Binghamton University, State University of New York, Binghamton, NY 13902, USA
[b] Department of Computer Science and Engineering, University at Buffalo, State University of New York, Buffalo, NY 14260, USA

A B S T R A C T

Continuous vital signs monitoring plays a critical role in proactively managing people's health in their daily lives. Thus, more affordable, portable, convenient, and reliable health monitoring systems are highly demanded. In this study, we present a new approach capable of extracting photoplethysmogram (PPG) from facial videos using the Laplacian pyramid based signal amplification technique, with the goals of improving the estimation accuracy and enhancing the robustness to varying environmental and operational conditions. To validate our approach in the day-to-day setting, 10 videos from each of the 10 subjects were recorded under various lighting and illumination conditions at different places, over a period of one month. All the results are compared against the popular PPG extraction technique based on the region of interest (ROI) averaging method, as well as the PPG waveforms extracted with the standard pulse oximeter. It is shown that the proposed approach can generate comparable results with improved SNR and a high level of consistency and stability. The average accuracy of the proposed approach for heart rate estimation reaches 93.5%, in comparison with 87.9% for the ROI averaging method.

1. Introduction

The physiological functions of the human body are usually associated with interpretable representations, such as electrical, acoustic, kinematic, or chemical properties (Sörnmo & Laguna, 2005). The measurements of those properties may not provide direct physiological assessment, but they convey useful information with suitable decoding mechanisms. The meaningful interpretations of those measurements reflect the physiological properties of the associated biological system and are hence proven to be very helpful in monitoring and diagnosing pathological conditions. A variety of electrophysiological signals have been extensively studied and widely used for interpreting biological functions, such as electroencephalogram (EEG), electrocardiogram (ECG), photoplethysmogram (PPG), electromyogram (EMG), and electrooculogram (EOG). These bioelectrical signals, individually or collectively, can provide the information of vital signs, including body temperature (BT), blood pressure (BP), heart rate (HR), respiration rate (RR), oxygen saturation (SpO_2), and heart rate variability (HRV).

In the today's world, it is crucial to continuously monitor the status of the body's life-sustaining conditions to determine an individual's health status in daily life or clinical settings. The vital signs are usually monitored by medical professionals periodically. However, without continuous monitoring of those vital signs, some minor symptoms could go undetected and eventually result in clinical deterioration (Acutely ill adults in hospital, 2007). The observable changes in vital signs are often present in a patient for 8 to 24 hours before a sudden life-threatening event (Weenk et al., 2017), such as the sudden cardiac arrest (SCA) and heart attack. Continuous monitoring of vital signs can allow experts to take precautionary actions to prevent such life-threatening events. However, most of the existing vital signs monitors adopted in the clinical settings are neither affordable nor portable, which has been

* Corresponding author.
E-mail addresses: opatil1@binghamton.edu (O.R. Patil), wwang49@buffalo.edu (W. Wang), ygao36@buffalo.edu (Y. Gao), wenyaoxu@buffalo.edu (W. Xu), zjin@buffalo.edu (Z. Jin).

https://doi.org/10.1016/j.smhl.2018.07.024

Available online 07 July 2018

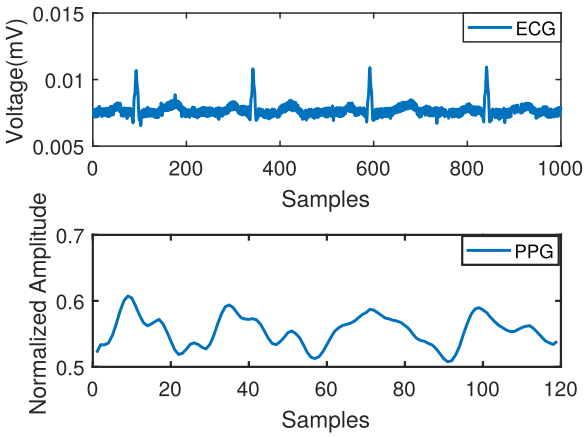

Fig. 1. Example of ECG and PPG waveforms.

prohibitive for their use in individual-centered pervasive healthcare. Therefore, it is important to develop a low-cost monitoring system that can remotely and continuously measure vital signs with ease of use and convenience. The studies on remote or non-invasive monitoring of HR and RR have been extensively reported in literature, using different methods such as laser Doppler (Scalise, 2012), RADAR (Kao, 2013), thermal imagery (Garbey, Sun, Merla, & Pavlidis, 2007), and microwave Doppler radar (Obeid, Zaharia, Sadek, & El Zein, 2012). Nevertheless, these techniques suffer from the following limitations: i) heavy equipment use, ii) high cost, iii) necessity of well-trained manpower, and iv) long measurement time. An ideal vital sign monitoring system should possess the following properties: i) accurate measurement, ii) ease and comfortableness of use, iii) low cost, iv) minimal calibration effort, and iv) noninvasiveness.

