

DISTANT MEASUREMENT OF PLETHYSMOGRAPHIC SIGNAL IN VARIOUS LIGHTING CONDITIONS USING CONFIGURABLE FRAME-RATE CAMERA

Jaromir Przybyło, Eliaz Kątoch, Mirosław Jabłoński, Piotr Augustyniak

AGH University of Science and Technology, Faculty of Electrical Engineering Automatics, Computer Science and Biomedical Engineering, Al. Mickiewicza 30, Kraków, Poland (✉ przybylo@agh.edu.pl, +48 12 617 3873, katoch@agh.edu.pl, mjk@agh.edu.pl, august@agh.edu.pl)

Abstract

Videoplethysmography is currently recognized as a promising noninvasive heart rate measurement method advantageous for ubiquitous monitoring of humans in natural living conditions. Although the method is considered for application in several areas including telemedicine, sports and assisted living, its dependence on lighting conditions and camera performance is still not investigated enough. In this paper we report on research of various image acquisition aspects including the lighting spectrum, frame rate and compression. In the experimental part, we recorded five video sequences in various lighting conditions (fluorescent artificial light, dim daylight, infrared light, incandescent light bulb) using a programmable frame rate camera and a pulse oximeter as the reference. For a video sequence-based heart rate measurement we implemented a pulse detection algorithm based on the power spectral density, estimated using Welch's technique. The results showed that lighting conditions and selected video camera settings including compression and the sampling frequency influence the heart rate detection accuracy. The average heart rate error also varies from 0.35 beats per minute (bpm) for fluorescent light to 6.6 bpm for dim daylight.

Keywords: photoplethysmography, remote patient monitoring, heart rate detection, video signal processing

© 2016 Polish Academy of Sciences. All rights reserved

1. Introduction

During the last decade we witness significant improvement in the development of camera sensors in the aspects of power economy, sampling rate and resolution. Imaging sensors are built into objects of everyday use including TV sets, laptops, smartphones or security systems. These devices can also be applied as valuable sources of health-related information like the heart rate. Worldwide research proved that it is feasible to determine the heart rate by analyzing a plethysmographic signal captured with a regular video camera as image sequences. However, there are no studies that reveal the influence of different sampling rates, lighting conditions and video camera settings on reliability of the heart rate detection.

In our research we examine various aspects of plethysmographic signal acquisition including the lighting conditions, frame-rate, compression and video camera settings and their influence on detection of the heart rate. Noninvasive heart rate monitoring plays a very important role particularly in ambient-assisted living systems for older people. However, heading to an indoor maintenance-free application of seamless cardiovascular monitoring must be preceded by a thorough study of its reliability.

Due to applying a high-quality widely-configurable camera designed for machine vision, we were able to emulate limitations of various parameters present in consumer-grade imaging devices. We adjusted the sensitivity as well as the spatial resolution and the field of view. Moreover, the programmability of the camera frame-rate (up to 200 frames per second – FPS)

enabled to emulate the behaviour of low-frame-rate cameras and study the impact of fluorescent light on the quality of video signal.

The remaining part of the paper is organized as follows. Section 2 presents related works. Section 3 describes videoplethysmographic signal processing rules and presents the experiment equipment and setup. Section 4 presents and discusses the results. Section 5 concludes the paper.

2. Related works

Verkruysse *et al.* [1] were probably the first, who drew attention of the researchers to the possibility of contactless measurement of plethysmographic signals in ambient light (over 1 lx) with a consumer-level video camera. Their predecessors for years have used dedicated light sources in the red or infra-red range, what was adequate for pulse-oximetry due to different absorption of these colours by oxygenated and deoxygenated hemoglobin. The authors collected a footage of several minutes with simple Canon Powershot cameras at 15 or 30 FPS and 640×480 or 320×240 resolutions in daylight or its combination with normal artificial fluorescent light. They found that the strongest plethysmographic signal was obtained in the green (G) channel, although the respiration rate (RR) signal was sometimes more pronounced in the red (R) or blue (B) channels.

Application of the Blind Source Separation technique by Poh *et al.* [2] enabled the videoplethysmographic system to tolerate motion artifacts. The recordings were made with a basic webcam embedded in a laptop in a 24-bit RGB color space at 15 FPS with a pixel resolution of 640×480 . Their algorithm uses the face detection (OpenCV) and automated face tracker to point and follow a region of interest. Three independent colour components were decomposed with Independent Component Analysis (joint approximate diagonalization of eigenmatrices [3]), and the second component is found to always contain the pulse wave. The authors not only demonstrated effective heart rate detection both at rest and during motion, but also performed simultaneous heart rate measurements of multiple participants.

The use of regular RGB cameras was challenged by Jeanne *et al.* [4]. They explored the possibility of touchless monitoring the heart rate for automotive applications and found that in highly dynamic light conditions *infra-red* (IR)-based detection (in a range of 700 nm to 1000 nm) is much more robust. Their video recordings were made at 20 frames per seconds with a resolution of 120×180 pixels, and the results show the *root-mean-square error* (RMSE) less than 1 BPM and the correlation above 0.99 with reference measurements.

