

## RELIABILITY OF PULSE MEASUREMENTS IN VIDEOPLETHYSMOGRAPHY

Jacek Rumiński

Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Narutowicza 11/12, 80-233 Gdańsk, Poland  
(✉ jacek.ruminski@pg.gda.pl, +48 58 347 2678)

### Abstract

Reliable, remote pulse rate measurement is potentially very important for medical diagnostics and screening. In this paper the *Videoplethysmography* was analyzed especially to verify the possible use of signals obtained for the YUV color model in order to estimate the pulse rate, to examine what is the best pulse estimation method for short video sequences and finally, to analyze how potential PPG-signals can be distinguished from other (*e.g.* background) signals. The presented methods were verified using data collected from 60 volunteers.

Keywords: videoplethysmography, vital signs measurements, video processing, image processing, smart glasses.

© 2016 Polish Academy of Sciences. All rights reserved

### 1. Introduction

Vital signs are measurements of the body's most fundamental functions. The three main vital signs, often monitored by healthcare professionals, include: body temperature, heart/pulse rate, and respiration rate. Typically, heart rate or pulse rate is clinically calculated using either *Electrocardiography* (ECG) or *Photoplethysmography* (PPG) signals. Traditional PPG measures changes in skin absorption of light using a finger or earlobe probe. The measured arterial pulsation (arterial blood volume change) is caused by pressure pulse. It is typically detected by a photodiode as a response to excitation generated from a light source (*e.g.* a light-emitted diode). However, when the direct contact with the skin is not possible (*e.g.* due to skin damage) other methods of pulse rate estimation are required. In [1] authors demonstrated the possibility of PPG signal measurement using a monochrome camera located 40 cm from the subject's finger. Verkruyssen *et al.* examined a remote PPG method using only ambient light as the source and an RGB camera as the detector [2]. They showed that the green channel contained the strongest PPG signal, corresponding to an absorption peak by (oxy-) hemoglobin, but the red and blue channels also contained some PPG-related information. Poh *et al.* proposed a robust method for computation of the pulse rate from digital color video recordings of the human face [3]. They assumed, that R, G, and B image component signals measured using the *Videoplethysmography* (VPG) are actually a (linear) combination of the pulse signal and other signals (including noise). They proposed to use the *Independent Component Analysis* (ICA) method (JADE algorithm) to remove correlations between and higher order dependence on RGB signals. The experiments performed for 12 subjects showed that the *Root Mean Square Error* (RMSE) was reduced from 6.00 beats per minute (bpm), obtained from the analysis of G – green – trace, to 2.29 bpm.

Later, many different ICA algorithms have been evaluated including FastICA, Jade, RADICAL [4]. The results showed that some algorithms (*i.e.* FastICA, RADICAL) gave better results than other (JADE, Robust ICA) [4, 5]. However, in [5] the authors showed that sometimes there were problems with ICA algorithms, for example, FastICA failed to properly

initialize. It was also shown that using a camera embedded in a laptop computer the use of green trace, bypassing the ICA stage, gave better results.

Earlier in [6] we proposed a method optimized for fast pulse rate estimation. The *Principal Component Analysis* (PCA) algorithm was used to eliminate correlations between RGB signals. Additionally, we compared signals obtained from the entire face region (used in most other works) with signals obtained from the forehead region. The results proved that the accuracy of the proposed method was not worse than that using the ICA algorithm for entire face data.

Video-based pulse rate estimation is a potential application for smart glasses. Smart glasses are mobile, wearable devices that can provide valuable information to the user extending human senses and capabilities of information processing. In recent years, many devices were proposed, including Google Glass, Epson Moverio BT-200, *etc.* [7]. Many demonstrations of potential roles of smart glasses in healthcare were presented. For example, Evena Medical showed the use of smart glasses to improve visualization of veins in/under the skin using IR illumination [8]. In the eGlasses project ([www.eglasses.eu](http://www.eglasses.eu)) we are developing an open smart glasses platform that can be easily modified using additional hardware and software [9]. For example, using a visible-light camera, a thermal camera and a thermal sensor, we can potentially estimate three basic vital signs. For an identified patient (*e.g.* using face recognition [10] or using graphical markers like QR-code [11, 12]) the measured data can be automatically uploaded to the healthcare information system. The application of smart glasses to pulse rate estimation requires analytic methods that can be efficient and can use short time windows of analyzed signals.