PPG is an important physiological signal that can provide the information about the HR, BP, and SpO_2 levels with suitable algorithms (Spetlík, Cech, & Matas, 2018). Each time a heart muscle contracts, blood is ejected from the ventricles and a pulse of pressure is transmitted through the circulatory system. This pressure pulse or blood volume pulse (BVP) when traveling through the vessels is measurable at various points of the peripheral circulatory system and the mechanism used to measure and record this pressure pulse is called photoplethysmography (Elgendi, 2012). PPG provides a descriptive analysis of blood flow through the superficial arterial structure. Two variations of PPG are often used, namely the *transmission mode* and the *reflectance mode*. The transmission mode measures PPG from the changes in the transmitted light (transmitted through tissues, e.g., fingers) caused by pulse waves, while the reflectance mode measures the pulsatile changes in the backscattered light using a unit consisting of the light source and the receiver on the same side (Allen, 2007). The general representations of the PPG waveform and the corresponding ECG waveform are shown in Fig. 1. Given the significance of PPG and its close association with the cardiovascular system, in this study, we propose a new, non-invasive, and non-contact PPG extraction method using the Laplacian pyramid representation to reveal the variations in color segments of an image in a recorded facial video, based on the consumer-grade webcams. In summary, the proposed approach has the following advantages:

1. Enhanced robustness to the varying recording environments and resistance to environmental noises: human subjects were required to record the videos under various environmental conditions including different illuminations and backgrounds;
2. Minimized estimation time: the proposed technique uses only 10 seconds video data to be able to extract reliable PPG and HR;
3. Improved accuracy: the proposed technique can achieve the higher signal-to-noise ratio (SNR) and accuracy compared with the existing Region-of-Interest (ROI) averaging based methods.

This paper is organized as follows: Section 2 introduces the state-of-the-art research efforts in remote monitoring of vital signs based on PPG measurements. Section 3 describes the theoretical backgrounds of the PPG extraction from facial videos and the Laplacian pyramid based image down-sampling and reconstruction. Section 4 provides the specific experimental methods adopted in this work to extract PPG signals and HR. Experimental set-up and results are discussed in Section 5. The conclusion and future work are summarized in Section 6.

2. Related work

The approaches of non-contact PPG monitoring have been widely explored in many studies based on different techniques, such as infrared (IR) light emitting diode (LED) (Shi et al. 2009; Saquib, Papon, Ahmad, & Rahman, 2015; Shi et al., 2010) and camera (Kumar, Veeraraghavan, & Sabharwal, 2015; Zheng, Hu, Chouliaras, & Summers, 2008). Zheng et al. (2008) demonstrated the feasibility of imaging PPG, where a digital CMOS camera and three LEDs were used to detect blood volume changes in tissues underneath the skin. This study proposed the potential use of a camera in recording PPG from specific areas of the human body. The selected skin area was illuminated by LED of specific wavelengths (660 nm and 940 nm) and the skin area was analyzed by a CMOS

camera with 24 to 600 frames per second (fps) which was placed 5 cm away from the illuminated skin region. The experiment was carried out in an optical darkroom to reduce the surrounding light effects. Both transmission mode and reflectance mode experimental setups were conducted by illuminating the subject's finger and upper arm respectively. The results showed promising recordings of the PPG from different ROIs at different LED wavelengths.

A novel approach for continuous monitoring of PPG and HR by installing a PPG sensor on the mattress was presented in (Wong, Pickwell-MacPherson, & Zhang, 2010). The experiments were conducted on 15 individuals in a laboratory environment. The ECG of the subject was recorded simultaneously in order to validate the results by placing electrodes on the subject's right and left forearms. One PPG sensor is clipped with the index finger (contact) and another one is placed underneath the white sheet (non-contact) on the mattress by putting a set of sheets in between the sensor and subject's back as a separator. With the given experimental settings, subject's left trapezius muscles were localized by the PPG sensor. Subjects were asked to wear thin clothes such as T-shirts and rest in the supine position on the mattress for the experiment. The first segment of the experiment was conducted for 6 minutes followed by a second segment of 1 minute, when the subject was asked to sit upright, and the experiment was then continued in the third segment for another 4 minutes in the supine position. The overnight data were recorded for one subject using the same experimental setup. The high correlation among the HRs resulted from ECG and back PPG ($r = 0.9877$), from ECG and finger PPG ($r = 0.9995$), and from back PPG and finger PPG ($r = 0.9878$), respectively, showed the reliability and consistency of the experiment. This study showed promising accuracy for HR monitoring in comparison with contact measurement techniques.

A video-based pulse monitoring system implemented on a mobile service robot was proposed by Stricker, Müller & Gross (2014). This approach aimed to provide a guiding robot to do physical and cognitive tasks and to measure important vital signs. The PPG signals were analyzed from the facial region of a subject through the following steps: ROI selection, signal extraction, and signal processing. In total, five alternative approaches in selecting ROI were evaluated based on the robustness of the system against head and facial movements. The first approach selected the fixed region in an image as ROI and used it for further processing. The Viola-Jones face detection (Viola and Jones, 2004) and the face tracking (Kolarow et al. 2012) algorithms were performed in the second approach. The face tracking approach was further improved by extracting face mask based on skin color segmentation in the third criteria. An improved face tracking algorithm based on deformable model fitting by regularized landmark mean shift (RLMS) was used in the fourth experimental setup to evaluate efficient face detection and tracking. Forehead region was tracked to compensate the facial movements in the fifth approach. The videos were recorded at 640 × 480 resolution with 30 fps from 10 subjects in six experimental setups. SNR was calculated for each approach in different environmental scenarios such as steady, talking, slow translation, fast translation, small rotation, and medium rotation. The average accuracy of the RLMS approach was higher than any other approaches.

Another study on distance measurement of PPG and other vital signs based on camera was performed by Kumar et al. (2015). The combinations of color change signals were obtained from different facial regions using the weighted average method. The weights were selected automatically based on incident light intensity. The average pixel intensity from different regions was combined based on the maximum ratio combining (MRC) to improve SNR in the final PPG estimate. Each ROI was separately tracked by using the Kanade Lucas Tomasi (KLT) algorithm to remove motion artifacts. A camera with the resolution of 1280 × 1024 at 30 fps was used to record videos and a pulse oximeter was used to validate the results. All experiments were carried out under ambient fluorescent light conditions. The SNR for the face averaging method was presented for different skin tone subjects.

Along with these representative works in non-contact PPG monitoring research, several other techniques were also proposed in relevant studies, as listed in Table 1.