McDuff *et al.* [5] proposed extraction of the pulse wave form videos captured with a digital single-lens reflex camera at 30 FPS, with 960×720 resolution. The processing employed Independent Component Analysis and successive signal interpolation and differentiation to distinguish systolic and diastolic peaks. The system was tested with varying mixtures of sunlight and indoor illumination, and also compared the extended colour space (ROGCB – red-orange-green-cyan-blue) and its subspaces with additional orange and cyan sensors.

Another work [6] reports on the use of distant pulse rate measurement in neonates. The system was based on a *high definition* (HD) webcam, while the final image resolution was only 640×480 . The authors used motion compensation, controlled ambient light (300 lx) and separate green channel videos to obtain an average accuracy of 2.52 bpm for all 8 examined objects. They also found that continuous *pulse rate* (PR) monitoring can be improved by selecting and tracking multiple *regions of interest* (ROIs) from video frames to generate respective time-series signals.

In the papers [7] and [8] facial video recordings were made on patients undergoing electrical cardioversion for treatment of *atrial fibrillation* (AF). The authors proposed a Pulse Harmonic Strength concept, based on the ratio of harmonic components (*i.e.* between 0.05 Hz and 3 Hz)

and the total signal energy, to effectively minimize the AF detection error. They used a *Video Graphics Array* (VGA) and *high-definition* (HD) cameras with 15/30 FPS and ambient light of the cardioversion room. The synchronization was made by a flash light. The reported uniqueness of the system consisted in its ability to capture the beat-to-beat interval variation.

The problem of a moving object was solved by Li *et al.* [9] who employed detection and tracing of selected facial features. The proposed system can continuously capture pulse-related absorption of green light by the skin, predicts the location of ROI based on 2D geometric transformation of the face, and is immune to HCI-realistic lighting conditions (*e.g.* when the object watches movies on a screen). Besides compensating the object's motion, a novel method detects non-rigid face deformations (*e.g.* facial expressions) and prevents the heart rate detection algorithm from misleading by transient segments of frames. The system was tested on the MAHNOB-HCI database [10] containing 20 frontal face videos recorded for each object with a resolution of 780×580 pixels at 61 FPS, during his watching movie clips on a computer screen.

An interesting approach has been reported in [11], where the heart rate and beat lengths are extracted from videos by measuring a subtle head motion caused by the Newtonian reaction to the influx of blood at each beat. The described method consists in tracking selected features on the head and further post-processing (made with Principal Component Analysis – PCA) applied in order to decompose their trajectories into a set of component motions. The authors revealed that an important factor which affected their results was the camera speed (expressed in FPS), so that they used the cubic interpolation to increase the sampling rate. Another factor pointed out in that paper was a suboptimal lighting that could affect the algorithm of face feature tracking.

The authors of the work [12] proposed to increase the bit-width resolution in order to improve the pulse recognition ratio. The acquisition was performed with low ratios, *i.e.* 12–30 FPS, assuming that the signal spectrum did not contain more than 2–4 harmonics of a human heart rate. The authors also report that ambient light flickering (100 Hz) can produce alias frequency components close to the cardiac frequency heart rate and thus worsen the quality of the measurement. They proposed a proprietary algorithm to identify and cancel these components. Sugita *et al.* [13] captured images at 140 FPS under a controlled illumination. They used a method of pulse transit time for estimation of blood pressure variations.

Contributing to the research in this area, we aim to examine technical details of the video plethysmography. Our primary goal is to evaluate the possibility of heart rate measurements using industrial or surveillance cameras with a prospect of future hardware implementation (real-time). We aim at analysing the impact of an increased time resolution by performing acquisition of video sequences with an increased frame rate (up to 200 FPS) that enables to make measurement in the frequency domain more precise (and fulfil the Shannon sampling theorem in the presence of flicker frequencies). In this way, we avoid an additional processing stage during the analysis and prepare real-time implementation in the future. The use of different camera settings (frame rate, compression, resolution and spectral sensitivity) and different light sources (fluorescent, dim daylight, infrared light, incandescent light bulb) reveals optimal conditions for distant pulse monitoring [15] and limitations of the method in natural indoor conditions.