So far, estimation of the pulse rate has been mainly based on filtered RGB signals analyzed in the frequency domain [2–5] for the highest peak of the estimated *Power Spectral Density* (PSD) or just in the amplitude spectrum. The method has many limits (*e.g.* limited spectral resolution, “false” pulse rate results for not PPG-signals, *etc.*). Therefore, it is important to analyze different methods for fast and reliable estimation of the pulse rate from video sequences, especially for short-time analysis windows.

In this paper, we would like: to verify the possible use of signals obtained for the YUV color model to estimate the pulse rate, to examine what is the best pulse estimation method for short video sequences and to analyze how potential PPG-signals can be distinguished from other (*e.g.* background) signals.

## 2. Method

### 2.1. Data acquisition

The method is based on continuous measurement of frames from the face region using a digital camera. It is assumed that the face is illuminated by the ambient light source. It is further assumed that the reflected light features, measured by the digital camera, are affected by the properties of skin and skin surface changes. If the parameters of the object and of measurement system are constant (*e.g.* the same illumination), then no differences between measured signals should be obtained for all frames (except an influence of noise). However, in the presence of blood flow there is a dynamic change of arterial blood volume caused by the pressure pulse. This modifies the measured characteristics of light reflected from the face skin. The reflected light is measured by a matrix of detectors of the color digital camera and represented in the RGB or YUV color space. A graphical presentation of the data acquisition setup is shown in Fig. 1a.



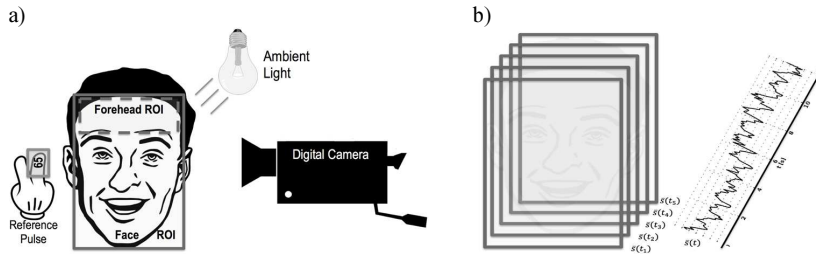


Fig. 1. a) The measurement setup; b) the signal composition procedure.

As a result of the measurement, a sequence of images (video frames) is recorded and encoded in the available video format. The sequence is measured using the configured (possible) sampling rate,  $f_s$ , during the total acquisition time  $T_a$ .

In this study the Logitech Webcam 9000 Pro was used for experiments. The camera has a CMOS optical sensor with maximum resolution of  $1600 \times 1200$  (F/2.0, focal length 3.7 mm) and with theoretically maximum sampling rate  $f_s = 30$  fps (30Hz). However, due to the planned application of the method (smart glasses), the image sequences were acquired with resolution of  $800 \times 600$  using a sampling rate,  $f_s$ , from 15 to 25 fps. Additionally, since most currently available smart glasses models (including the eGlasses prototype) use the Android OS, the YUV color format was chosen for video encoding. This color model is default for the Android OS (NV21 encoding – YUV 4:2:0). Using the Android SDK the captured frames are stored in a data buffer with YUV format, so it is practically more efficient (has better performance) in processing YUV data than RGB data (no format conversion)

As a result of the data acquisition procedure the sequence of YUV frames is available for further processing. In the pre-processing step the face ROI and forehead ROI are manually or automatically selected (or detected). The face ROI is a rectangle covering the entire face region (Fig. 1a). Similarly, the forehead ROI is a rectangle covering the entire visible forehead region of the face. For the needs of this study both ROIs were selected manually by the author using the software developed in Java.

Next, the average value is calculated for each color component of each frame as:

$$\bar{s}(t_k) = \frac{1}{N_{ROI}} \sum_{i=r_s}^{r_e} \sum_{j=c_s}^{c_e} C(i, j), \quad (1)$$

where:  $N$  – the number of pixels in the ROI;  $r_s, c_s$  – the first (start) row and column of the ROI rectangle;  $r_e, c_e$  – the last (end) row and column of the ROI rectangle;  $C(i, j)$  – the pixel value of a color component.