These different methods have shown promising outcomes in non-contact PPG monitoring. However, they still have limitations preventing their applicability in the day-to-day continuous measurements. For instance, the subject has to remain still during the entire measurement course to ensure accurate processing. Other limitations include the relatively longer measurement period, the controlled environments (e.g., ROI is illuminated with artificial lights or experiments are performed in the closed lab setting), use of high-end instrument (e.g., expensive cameras), and insufficient validation using data recorded over a short period. Most of the prior studies employed the ROI averaging method that takes the average of a specific color segment over the selected ROI in each frame to extract PPG and HR (Kumar et al., 2015; Poh, McDuff, & Picard, 2011; Zheng et al., 2008). This method shows the potential for HR and PPG extraction, but the subtle changes in color segments are very small that will likely result in constraints for recording setups such as minimum movement of ROI and no light variation. In order to improve the accuracy of current monitoring systems and more importantly, to enhance the robustness and resistance to the ever-changing environmental effects, a low-cost, effective PPG extraction method using cameras based on Laplacian pyramid (Burt & Adelson, 1983; Wu et al., 2012) is proposed and evaluated in this paper.

Table 1
Overview of the state-of-the-art for non-contact vital signs monitoring.

Literature	# of Subjects	Vital signs	Data acquisition
Verkruysse, Svaasand, and Nelson (2008)	1	HR, RR	Camera
Poh et al., (2011)	12	HR, RR, HRV	Camera
Yu et al. (2011)	12	HR, RR	Camera
Wieringa, Mastik, and van der Steen (2005)	7	HR	LED
Humphreys, Ward, and Markham (2007)	10	SpO_2	LED

3. Methods

3.1. PPG extraction from facial video

The absorbance component of arterial blood versus the surrounding tissue gives a distinctive pulsatile nature of arteries, which was first discovered and exploited by T. Aoyagi in 1974 (Aoyagi, 1974) and has thus marked a potential of atrial bedside pulse oximetry for monitoring HR and SpO_2 levels. Based on the same principle, the path length of the incident light on the facial area changes in proportion to the changes of the blood volume in the facial blood vessels (Poh et al., 2011). These subsequent changes depict and reflect the cardiovascular events with suitable extraction algorithms. These reflected signals can be recorded by analyzing color segments (e.g., *RGB)* in each frame $S(1,2,3, \ldots,f)$ where S and f are the reflected intensity for a video sequence and the total number of frames of a facial video respectively. The reflected intensity records over a two-dimensional grid (x, y) of a camera sensor. So, for separate color segments, intensity can be given as $S_r(x, y, f)$, $S_g(x, y, f)$, and $S_b(x, y, f)$ for the *RGB* color model. The measured intensity can be formalized with the reflected intensity from the skin and the intensity/noise of the surrounding ambient light:

$$S(x, y, f) = A(x, y, f)B(x, y, f) \tag{1}$$

where A represents the surrounding ambient light and B is the intensity of light reflected from the skin. To further decompose B, the reflected light intensity from the skin surface and the subskin surface can be considered. The intensity of the light reflected from the subsurface region is essentially the signal of interest for extracting PPG signals. The magnitude of this signal is very low and contains noises due to the surrounding light and the reflected light from the skin surface. It is thus very challenging to extract this low magnitude signal effectively and efficiently from the recorded video.

3.2. Image pyramid coding

Images are represented by pixel values with correlation characteristics in neighboring pixels (Burt & Adelson, 1983). This form contains a lot of redundant information such as noise. In order to extract subtle changes in color segments of an image due to BVP, a compressed representation has to be filtered out from the original image. This compressed image reduces the redundant information with an accurate representation of features related to subtle changes. The transform technique (Cherifi, Beghdadi, & Belbachir, 2010) and predictive technique (Lucas, da Silva, de Faria, Rodrigues, & Pagliari, 2017) have been proposed to extract the compressed representation. The future random variables are predicted based on statistical analysis of the past and present variables. Specifically, for image compression, a pixel value is predicted based on the observable single pixel value or a group of pixel values in previous iterations. This method evaluates the difference between the observable pixel value and the predicted pixel value and then transmits this residual signal for further processing.

The Laplacian pyramid method (Burt & Adelson, 1983; Wu et al., 2012) combines the two image compression techniques namely, the transform and predictive techniques to minimize the redundant information from an image as well as the pixel correlations. A low-pass filter is applied to an original image I_0 and a resulting image I_1 is formed. This step filters the low-scale noise from an image and gives the reduced version of an original image. This step is similar to convolution with the group of local weighting functions. In the next iteration, image I_2 is generated by filtering I_1 in a similar way. The original and generated images I_0, I_1, I_2, …., I_n are together called as a Gaussian pyramid due to its resemblance to the Gaussian probability distribution. Suppose that the size of the original image I_0 is $x \times y$ where x and y represent the number of columns and rows in the image array. I_0 is the bottom level image in Gaussian pyramid with I_1 at level one and so on. The light intensity for each pixel is represented by L ranging from 0 to $R - 1$. Each corresponding value in the image I_1 is a weighted representation of the specific window in the image I_0. In the same way, I_2 is obtained from I_1 by evaluating the similar weighting algorithm. This can be formalized as:

$$I_n(i, j) = \sum_{a=s_1}^{s_2} \sum_{b=s_1}^{s_2} w(a, b) I_{n-1}(2i + a, 2j + n) \tag{2}$$