3. Materials and Methods

3.1. Setup and acquisition of high-quality videos

To estimate optimal conditions for distant pulse monitoring and limitations of the method in real-world settings, we have recorded several video sequences. The experimental setup

consisted of a monochrome acA200-165 μm camera with a 6 MegaPixel 25 mm focal length lens. The camera's sensor has extended capabilities in the range of infra-red spectrum. This property is advantageous for video acquisition in low-light conditions and when using invisible IR illumination. Its quantum efficiency exceeds 35% in a spectrum range of 400 nm – 800 nm, whereas its maximum values of 65% appear between 500 nm and 650 nm. Beyond 800 nm, the quantum efficiency falls to 5% at 1000 nm. The nominal sensitivity is $5.56 \text{ V}\cdot\text{lux}^{-1}\cdot\text{s}^{-1}$. However, this parameter applies to the pixel array of the sensor, which is mounted behind the lens reducing the illumination accordingly to its properties and actual aperture settings. According to the specification provided by the manufacturer, the Signal-to-Noise-Ratio of the sensor does not fall below $SNR_{max} = 41.3 \text{ dB}$. This value has been verified in the experiment, by analysing noise in a dark image acquired with the blinded lens. An influence of light flickering and other phenomena (e.g. vibrations, a shutter motion, etc.) on the noise characteristics of the sensor has been thoroughly studied and reported in details in the paper [16]. The device captured images in the global shutter mode with a free-run trigger, thus asynchronously with light pulsations caused by the AC power supply.

Advanced features of the camera make the setup highly flexible for acquisition of image frames with various temporal resolutions of up to 200 FPS and adjustable dimensions of image frames. The device provides both automatic and manual setup of exposure and analogue gain parameters that enable to individually adjust to particular lighting conditions. The manual setup was used in order to keep all parameters fixed during subsequent experiments. The proprietary camera software provides strict control of the data flow and recording raw uncompressed videos into a hard disk. This prevents from an accidental loss of data and artefacts that could deteriorate the results of further frequency analysis.

The camera we intentionally used does not include an anti-flicker subsystem that is usually embedded into consumer-grade devices making them immune to periodical fluctuations of the light intensity. Such a system prevents the video signal from appearing the flicker effect by automatically adjusting the trigger, so that setting it matches the phase of mains supplying the light. Another possible option is extending the exposition time to 40 ms to match the interval of powerline-dependent light pulsation. Both enable to acquire the video with minimized artefacts, yet available ranges of acquisition parameters (frame rate, exposure) are limited [17].

The camera was positioned approximately 2.8 m away from the volunteer, in such a way that the face of the monitored person occupied the largest part of focal plane of the camera, as is shown in Fig. 1. Five sequences have been recorded in a greyscale uncompressed 8-bits-per-pixel format, an image resolution of 640×480 pixels and a frame rate of 200 FPS in the following conditions:

- Sequence S_FLU18: fluorescent artificial light (18 tubes);
- Sequence S_FLU06: fluorescent artificial light (6 tubes);
- Sequence S_DAY: dim daylight (indirect daylight illumination from a north-directed window, January, 3:30 PM, cloudy weather);
- Sequence S_IR: infrared IR light;
- Sequence S_BLB: incandescent light bulb.

To record the real value of the pulse (ground truth) a portable OxiMax N-65 (by Nellcor) pulse oximetry monitor has been used. The monitored person was seated motionless during the pulse and video acquisition (approximately 30 seconds of recording). In each frame three distinct ROIs were selected – ROI no. 1 on the upper part of the face, ROI no. 2 on the scene background area and ROI no. 3 on the whole face. The face ROIs were used for heart rate detection, and the background ROI was selected for the reference. An overview of the method used in our experiments is shown in a block diagram (Fig. 2a).

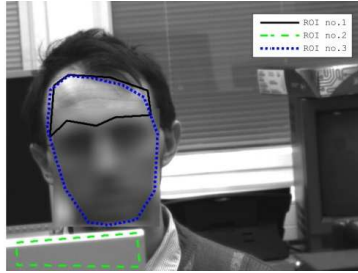


Fig. 1. An example of human head image with marked ROIs.

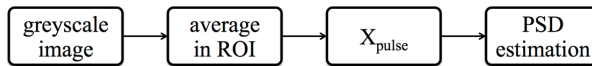


Fig. 2a. An overview of the used method.

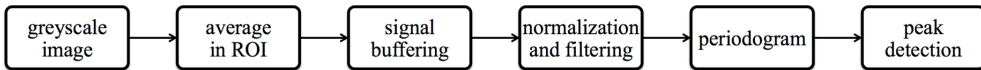


Fig. 2b. An overview of the video processing algorithm.

In the first step, the mean value of pixels is calculated inside a selected *region of interest* (ROI) of each image of the video sequence, resulting in the signal $x_{pulse} = [x_1, x_2, \dots, x_i]$ (where i is the image frame index in the analysed video sequence). Then, we calculate the power spectrum density (PSD) estimate of the input signal with the use of Welch's overlapped segment averaging estimator (MATLAB 'pwelch' routine [14]). Subsequently, we identify the dominant frequencies for the entire video sequences and estimate the influence of video compression on the possible occurrence of artefacts or degradation of pulse data reliability. The results are presented in Sections 4.1–4.3.

3.2. Video processing algorithm

Our aim is enabling to precisely monitor the patient's vital signs in real-time and various lighting conditions, not only in daylight but also in artificial light or dusk. This implies that the RGB video data will not always be available (*i.e.* if a camera has a greyscale sensor or switches to the night mode in which only the luminance component is available). Most of the current state-of-the art pulse wave extraction methods are based on the RGB video data, and therefore further examination is needed in order to estimate the influence of different lighting conditions on heart rate detection in greyscale videos.