In this study, the original data format of each frame was YUV. Additionally, the data were also converted to the RGB color model. The goal of the operation was to compare the pulse analysis techniques for data represented in RGB and YUV data models. After calculation of average values the samples of digital signals were obtained (Fig. 1b):

$$\bar{s}(t_k) = \{ \bar{s}(t_1), \bar{s}(t_2), \bar{s}(t_3), \dots, \bar{s}(t_K) \}, \quad (2)$$

where  $K$  is the number of frames in the processed video (image sequence).

In the next step the digital signals (for each color component, each color model, each ROI) were normalized by removing the mean value:

$$\bar{s}_n(t_k) = \bar{s}(t_k) - \mu(\bar{s}(t)). \quad (3)$$

The normalized signals were optionally filtered using the moving average operation in the time domain. The window size was experimentally set as  $\text{round}(f_s/5)$ , e.g. 3 for  $f_s = 15$  fps.

## 2.2. Experiments

The data acquisition was performed in two laboratory rooms with the presence of ceiling fluorescent light and, partly, a side natural illumination (overcast). In total, 60 healthy volunteers (Caucasian heritage) participated in the data acquisition step (males: 37, females: 23, average age:  $27.28 \text{ y} \pm 7.18 \text{ y}$ ). During the measurements the subjects were sitting still at a distance of 0.4–0.8 m from the front of the camera. The participants were asked to rest for a while, before starting the data acquisition. Measurement of the reference pulse rate was performed using a Sanitas SPO25 finger pulse oximeter (Hans Dinslage GmbH) during the entire data acquisition time,  $T_a$ . The device was additionally verified using an AsCARD electrocardiograph (AsCARD MrGrey v.201, Aspel) and another pulse oximeter with a finger probe (Vital sign monitor, MD2000A, Beijing Choice Electronic Technology, Co. Ltd.). A simple cross test was performed for one volunteer connected to all devices observing the measured heart/pulse rate values for one minute. The average difference between devices was 2 bpm. This is a typical variation observed due to differences in heart/pulse rate estimation procedures (what is discussed later).

During the video acquisition the pulse rate values were observed on the finger pulse oximeter and were recorded. The video acquisition time was set to 30 s. However, only last 12 s were chosen for further analysis (e.g. the number of samples  $N = 180$  for  $f_s = 15$  Hz or  $N = 300$  for  $f_s = 25$  Hz). The first 18 s were used for additional stabilization of the measurement conditions (e.g. pulse rate at rest). Using the smart glasses or another portable data acquisition system the processing time should be performed in short time windows (e.g. to reduce an object's movements to increase computational performance). For example, future applications can include measurement of vital signs of a patient during a routine interview (a less stressful procedure than that using connected devices) or in screening conditions (e.g. at airport).

Two additional ROIs were chosen to measure the background signals. The first ROI was selected in the subject's hair region while the second ROI was selected for the background ceiling area. The areas of each ROI were set similar to those for the forehead. The signals obtained from these ROIs were used to analyze similarities between signals from the target ROIs (expected pulse wave) and from the background ROIs.

Concluding, 18 signals for each subject were extracted: 6 for the face ROI (for  $Y, U, V, R, G, B$ ), 6 for the forehead ROI (for  $Y, U, V, R, G, B$ ), 3 for the hairs ROI (for  $Y, U, V$ ) and 3 for background ROI (for  $Y, U, V$ ). In total, 1080 signals were further analyzed.

## 2.3. Post-processing

The signals collected during the experiments were further processed to: reduce the baseline, extract uncorrelated signals from  $R, G, B$  signals (PCA), extract independent signals from  $R, G, B$  signals (ICA), and reduce spectrum of signals.

The baseline was removed using a 9th-order high pass digital Butterworth filter with a cutoff frequency set to about 40 bpm. The filter was dynamically designed using the *butter* function from the Octave signal processing toolkit. After the baseline removal the signals obtained for face and forehead ROIs were analyzed to verify the stationarity hypothesis. It was analyzed using the *augmented Dickey-Fuller* (ADF) test at the confidence level of 95% ( $p < 0.01$ , with the stationarity as the alternative hypothesis) [13].