where n is the number of the levels in the pyramid, a and b represent the sizes of the window, and i and j represent the specific nodes ($1 \leq i < x$ and $1 \leq j < y$). It has been shown that the selection of window size does not affect the performance significantly (Dollár, Appel, Belongie, & Perona, 2014). Further, in order to reverse this procedure, i.e., to obtain an original image from a compressed representation, new pixel values are interpolated between the predicted values. So, $I_{n,l}$ which is of the same size as I_{n-1} can be obtained from I_n where l is the number of the level in expanding image. So, $I_{n,0} = I_n$. The Laplacian pyramid is constructed to further reduce the image in order to obtain the compressed feature set. The residuals that are retained by subtracting an expanded image I_1 from I_0 constitute the bottom level for the Laplacian pyramid. In the next iteration, I_1 is used along with the expanded I_2 to generate the next level error image. So, Laplacian pyramid can be defined as a sequence of error images E_0, E_1, E_2,….,E_n from the corresponding two levels of Gaussian pyramid. Thus,

$$En = In - I(n + 1), 1 \tag{3}$$

The subtracting technique reduces most of the dependency among pixels in an image. The entropy of the pixel distribution is the minimum number of steps for each pixel to recover the original pixel value from a compressed image where the neighboring pixels are statistically independent. The optimal value of the entropy can be selected by quantizing pixel values in a compressed image in each layer of the Laplacian pyramid. The stepwise representation of the Laplacian pyramid procedure for decoding and encoding an image is shown in Fig. 2.

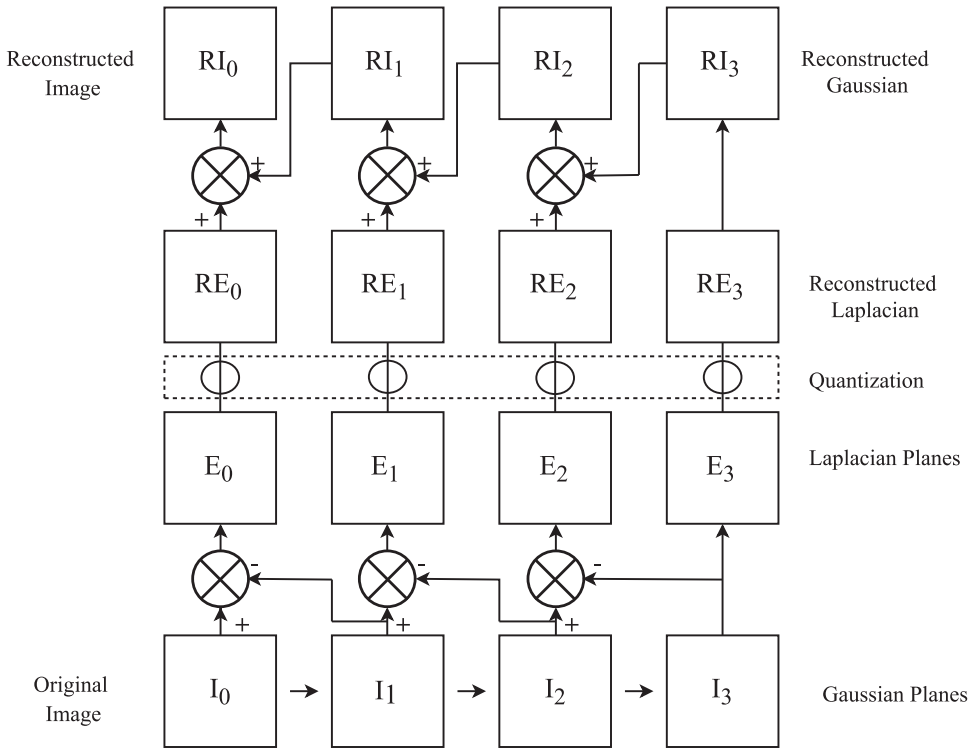

Fig. 2. Illustration of Laplacian pyramid construction (Burt and Adelson, 1983).

4. Processing flow

The overall algorithm is composed of the following steps: face detection, ROI selection, signal filtering along with Laplacian pyramid and amplification, color channel selection and signal extraction, Independent Component Analysis (ICA) application, bandpass filtering for specific narrowed range, moving window sampling, and PPG and HR extraction. The flow diagram for PPG extraction is shown in Fig. 3.

4.1. Face detection and ROI selection

As the first step in the processing flow, face detection is carried out in the first frame of the recorded video followed by tracking in the consecutive frames and the forehead region is selected as ROI. The selection of the forehead region for PPG extraction is based on the assumption that the forehead region is least affected by facial movements such as talking, smiling, or laughing. The Viola-Jones algorithm along with the KLT algorithm are used for face detection and face tracking (Tommasini, Fusiello, Trucco, & Roberto, 1998; Viola & Jones, 2004).

4.2. Signal amplification

Our approach aims to filter out all the corresponding noises and extract subtle changes in color segments in time series at the specific spatial coordinates. These subtle changes are with significantly low signal strength because the vascular structure underneath the facial skin contains only 2%-5% of the total blood volume and only $\sim$ 5% of the total blood volume changes in proportion to cardiac pulse (Hu, Azorin-Peris, & Zheng, 2013). In order to enhance the subtle color segment changes, each channel is extracted and magnified in a specific frequency band of interest with the construction of Laplacian pyramid and amplification factor α (Wu et al., 2012). All the frames are spatially low-pass filtered to increase the SNR by convolving the number of pixels and reducing redundant information with Laplacian pyramid construction as described in Section 3. The down-sampling of the video sequence provides low-pass filtering and better computational efficiency by reducing quantization.