An overview of our video processing algorithm used in deriving estimates of the heart rate signal is shown in a block diagram (Fig. 2b). This algorithm has already been tailored to a need of future real-time implementation.

As described in the previous section, the first step is calculation of the mean value of pixels inside a selected *region of interest* (ROI). Then, from a sequence of consecutive frames, the signal x_{pulse}^t is formed in a buffer of finite length N . Besides the pulse, the signal may also contain noise and the respiratory component. Therefore, the next step consists in band-pass filtering of the buffer contents with a least-square linear-phase FIR filter. Simultaneously, the DC offset is removed from the signal. Then, the periodogram estimate has been computed for

the filtered signal. In order to find the pulse frequency, the greatest frequency peak is localized in the periodogram. To obtain an accurate estimate of pulse wave, a central moment in the neighbourhood of the peak is computed. The results are presented in Section 4.4.

3.3. Selection of frequency resolution

An important aspect of the algorithm is the appropriate selection of the frequency resolution Δf and temporal resolution T . The frequency resolution of the power spectrum is defined as (1), and it determines the algorithm accuracy:

$$\Delta f = \frac{F_s}{N}, \tag{1}$$

where N is the buffer length and F_s is the sampling frequency.

An example of analysis is shown in Table 1. The larger the buffer length the higher the frequency resolution. However, a long buffer results in a considerable signal delay (*i.e.* deteriorates the temporal resolution) which might be impractical in real scenarios.

Table 1. The relation between the frequency (Δf) and temporal (T) resolutions for different frame rates and buffer lengths.

F_s [Hz]	N	Δf [Hz]	Δf [bpm (beats per minute)]	T [s]
200	4096	0.049	2.94	20.48
200	2048	0.098	5.86	10.24
200	1024	0.195	11.72	5.12
30	1024	0.029	1.76	34.13
30	512	0.059	3.52	17.07
30	256	0.117	7.03	8.53

4. Results and discussion

For each test sequence, the following analyses have been performed:

- Test 1: Estimation of the power spectrum density (PSD) for all 3 ROIs and computation of the *signal to noise ratio* (SNR).
- Test 2: Estimation of PSD similarly to Test 1, but for compressed S_FLU18 and S_DAY source video sequences (see Section 3.1). The videos have been compressed using MJPEG encoding with high quality settings.
- Test 3: Estimation of PSD similarly to Test 1, but for S_FLU18 and S_DAY source video sequences down-sampled to 28.5 FPS (every 7-th frame taken), a bit rate of 200 kB/s.
- Test 4: Estimation of the heart rate using the proposed algorithm (see Section 3.2).

4.1. Estimation of power spectrum density for high-quality videos

The power spectrum density has been calculated using Welch’s overlapped segment averaging estimator with the rectangular window of length $N = 2048$ shifted by one sample through the entire sequence.

The window length of 2048 samples results in the frequency resolution $\Delta f = 0.0977$ Hz/bin (which is equivalent to 5.8594 bpm – see Table 1) and the temporal buffer window $T = 10.24$ s. Respective PSD estimates for all test sequences, cropped to a range of 1–400 bpm, are shown in Figs. 3 to 7. The frequency range is given in beats per minute (bpm), being equal to $60 * f$.

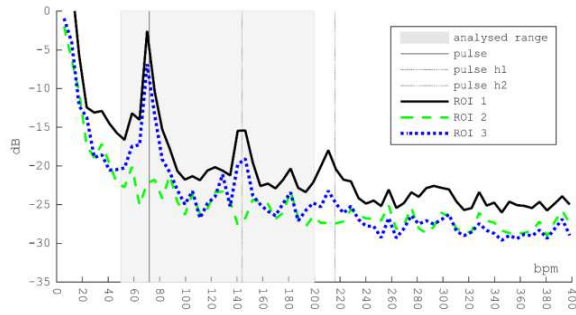


Fig. 3. The power spectrum density estimate of S_FLU18 sequence.

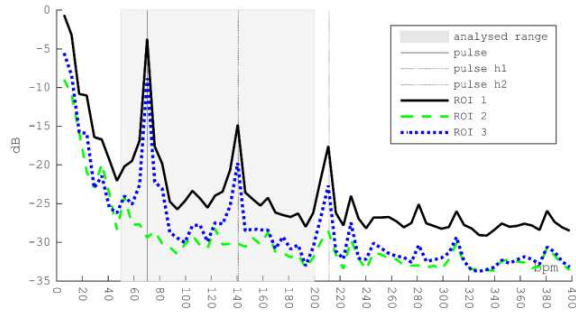


Fig. 4. The power spectrum density estimate of S_FLU06 sequence.