It is known that between RGB and YUV there is a linear dependence, e.g. according to the ITU-R BT.601 specification:



$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}. \quad (4)$$

It can be observed that main components of  $Y$  and  $V$  signals are from  $R$  and  $G$  components, where the *Photoplethysmographic* features are the most significant. So a hypothesis can be assumed that mainly  $Y$  and  $V$  signals should contain the required pulse-related information. Another hypothesis can be assumed that linear decomposition of  $R$ ,  $G$ , and  $B$  signals could lead to  $Y$ ,  $U$ , and  $V$  signals. It means that  $Y$ ,  $U$ ,  $V$  signals should correlate with the results of PCA analysis and maybe with the results of ICA analysis. Therefore, in the preprocessing step,  $R$ ,  $G$ ,  $B$  signals from the face/forehead ROI were processed using PCA, FastICA, Jade ICA, Kernel ICA, and RADICAL ICA to verify both hypotheses and additionally compare different ICA algorithms for the needs of future use. In this study we used Matlab/Octave implementations of the methods provided mainly by the authors: PCA and FastICA from [14], Jade ICA from [15], Kernel ICA from [16], and RADICAL from [17].

The signals obtained after PCA/ICA processing were compared to  $Y$ ,  $U$ ,  $V$  signals, searching for the maximum similarity. Stationarity of all obtained signals was positively tested (ADF < 95%) and then the spectral similarities were analyzed using the magnitude squared coherence estimate function. This *Cross-Spectral Coherence* (CSC) is a function of frequency  $\gamma_{xy}(f)$  that indicates how well  $x$  corresponds to  $y$  at each frequency [18, 19]. The coherence is a function of the power spectral densities,  $P_{xx}(f)$  and  $P_{yy}(f)$ , of signals 1 and 2, and the cross-power spectral density (cross-PSD),  $P_{xy}(f)$ , of signals  $x$  and  $y$ :

$$\gamma_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)}. \quad (5)$$

If signal 1 fully corresponds to signal 2 at  $f$  frequency then the value of the coherence function is 1. The similarity can be typically analyzed for characteristic peak values observed for PSD or for a given spectral range, from  $f_1$  to  $f_2$ . Then the similarity measure can be defined as:

$$AUC\gamma(f_1, f_2) = \frac{1}{L} \sum_{f_1}^{f_2} \gamma_{xy}(f), \quad (6)$$

where  $L$  – the number of frequencies between  $f_1$  and  $f_2$ .

For example, the CSC between any pair of ECG leads will give values greater than 0.9 at most physiologically important frequencies for healthy subjects (1–10 Hz) [20].

Another technique used in the preprocessing phase was band pass filtration for further pulse rate estimation. A Hamming window-based FIR filter was designed, mainly to filter high frequencies (> 4 Hz, 240 bpm). The *fir1* and *filtfilt* Matlab/Octave functions were used.

Next, different techniques for pulse rate estimation were examined. Most papers in the literature use an FFT-based method searching for the highest peak in the frequency spectrum (the method referenced below as *ePR\_sp*). The method has two main disadvantages. First, for short signals it has a low frequency resolution (e.g. for  $N = 180$ ,  $f_s = 15$  Hz, a spectral resolution is  $\Delta f = f_s/N = (15 \text{ Hz} / 180) = 0.0834$  Hz;  $0.0834 \text{ Hz} * 60 \text{ s} \approx 5$  bpm). Secondly, it practically always returns a result even for a noise signal, so for the filtered signals “false” pulse rate values could be returned. The problem is that obtaining only the frequency value we cannot guarantee that we have reliably found a proper pulse rate estimate from the analyzed signal. Therefore, other signal properties should be analyzed based on prior knowledge about



*Photoplethysmographic* signals. The characteristic property of PPG signal is also the waveform morphology. For example, the PPG signal, after moving average filtration, tends to resemble the sine function. We assumed that the VPG signal should dominate in the  $Y$ ,  $U$ , or  $V$  signals and therefore, after filtration, it should be similar to the sine function. If not, then either there is no VPG signal present in the processed signal, or its reliability is low. Additionally, other criteria can be used, for example, verifying if the processed signal is from the face/forehead region (face detection, forehead detection, skin detection, *etc.*).

In this study we calculated the estimate of pulse rate (“ $ePR$ ”) using the following methods:  $ePR_{sp}$  – dominated peak in the frequency domain,  $ePR_{ac}$  – the periodicity of peak locations for the autocorrelation function in the time domain,  $ePR_{zc}$  – the number of zero-crossings,  $ePR_{p1}$  – the number of peaks in the time domain, and  $ePR_{p2}$  – the periodicity of peaks.