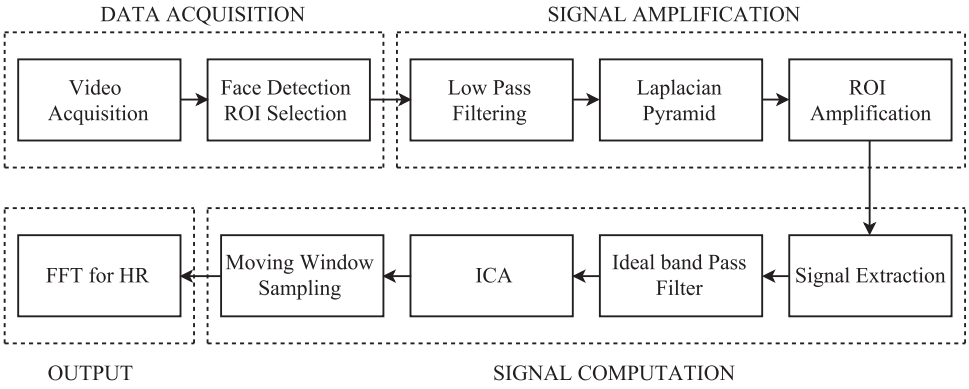

Fig. 3. System architecture for PPG and HR extraction.

The frame sequence is then decomposed as per the application into different spatial frequency bands. Each spatial band is processed along the time series by considering temporal location corresponding to the pixel value in the decomposed spatial frequency band. A bandpass filter is used to extract the frequency band of interest. The selection of frequency for bandpass filter is application based, e.g., frequency range of 0.5–4 Hz can be selected for the corresponding pulse rate of 30–240 beats per minute (BPM) and it can be narrow down to the specific range to evaluate accurate pulse rate, e.g., 0.8–1.6 Hz corresponding to normal human heart beat range of 50–100 BPM (Fleming et al., 2011). All the pixels in each level of frequency band exhibits uniform temporal stability. The extracted band of interest is multiplied with the amplification factor α to enhance the signal magnitudes. The amplification factor is attenuated beyond the cutoff frequency ω_c , corresponding to the spatial wavelength $\lambda_c = \frac{2\pi}{\omega_c}$ which significantly reduces noise in the specific frequency band of interest. The attenuation of the amplification factor can be selected as zero or can be scaled down to zero linearly.

A low-order IIR filter is used in the first stage of filtering as it is computationally efficient for real-time implementation. The filter selection can vary with different applications or tuned for the optimum outputs. The specific values for α and ω_c are tuned to boost subtle color segment changes and reduce noises.

4.3. Signal computation

The camera sensor records the red $S_r(t)$, green $S_g(t)$, and blue $S_b(t)$ color segments over time t for the frame sequence of the forehead region, reflecting the corresponding plethysmographic changes. All the pixels in the selected ROI do not ensure subsequent changes in pixel values in accordance with cardiovascular activities. As the signal is bandpass filtered for the maximum human heart beat range (30–240 BPM) in the signal amplification step, it is comprised of noises along with information in a specific frequency band related to the subject heart beat range. Therefore, combining color segment values for all the pixels to a single spatially averaged value over the ROI frame is beneficial for reducing artifacts like ROI movements or muscle movements.

$$I_c(t) = \frac{1}{\widehat{X}.\,\widehat{Y}} \sum_{x}^{\widehat{X}} \sum_{y}^{\widehat{Y}} S(t, x, y, c) \tag{4}$$

where $I_c(t)$ represents the spatially averaged intensity at time t, $\widehat{X}$ and $\widehat{Y}$ are the width and height of ROI respectively, c represents the extracted color channel (e.g., for red color segment $c = 1$), x and y are the pixel coordinates in a given frame.

The spatially averaged signals $I_c(t)$ were detrended (Tarvainen, Ranta-Aho, & Karjalainen, 2002) and normalized by performing the ICA as:

$$\widehat{I_c}(t) = \frac{I_c(t) - \mu_c}{\sigma_c} \tag{5}$$

where $\widehat{I_c}(t)$ is the normalized source signal, $c = 1, 2, 3$ represent the RGB color channels, μ_c and σ_c represent the mean and the standard deviation for $\widehat{I_c}(t)$. ICA is performed to extract the original source signal $\widehat{s_1}(t)$, $\widehat{s_2}(t)$, and $\widehat{s_3}(t)$ from the extracted RGB signals namely: $I_1(t)$, $I_2(t)$, $I_3(t)$. ICA is a technique used to reveal the underlying source signal by reducing statistical dependence from a given observed signal which is a linear mixture of independent signals (Hyvärinen, 2000). As the observed signals are the linear mixture of the underlying source signals, i.e., $\widehat{I_c}(t) = Ms_c(t)$ where M is a coefficient matrix, $\widehat{I_c}(t) = [\widehat{I_1}(t), \widehat{I_2}(t), \widehat{I_3}(t)]^T$, and $s_c(t) = [s_1(t), s_2(t), s_3(t)]^T$. ICA finds the inverse of matrix M such that $\widehat{s_c}(t) = W\widehat{I_c}(t)$ where W is an approximation of the inverse matrix of M and $\widehat{s_c}(t)$ is the extracted source signal. ICA is performed on the normalized signals by using Joint Approximate Diagonalization of Eigen matrices (JADE) (Cardoso, 1999). In order to further narrow down the frequency resolution, an ideal bandpass

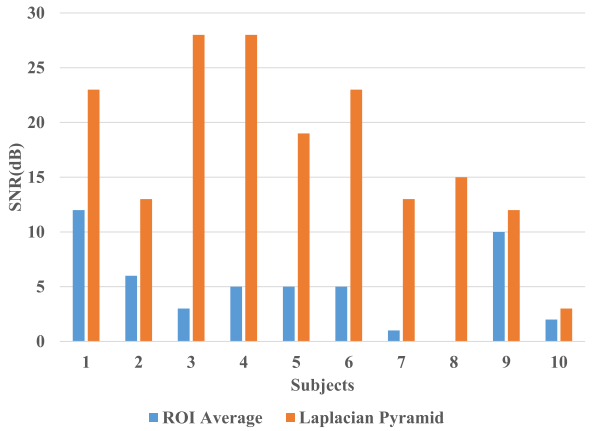

Fig. 4. SNR comparison for ROI averaging and Laplacian pyramid method.