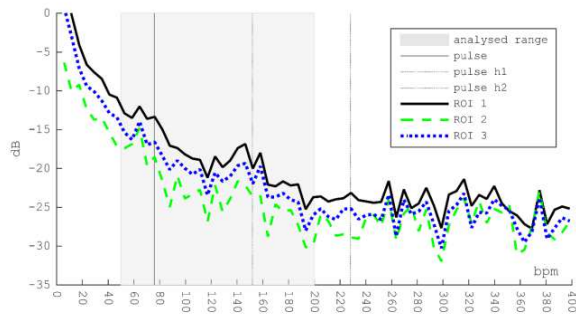


Fig. 5. The power spectrum density estimate of S_DAY sequence.

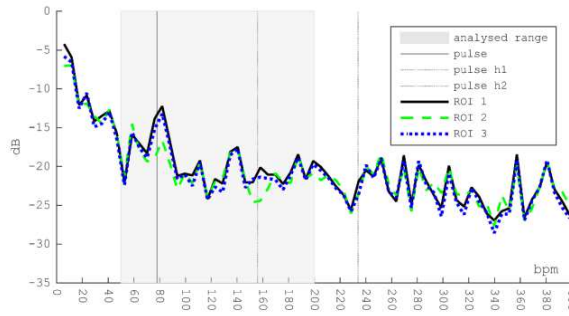


Fig. 6. The power spectrum density estimate of S_IR sequence.

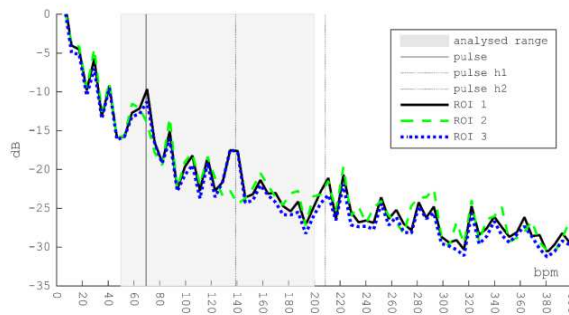


Fig. 7. The power spectrum density estimate of S_BLB sequence.

It can be observed that in all tested videos the pulse frequency peak (and its harmonics, depicted as *pulse h1* and *pulse h2*) is present in the ROI related to the face location. Its power is greater than that in the background ROI (green dotted line) and quantitatively expressed as *signal to noise ratio* (SNR) computed for actual values of the ground truth pulse frequency (Table 2). The algorithm used to compute the SNR was a slightly modified function from MATLAB Signal Processing Toolbox [14]. Also, the SNR is computed only within the range of 50–150 bpm. The function [14] searches the periodogram for the largest nonzero spectral component. However, this value is computed by our algorithm and we want to estimate SNR exactly for this frequency. Thus, we add an additional input argument to the original function – the detected frequency value.

Pulse peaks appear not always at the reference heart rate frequency – this is a result of the limited frequency resolution. Also, for S_DAY and S_IR sequences, the SNR (in dB) is negative, which means that the noise level is higher and might impede correct detection of the heart rate. Also, the value of SNR is greater for ROI 1 (forehead) than for ROI 3 (whole face).

Table 2. The signal to noise ratio (SNR) for the ground truth pulse frequency computed within the range of 50–150 bpm.

Sequence id	Pulse freq. [bpm]	SNR, ROI 1 [dB]	SNR, ROI 2 [dB]	SNR, ROI 3 [dB]
S_FLU18	72.0	5.012	-6.799	5.677
S_FLU06	70.5	8.759	-7.686	6.463
S_DAY	76.0	-2.973	-6.539	-5.835
S_IR	78.0	-2.424	-5.850	-3.063
S_BLB	69.5	2.055	-3.596	2.041

4.2. Estimation of PSD for high-quality videos with compression

The goal of this test is to assess the effect of video data compression, which is an industry standard in surveillance cameras considered for a target solution of heart rate monitoring. To this point we selected two video sequences – with fluorescent light (which contains flickering frequencies) and with dim daylight. The PSD has been estimated in the same way as in the previous experiment (see Section 4.1). Respective PSD estimates and values of SNR for selected test sequences, cropped to a range of 1–400 bpm, are shown in Figs. 8 and 9 and also summarized in Table 3.

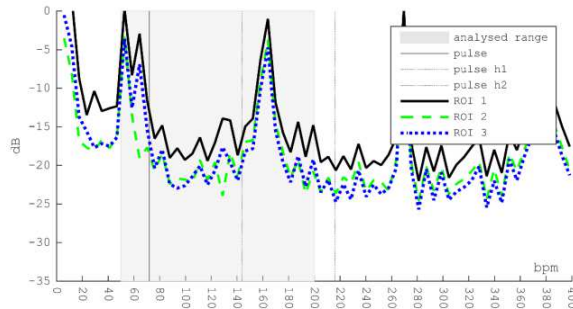


Fig. 8. The power spectrum density estimate of S_FLU18 sequence, MJPEG compression.