The autocorrelation sequence of a periodic signal has the same cyclic characteristics as the signal itself. Therefore, in the  $ePR_{ac}$  method the autocorrelation is calculated and the period is determined calculating average time differences between detected peaks. The  $xcorr$  octave function was used. As a peak detector we used a method searching for a local minimum and a local maximum, for which their difference is greater than the threshold value  $T$ :

$$d_j = \bar{s}_{fn}(t_{j+1}) - \bar{s}_{fn}(t_j), d_j > T, \quad (7)$$

where:  $\bar{s}_{fn}(t_j)$  – the filtered signal value of a local minimum at  $j$ ;  $\bar{s}_{fn}(t_{j+1})$  – the filtered signal value of a local maximum at  $j+1$ .

The threshold value  $T$  was calculated in two phases. In the first phase as:

$$T_1 = T_{K1} * (\max(\bar{s}_{fn}(t)) - \min(\bar{s}_{fn}(t))), \quad (8)$$

where  $T_{K1}$  was a scaling value set to 0.33.

As a result, the series of valleys and peaks (and slopes) was detected. Then the median of the peak-to-peak gradient values was calculated and used to define the  $T_2$  threshold. This scaled threshold ( $T_{K1} = 0.25$ ) was used in the second pass of the algorithm to detect final peaks of each signal.

Another pulse rate estimator is based on the number of zero-crossings ( $nZC$ ) of the filtered signal and it estimates the frequency as:

$$f_{zC} = \frac{nZC(\bar{s}_{fn}(t)) - 1}{2} * \frac{fs}{N}. \quad (9)$$

Similarly, the next pulse rate estimator is based on the number of peaks in the filtered signal and estimates the frequency as:

$$f_{p1} = \frac{nPK(\bar{s}_{fn}(t))}{2} * \frac{fs}{N}. \quad (10)$$

Finally, the last estimator is based also on detected peaks and calculation of the average value of peak-to-peak time period  $T_{avgPP}$  ( $f_{p2} = 1/T_{avgPP}$ ). The peaks are detected using the previously described method. The calculated frequencies are multiplied by 60 to obtain the final units of beats per minute (bpm).

The frequency values can be calculated practically for all signals obtained from the recorded videos (*e.g.* for the face ROI, for the background ROI, *etc.*). However, it is important to additionally describe how good is a signal or how similar is the measured signal to the a priori known PPG signal characteristics. Based on the nature of the PPG signal it can be assumed, that the expected signal is periodical and, after low pass filtration (*e.g.* removing average), similar to the sine function. It is expected that either the dominating, fundamental frequency in the frequency domain or the periodical character of the signal in the time domain will be found. Therefore, 5 signal parameters are used to characterize true PPG signals (face



ROI, forehead ROI) and other signals (hairs ROI, background ROI): the normalized peak energy (in the amplitude spectrum), the selective SNR (in the amplitude spectrum), mobility (in the time domain), complexity (in the time domain), and *Spectral Purity Index* (in the time domain).

The normalized peak energy in the frequency domain is calculated as the squared value of the highest peak to the sum of squared other values, scaled by the number of samples. The selective SNR measure is calculated similarly. However, additionally, the peaks – close to the highest peak – whose amplitude is at least 0.4 of the highest peak value, are added in the nominator. The high peaks represent frequencies of the *PPG signal* (present due to the spectral leakage) while the other frequencies are treated as *interferences*.

Very interesting signal measures, indicating some similarities to the sine function, are parameters calculated using the spectral moments: the Hijorth parameters (mobility, complexity) and the *Spectral Purity Index*. For the sampled signal  $s(t_k)$  of a continuous signal  $s(t)$  (the sampling period  $T_s$ ,  $t_i = i \cdot T_s$ ) the spectral moments can be defined in reference to the expected value of  $s(t)$  and its derivatives [21]:

$$\bar{m}_0 = 2\pi E[s^2(t)], \quad \bar{m}_2 = 2\pi T_s^2 E\left[\left(\frac{ds(t)}{dt}\right)^2\right], \quad \bar{m}_4 = 2\pi T_s^4 E\left[\left(\frac{d^2s(t)}{dt^2}\right)^2\right]. \quad (11)$$