filter is used with the frequency of 0.75–1.66 Hz corresponding to 45–100 BPM. The moving window integration is performed to obtain the intrinsic features of the waveform and is expressed as:

$$y(nT) = \frac{1}{w}[\widehat{s_c}(nT - (N-1)T) + \widehat{s_c}(nT - (N-2)T) + ... + \widehat{s_c}(nT)] \tag{6}$$

where y is the moving window averaged signal, w is the window size, n is the sampling point, and T is the sampling period. The size of the window is not arbitrary as it directly affects the accuracy. With the 2 seconds (60 sample points) window, the algorithm gives the optimal performance. The signal is further computed for HR detection in the following steps: Fast Fourier Transform (FFT) and peak detection. The signal is transformed into the frequency domain using FFT and the corresponding frequency for the highest peak is evaluated to determine HR.

5. Experiments and results

Our approach aims to extract PPG signals in the most generalized environment with a minimum extraction time, in comparison with prior studies where experiments were carried out in a controlled lab environment. This experiment was performed on 10 subjects (6 males and 4 females) of age 23–32 years old. Each subject recorded 10 videos, each 12–15 seconds long, over a period of one month. A 10 seconds video was selected from the recorded data for actual experimental procedure. The videos were made in daylight at different places with different surrounding light conditions. The videos were recorded with the minimum resolution of 640 × 480 at 30 fps. To validate our results, a pulse oximeter (iHealth Air, model PO3) has been used as a contact measurement device. All the experimental procedures are reviewed and approved by the Internal Review Board (IRB) of Binghamton University.

It has been well proven by prior studies that the ROI averaging method can extract PPG with very high similarity as the ground truth PPG waveform. We applied a similar method to extract PPG from the recorded video and compared the results with our proposed Laplacian pyramid based method. The ROI averaging method uses ICA as part of signal extraction (Poh et al., 2011). Hence , in other words, results are compared for ICA (ROI averaging method) and ICA + Laplacian pyramid (our proposed method). The SNRs for both methods are evaluated to compare the performance. The SNR is calculated by using the periodogram of the same size as the extracted PPG waveform.

The first six harmonics were excluded and the periodogram was evaluated with a Kaiser window function (Higham & Higham, 2016).

Fig. 4 shows the comparison results for both ROI averaging and Laplacian pyramid methods in terms of SNR of the signal

Table 2
True HR value and the mean HR values with the ROI averaging and the Laplacian pyramid based methods. % Acc. 1 and % Acc. 2 represent the percentage accuracy of ROI averaging and Laplacian methods for HR extraction respectively.

Subjects	True HR	ROI averaging	Laplacian pyramid	% Acc.1	% Acc.2
1	65	59	62	90.76	95.58
2	68	64	65	94.11	95.58
3	66	74	65	87.87	98.48
4	75	73	74	97.33	99.98
5	87	71	79	81.60	90.8
6	73	62	64	84.93	87.67
7	59	70	64	81.35	91.52
8	72	62	69	86.11	95.83
9	65	69	67	93.84	96.92
10	75	61	62	81.33	82.66

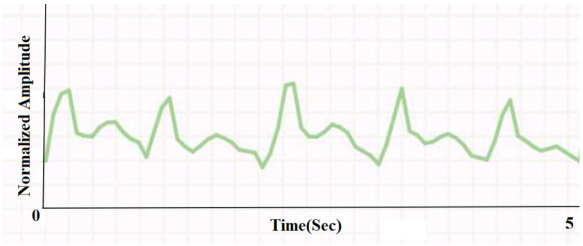

Fig. 5. PPG waveform acquired with a contact device (iHealth Air PO3 pulse oximeter).

extracted from a single video for each individual. Clearly, SNR for Laplacian pyramid based method is higher than SNR for the ROI averaging method, which demonstrates better noise-free signal extraction with the Laplacian pyramid method. The Laplacian pyramid method improves the SNR for PPG extraction by enhancing the subtle changes due to BVP in recorded video and filtering them for a specific range compared with the ROI averaging method where these changes are averaged over a selected ROI.

Table 2 shows the average HR of 10 trials of video recordings from 10 subjects extracted using both the ROI averaging and the Laplacian pyramid based methods. The overall accuracy with the Laplacian pyramid based method (ranging from 82.66% to 99.98%) is higher than the ROI averaging method (ranging from 81.33% to 97.33%). The standard deviations for HR estimation using the Laplacian pyramid based method for all the subjects range from ± 1.89 mean 74 to ± 4.39 mean 64. The accuracy for the HR monitoring shows the consistency with SNR levels plotted in Fig. 4. It can be observed that the maximum accuracy of 99.98% for subject 4 also corresponds to the highest SNR level and the lowest accuracy for subject 10 correlates with the lowest SNR level.