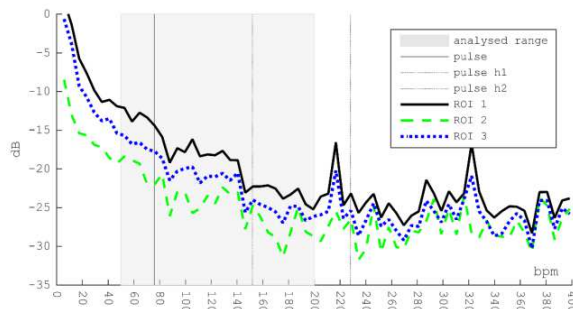


Fig. 9. The power spectrum density estimate of S_DAY sequence, MJPEG compression.

It can be noticed that MJPEG compression introduces significant distortions into the heart rate signal and the value of SNR is much lower. For S_FLU18 sequence (*i.e.* the video taken in fluorescent light) an extra peak at 164 bpm and a shift of the proper pulse peak to the left are observed, probably as a result of interference with the flickering frequency. For S_DAY sequence (*i.e.* the video taken in dim daylight) distortion peaks from compression appear above 200 bpm, which can be easily filtered out.

Table 3. The *signal to noise ratio* (SNR) for the ground truth pulse frequency computed within the range of 50–150 bpm (MJPEG compression).

Sequence id	Pulse freq. [bpm]	SNR, ROI 1 [dB]	SNR, ROI 2 [dB]	SNR, ROI 3 [dB]
S_FLU18	72.0	-5.590	-15.403	-5.759
S_DAY	76.0	-5.312	-8.210	-5.746

4.3. Estimation of PSD for high-quality videos down-sampled to 28.5 FPS

The goal of this test is to assess the effect of limiting the frame rate (*i.e.* using values typical for surveillance or web cameras). To this point we selected two video sequences – with fluorescent light (which contains flickering frequencies) and with dim daylight. The PSD has been estimated using Welch’s overlapped segment averaging estimator, with the rectangular window of length N . The window length of 512 samples lasts $T = 17.92$ s and results in $\Delta f = 0.0558$ Hz/bin (*i.e.* 3.3482 bpm). Respective PSD estimates and values of SNR for selected test sequences, cropped to a range of 1–400 bpm, are shown in Figs. 10–11 and in Table 4.

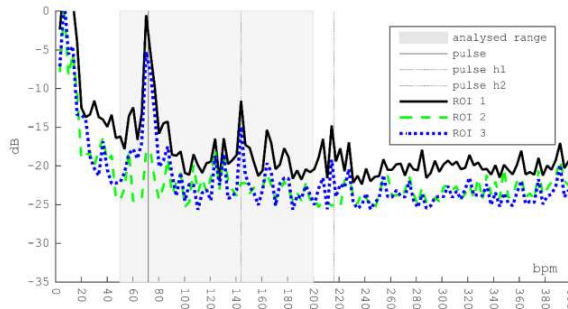


Fig. 10. The power spectrum density estimate of S_FLU18 sequence, downsampling.

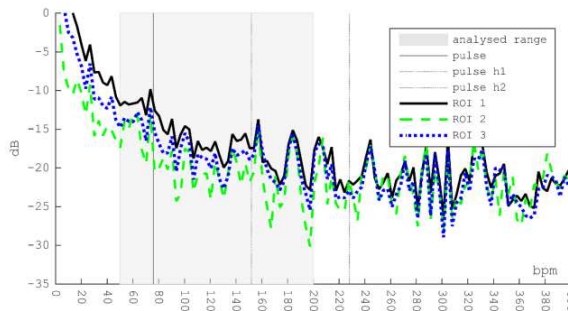


Fig. 11. The power spectrum density estimate of S_DAY sequence, downsampling.

Table 4. The *signal to noise ratio* (SNR) for the ground truth pulse frequency computed within the range of 50–150 bpm (down-sampled signal).

Sequence id	Pulse freq. [bpm]	SNR, ROI 1 [dB]	SNR, ROI 2 [dB]	SNR, ROI 3 [dB]
S_FLU18	72.0	2.956	-7.084	2.613
S_DAY	76.0	-5.715	-9.276	-7.297

4.4. Estimation of heart rate with use of proposed algorithm

To verify the accuracy of the proposed algorithm (see Section 3.2), all video sequences were used. Each of the sequences are processed as follows. In the first step the mean value of pixels inside the selected *region of interest* (ROI) for each video frame is calculated, forming the signal

x_{pulse}^i , where i is the frame number $i = 1 \dots L$. Then, this signal is partitioned in a buffer of finite length N , resulting in the signal x_{pulse}^t , where $t = i \dots L - N$.

The buffer contents are normalized and filtered using a band-pass least-square linear-phase FIR filter with low- and high-frequency cut-offs corresponding to 50 beats per minute (bpm) and 200 bpm, respectively. Then, the periodogram has been calculated on the buffer data. The lengths of $N = 2048$ and $N = 4096$ samples result in $\Delta f = 0.0977$ Hz/bin (5.8594 bpm) and $\Delta f = 0.0488$ Hz/bin (2.9297 bpm), respectively. The pulse frequency f_{pulse}^t is estimated by localizing the greatest frequency peak within the range of 50–200 bpm in the periodogram.