In this work the filtered normalized signals were used. Approximating the derivatives as:

$$\bar{s}_{fn}^{(1)}(t_i) = \bar{s}_{fn}(t_i) - \bar{s}_{fn}(t_{i-1}), \quad (12)$$

$$\bar{s}_{fn}^{(2)}(t_i) = \bar{s}_{fn}(t_{i+1}) - 2\bar{s}_{fn}(t_i) + \bar{s}_{fn}(t_{i-1}), \quad (13)$$

the spectral moments can be estimated as:

$$\bar{m}_o \approx \frac{2\pi}{K} \sum_{i=1}^K (\bar{s}_{fn}^{(o/2)}(t_i))^2, \quad \text{for } o = 0, 2, 4. \quad (14)$$

Then two Hijorth parameters and the spectral purity measure can be defined as:

$$\text{mobility} = \sqrt{\frac{\bar{m}_2}{\bar{m}_0}}, \quad \text{complexity} = \sqrt{\frac{\bar{m}_4}{\bar{m}_2} - \frac{\bar{m}_2}{\bar{m}_0}}, \quad \text{SPI} = \frac{\bar{m}_2^2}{\bar{m}_0 \bar{m}_4}. \quad (15)$$

In the experiments all parameters were calculated for all videos and for all ROIs: face, forehead, hairs, background. The *true* PPG signals were marked as 1 (*true PPG signal*), whereas others were marked as “0” (*other, noise*). Data mining algorithms were used to characterize the discriminative properties (1 – PPG signal, 0 – other) of these measures.

### 3. Results

#### 3.1. Role of YUV color model

In Fig. 2 an example of preprocessed signals for RGB and YUV color models are presented. Additionally, the results of using PCA and ICA algorithms applied to  $R$ ,  $G$ , and  $B$  signals are presented.

After the qualitative analysis of the calculated signals in the entire experiment it can be concluded that  $Y$ ,  $U$ ,  $V$  signals are very similar to those obtained after PCA analysis. The most interesting finding is that the signals very similar to the  $V$  signal are present in ICA results.

The quantitative analysis was performed by calculating cross-spectral coherence. First, the ADF test, performed for all  $Y$ ,  $U$ ,  $V$  signals from the experiment and for PCA signals, proved

that the signals were stationary (the highest – the worst – test value was  $v = -7.40$ , with the critical value  $CV(95\% \text{ level}) = -2.88$ ;  $v < CV(95\%)$ ). High pass filtration was first used to remove the baseline. The total average value of the  $AUCY(f_1, f_2)$  was  $0.73 \pm 0.2$ . It was calculated in the range of frequencies from 0.7 Hz to 7 Hz and for comparison of  $Y, U, V$  signals to corresponding (manually chosen) PCA signals. For a narrower bandwidth (0.7 Hz to 4 Hz, which is specific for pulse wave frequencies, 42 bpm – 240 bpm) the results were by about 10% better. It shows a high similarity of signals in terms of characteristic frequencies. Similar analysis performed between  $Y, U, V$  signals and signals obtained for two randomly chosen ICA algorithms showed a little bit worse results (what is expected, since ICA may produce independent signals – not only uncorrelated signals). For Jade ICA the cross-spectral coherence measure (0.7–7 Hz) was  $0.68 \pm 0.20$ , while for Kernel ICA –  $0.64 \pm 0.23$ . In some cases the results obtained for ICA algorithms applied to  $R, G, B$  signals were worse for the needs of pulse analysis than using  $Y, U, V$  or PCA (Fig. 3).