The morphological comparison between PPG acquired with the contact device and PPG extracted from the selected ROI region using the Laplacian pyramid based method for 5 seconds is shown in Figs. 5 and 6 respectively. Fig. 5 shows the PPG waveform recorded with the contact oximeter device that is considered as a ground truth data in order to compare the PPG waveform acquired with the proposed method shown in Fig. 6. Fig. 7 represents the morphological comparison of the PPG waveforms extracted with the ROI averaging method and the Laplacian pyramid based method. It is shown that the upper one is susceptible to artifacts such as motion or surrounding light variations. Although the morphological parameters for the PPG waveform extracted with the Laplacian pyramid method shows small deviations, they don't vary significantly. Therefore, a stable and reliable morphology is shown by the Laplacian pyramid method. Also, Figs. 6 and 7 confirm that the interbeat interval as well as peak-to-peak distance for the highest peaks are similar so that both methods can be useful for accurate HR monitoring. The error bars for HR estimated using the Laplacian pyramid based method for all subjects are plotted in Fig. 8, showing the precise monitoring for multiple videos recorded over a month.

6. Conclusion and future work

In this paper, we propose a new non-invasive and non-contact PPG estimation approach using facial videos, based on the Laplacian pyramid technique, which has shown significant improvement in accuracy and robustness compared with the previously

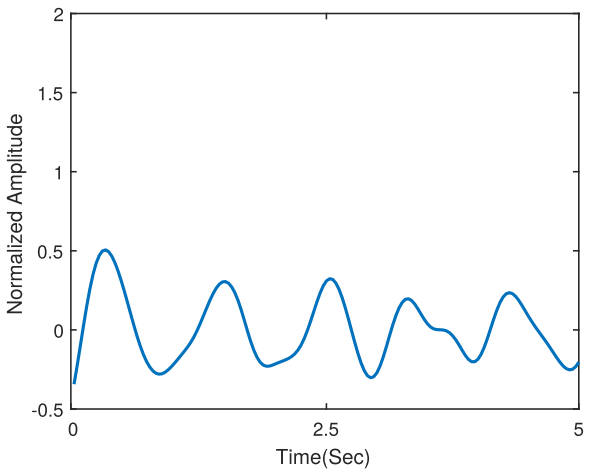

Fig. 6. PPG waveform evaluated with Laplacian pyramid based method.

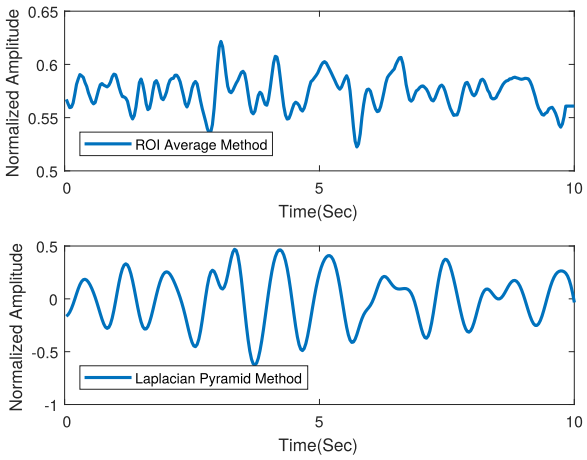

Fig. 7. PPG waveform for ROI average (top) and Laplacian pyramid based (bottom) methods.

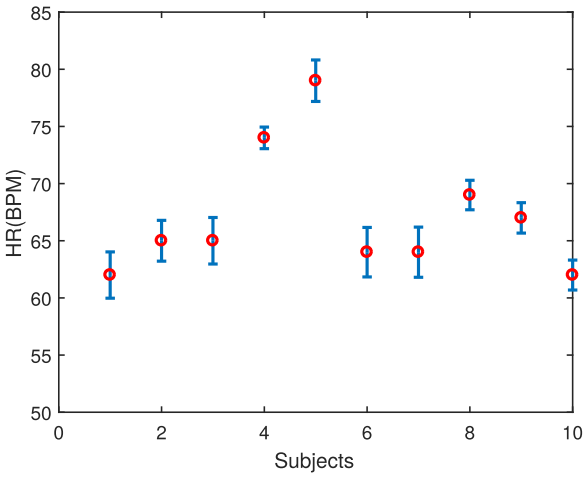

Fig. 8. Error bar for HR estimated with Laplacian pyramid based method.

used ROI averaging method. The experimental results show that the proposed approach is robust to the varying environmental conditions, demands less time for detection, and has been validated using the videos recorded over a period of one month. The morphological characteristics between PPG recorded with a contact device and with the Laplacian pyramid approach present a very high similarity level. The relatively small standard deviations validate the long-term reliability of the approach for monitoring HR and PPG. On the other side, the current limitations of this study include: large motion artifacts that can affect the accuracy as well as the selection of different parameters such as $\propto$, λ_c and the cutoff frequency band for an ideal bandpass filter needs tuned inputs. The future work will primarily focus on the implementation of real-time continuous monitoring of PPG and HR with this method. Also, expansion of the subject size as well as the number of videos for each individual to validate this approach for a broader scope.

Conflict of interest statement

None.

References

Sörnmo, L., & Laguna, P. (2005). *Bioelectrical signal processing in cardiac and neurological applications*, *8*. Amsterdam: Academic Press.

Acutely ill adults in hospital (2007). Recognising and responding to deterioration. Tech. Rep. Clinical Guideline [CG50], National institute for health and care excellence (July).

Weenk, M., van Goor, H., Frietman, B., Engelen, L. J., van Laarhoven, C. J., Smit, J., Bredie, S. J., & van de Belt, T. H. (2017). Continuous monitoring of vital signs using wearable devices on the general ward: Pilot study. *JMIR Mhealth Uhealth, 5*(7).

Scalise, L. (2012). Non contact heart monitoring. In R. Millis (Ed.). *Advances in electrocardiograms*Rijeka: InTech (Ch. 4).

Kao, T.Y.J., Lin, J. (2013). Vital sign detection using 60-GHz Doppler radar system. In Proceedings of the IEEE International wireless symposium. (IWS), pp. 1–4.

Garbey, M., Sun, N., Merla, A., & Pavlidis, I. (2007). Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. *IEEE Transactions on Biomedical Engineering, 54*(8), 1418–1426.