The dominant frequency f_{pulse}^t calculated for each buffer (actually, it is calculated for each video frame), is compared with the value of ground truth of a heart rate, measured with a portable pulse oximeter. The estimation error for i -th frame is computed as the absolute value of difference between the ground truth and detected pulse frequency (2):

$$\Delta P f_i = abs(P f_{ground_truth} - P f_{detected}). \tag{2}$$

The overall detection accuracy values for all video sequences and face ROIs are summarized in Table 5 and Table 6. It should be noted that due to the frequency resolution Δf , the detected pulse can be accepted as correct if it is within the uncertainty range of $\pm \Delta f$. This results in a correct detection rate of 100% for the sequence with fluorescent light (*i.e.* S_FLU sequence) and only 53% for IR light (*i.e.* S_IR sequence). In order to facilitate comparison of the accuracy of all sequences, a visual representation of error distributions (a box plot) is shown in Fig. 13.

Table 5. The pulse wave detection results for the buffer length of 2048 samples and the sequence length of 3951 frames.

Sequence id	Average error [bpm]	Maximum error ROI 1 [bpm]	Average error [bpm]	Maximum error ROI 3 [bpm]
S_FLU18	1.53	4.54	1.60	4.06
S_FLU06	0.35	2.07	0.33	1.53
S_DAY	6.59	65.91	8.07	78.25
S_IR	4.86	104.17	6.29	104.58
S_BLB	2.54	11.43	4.21	18.44

Table 6. The pulse wave detection results for the buffer length of 4096 samples and the sequence length of 1903 frames.

Sequence id	Average error [bpm]	Maximum error ROI 1 [bpm]	Average error [bpm]	Maximum error ROI 3 [bpm]
S_FLU18	0.91	2.03	0.95	1.59
S_FLU06	0.40	1.06	0.40	1.30
S_DAY	4.62	19.24	7.44	38.07
S_IR	2.53	3.95	2.57	3.89
S_BLB	1.63	8.64	2.45	9.36

The presented results confirm correctness of the proposed algorithm. However, the detection is incorrect for some video sequences (especially the videos acquired in low lighting conditions, *e.g.* S_DAY). Similarly, for S_IR sequence and the window length of 2048, the maximal detection error is very high. However, this is caused mainly by outliers (Fig. 12).

The obtained results presented so far can be further improved by using a more sophisticated peak detection algorithm (*e.g.* interpolation of the detected peak's neighbourhood values) or by taking into account the temporal context (*i.e.* exclude spurious detections).

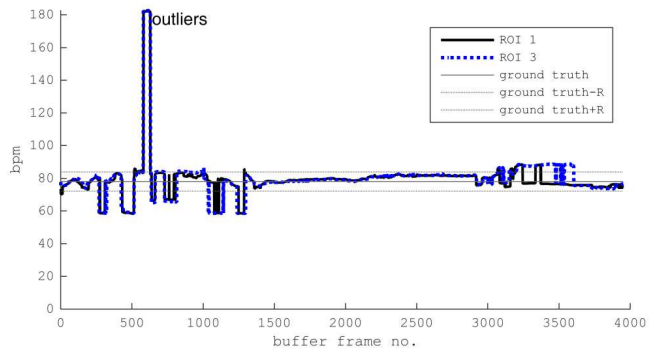


Fig. 12a. The detected heart rate values for S_IR video sequence and the window length of 2048.

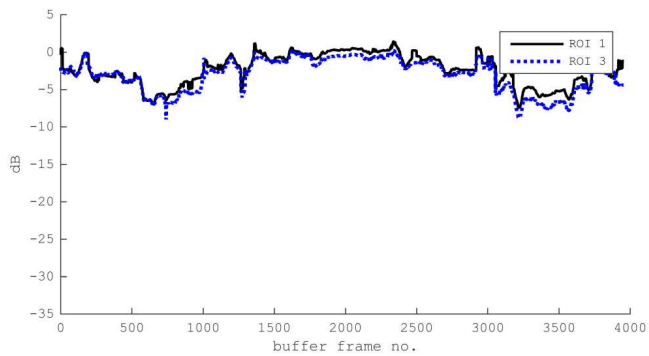


Fig. 12b. The detected SNR values for S_IR video sequence and the window length of 2048.

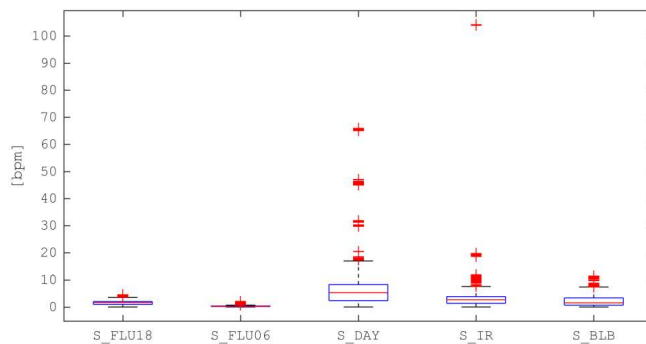


Fig. 13. Box plots of the heart rate detection errors for all video sequences and the buffer length of 2048.

5. Conclusions

We showed that various lightning conditions and selected video acquisition and transmission settings influence the reliability of videoplethysmographic signals. We found that the value of SNR is negative for dim daylight and infrared light, which may suggest that the illuminance level plays important role in correct pulse wave detection. Also, the MJPEG compression introduces high distortions and consequently may degrade the algorithm performance. On the other hand, a lower frame rate does not significantly affect the heart rate detection. However, it might produce aliasing frequencies. An important aspect of the pulse wave detection is an appropriate selection of the buffer length, which influences the frequency and temporal resolution values.

The experiments described in this paper show only a most representative fraction of research carried out in the area of contactless vital sign monitoring. Further research will focus on:

- evaluation of the proposed algorithm regarding the video-footage acquired with surveillance cameras and webcams (to check whether such an equipment can be used for reliable heart rate monitoring);
- developing a method of cancellation of noise frequencies introduced by the MJPEG compression;
- extending the proposed algorithm to detection of other vital signs (e.g. the breathing rate),
- extending the proposed algorithm to work in real conditions (handling head motions, face tracking);
- extending and evaluating a combination of two approaches – the frequency analysis (described in this paper) and the algorithm proposed in [11].

Acknowledgements

This scientific work is supported by the AGH University of Science and Technology in year 2016 as a research project No. 11.11.120.612.

References

- [1] Verkruyse, W., Svaasand, L.O., Nelson, J.S. (2008). Remote plethysmographic imaging using ambient light. *Optics express*, 16(26), 21434–21445.
- [2] Poh, M.Z., McDuff, D.J., Picard, R.W. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18(10), 10762–10774.
- [3] Cardoso, J.F. (1999). High-order contrasts for independent component analysis. *Neural Comput.*, 11(1), 157–192.
- [4] Jeanne, V., Asselman, M., den Brinker, B., Bulut, M. (2013). Camera-based heart rate monitoring in highly dynamic light conditions. *Connected Vehicles and Expo (ICCVE), 2013 International Conference on*, 798–799.
- [5] McDuff, D., Gontarek, S., Picard, R.W. (2014). Remote detection of photoplethysmographic systolic and diastolic peaks using a digital camera. *IEEE Transactions on Biomedical Engineering*, 61(12), 2948–2954.
- [6] Mestha, L.K., Kyal, S., Xu, B., Lewis, L.E., Kumar, V. (2014). Towards continuous monitoring of pulse rate in neonatal intensive care unit with a webcam. *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, 3817–3820.
- [7] Couderc, J.P., Kyal, S., Mestha, L.K., Xu, B., Peterson, D.R., Xia, X., Hall, B. (2014). Pulse Harmonic Strength of facial video signal for the detection of atrial fibrillation. *Computing in Cardiology Conference (CinC)*, 661–664.
- [8] Couderc, J.P., Kyal, S., Mestha, L.K., Xu, B., Peterson, D.R., Xia, X., Hall, B. (2015). Detection of atrial fibrillation using contactless facial video monitoring. *Heart Rhythm*, 12(1), 195–201.

- [9] Li, X., Chen, J., Zhao, G., Pietikainen, M. (2014). Remote heart rate measurement from face videos under realistic situations. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 4264–4271.
- [10] Soleymani, M., Lichtenauer, J., Pun, T., Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1), 42–55.
- [11] Balakrishnan, G., Durand, F., Guttag, J. (2013). Detecting pulse from head motions in video. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 3430–3437.
- [12] Tarassenko, L., Villarroel, M., Guazzi, A., Jorge, J., Clifton, D.A., Pugh, C. (2014). Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiological measurement*, 35(5), 807.
- [13] Sugita, N., Obara, K., Yoshizawa, M., Abe, M., Tanaka, A., Homma, N. (2015). Techniques for estimating blood pressure variation using video images. *Engineering in Medicine and Biology Society (EMBC), 37th Annual International Conference of the IEEE*, 4218–4221.
- [14] MATLAB and Signal Processing Toolbox and Image Processing Toolbox, Release 2016a The MathWorks, Inc., Natick, Massachusetts, United States.
- [15] Przystup, P., Bujnowski, A., Ruminski, J., Wtorek, J. (2013). A multisensor detector of a sleep apnea for using at home. *Human System Interaction (HSI), The 6th International Conference on*, 513–517.
- [16] Sur, F., Grediac, M. (2014). Sensor noise measurement in the presence of a flickering illumination. *Image Processing (ICIP), IEEE International Conference on*, 1763–1767.
- [17] Trzuppek, M., Ogiela, M.R., Tadeusiewicz, R. (2011). Intelligent image content semantic description for cardiac 3D visualisations. *Engineering Applications of Artificial Intelligence*, 24(8), 1410–1418.