Obeid, D., Zaharia, G., Sadek, S., & El Zein, G. (2012). Microwave doppler radar for heartbeat detection vs electrocardiogram. *Microwave and Optical Technology Letters, 54*(11), 2610–2617.

Špetlík, R., Čech, J., Matas, J. (2018). Non-contact reflectance photoplethysmography: progress, limitations, and myths . In *Proceedings of the 13th IEEE international conference on automatic face & gesture recognition* (pp. 702–709). (FG), IEEE .

Elgendi, M. (2012). On the analysis of fingertip photoplethysmogram signals. *Current Cardiology Reviews, 8*(1), 14–25.

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

Shi, P., Hu, S., Echiadis, A., Peris, V.A., Zheng, J., Zhu, Y. (2009). Development of a remote photoplethysmographic technique for human biometrics. In Design and Quality for Biomedical Technologies II, Vol. 7170, In *Proceedings of the international society for optics and photonics* (p. 717006).

Saquib, N., Papon, M.T.I., Ahmad, I., Rahman, A. (2015). Measurement of heart rate using photoplethysmography. In *Proceedings of the 9th international conference on networking systems and security* (pp. 1–6). (NSS), IEEE.

Shi, P., Azorin-Peris, V., Echiadis, A., Zheng, J., Zhu, Y., Cheang, P. Y. S., & Hu, S. (2010). Non-contact reflection photoplethysmography towards effective human physiological monitoring. *Journal of Medical and Biological Engineering, 30*(3), 161–167.

Kumar, M., Veeraraghavan, A., & Sabharwal, A. (2015). DistancePPG: Robust non-contact vital signs monitoring using a camera. *Biomedical Optics Express, 6*(5), 1565–1588.

Zheng, J., Hu, S., Chouliaras, V., Summers, R. (2008)., Feasibility of imaging photoplethysmography. In *Proceedings of the international conference on biomedical engineering and informatics* (pp. 72–75). (BMEI), Vol. 2, IEEE.

Wong, M. Y. M., Pickwell-MacPherson, E., & Zhang, Y. (2010). Contactless and continuous monitoring of heart rate based on photoplethysmography on a mattress. *Physiological Measurement, 31*(7), 1065.

Stricker, R., Müller, S., Gross, H.-M. (2014). , Non-contact video-based pulse rate measurement on a mobile service robot. In *Proceedings of the 23rd IEEE international symposium on robot and human interactive communication. (ROMAN)* (pp. 1056–1062). IEEE.

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

Kolarow, M., Brauckmann, M., Eisenbach, K., Schenk, E., Einhorn, K., Debes, H. (2012). Gross, Vision-based hyper-real-time object tracker for human robot interaction. In *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems. (IROS)*. (pp. 2108–2115).

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

Poh, M.-Z., McDuff, D. J., & Picard, R. W. (2011). Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering, 58*(1), 7–11.

Yu, S., Hu, S., Azorin-Peris, V., Chambers, J. A., Zhu, Y., & Greenwald, S. E. (2011). Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise. *Journal of Biomedical Optics, 16*(7), 077010.

Wieringa, F. P., Mastik, F., & van der Steen, A. F. (2005). Contactless multiple wavelength photoplethysmographic imaging: A first step toward “SpO2 camera” technology. *Annals of Biomedical Engineering, 33*(8), 1034–1041.

Humphreys, K., Ward, T., & Markham, C. (2007). Noncontact simultaneous dual wavelength photoplethysmography: A further step toward noncontact pulse oximetry. *Review of Scientific Instruments, 78*(4), 044304.

Burt, P. J., & Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications, 31*(4), 532–540.

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), 1–8.

Aoyagi, T. (1974). Improvement of the earpiece oximeter. *Abstracts of the Japanese Society of Medical Electronics and Biological Engineering, 1974*, 90–91.

Cherifi, D., Beghdadi, A., & Belbachir, A. (2010). Color contrast enhancement method using steerable pyramid transform. *Signal, Image and Video Processing, 4*(2), 247–262.

Lucas, L. F. R., da Silva, E. A. B., de Faria, S. M. M., Rodrigues, N. M. M., & Pagliari, C. L. (2017). Prediction techniques for image and video coding. *Efficient predictive algorithms for image compression* (pp. 7–33). Cham: Springer International Publishing, Springer Science & Business Media B.V.

Dollár, P., Appel, R., Belongie, S., & Perona, P. (2014). Fast feature pyramids for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(8), 1532–1545.

Tommasini, T., Fusiello, A., Trucco, E., Roberto, V. (1998). Making good features track better. In *Proceedings of the conference on computer vision and pattern recognition. (CVPR’98)* pp. 178–183. IEEE.

Hu, S., Azorin-Peris, V., & Zheng, J. (2013). Opto-physiological modeling applied to photoplethysmographic cardiovascular assessment. *Journal of Healthcare Engineering, 4*(4), 505–528.

Fleming, S., Thompson, M., Stevens, R., Heneghan, C., Plüddemann, A., Maconochie, I., Tarassenko, L., & Mant, D. (2011). Normal ranges of heart rate and respiretory rate in children from birth to 18 years of age: A systematic review of observational studies. *The Lancet, 377*(9770), 1011–1018.

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.

Hyvärinen, E. Oja (2000). Independent component analysis: Algorithms and applications. *Neural networks, 13*(4-5), 411–430.

Cardoso, J.-F. (1999). High-order contrasts for independent component analysis. *Neural Computation, 11*(1), 157–192.

Higham, D. J., & Higham, N. J. (2016). *MATLAB Guide* (3rd Ed.). Society for Industrial and Applied Mathematics.