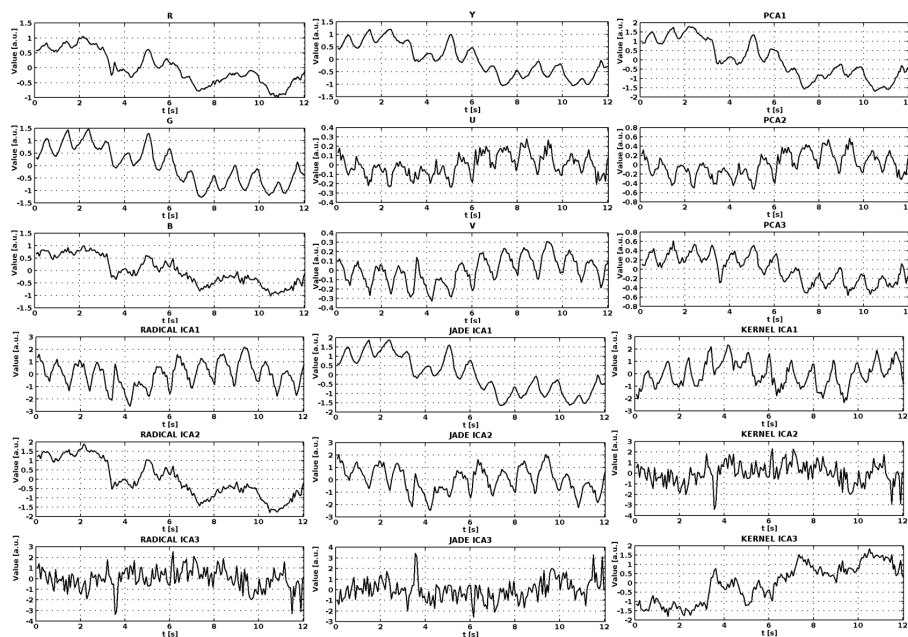


Fig. 2. An example of pre-processed signals obtained for: (top, left): a) RGB color model; b) YUV color model; c) PCA components from  $R, G, B$  signals; (bottom, left): RADICAL ICA, JADE ICA, KERNEL ICA.

### 3.2. Pulse rate estimation

The Mean Square Error (MSE) was calculated for all 5 pulse rate estimation methods and all (240) signals. The results are presented in Table 1.

The mean difference values between estimated and reference pulse rates obtained for the Forehead ROI (V signals) for best estimators were (mean  $\pm$  std. dev.):  $ePR_{sp}$ :  $1.85 \pm 1.52$ ,  $ePR_{ac}$ :  $1.81 \pm 1.55$ ,  $ePR_{p2}$ :  $2.67 \pm 2.87$  bpm. Examples of pulse rate estimations for 4 methods are presented in Fig. 4. Additionally, the minimum pulse rate values measured with the reference pulse oximeter for each subject are shown as (PR-REFERENCE).





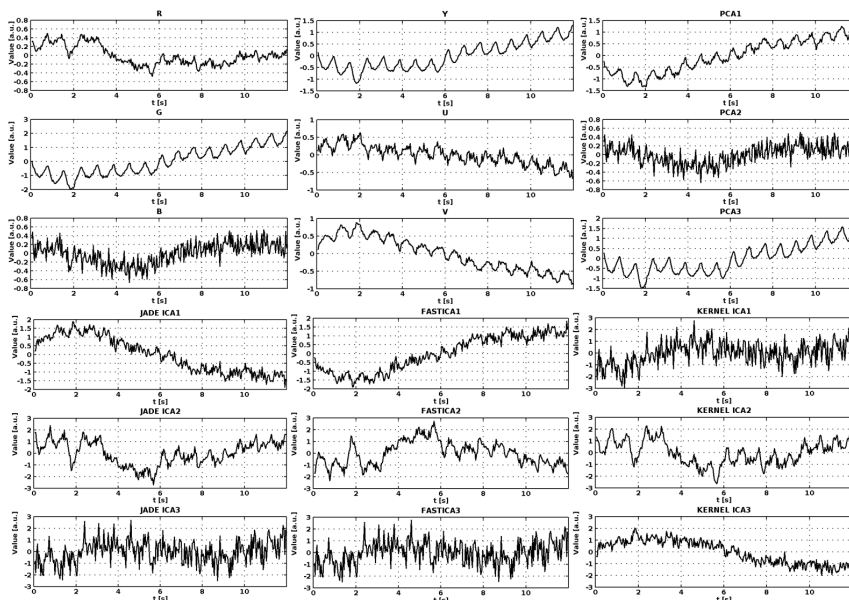


Fig. 3. An example of pre-processed signals obtained for: (top, from left): a) RGB color model; b) YUV color model; c) PCA components from  $R, G, B$  signals; (bottom, from left): JADE ICA, Fast ICA, and KERNEL ICA.

Table 1. The calculated values of MSE for different methods and signals.

|                 | $ePR_{sp}$ | $ePR_{ac}$ | $ePR_{zc}$ | $ePR_{p1}$ | $ePR_{p2}$ |
|-----------------|------------|------------|------------|------------|------------|
| Y, Face ROI     | 0.973      | 0.649      | 0.731      | 1.076      | 1.127      |
| V, Face ROI     | 0.571      | 0.468      | 0.787      | 2.034      | 0.877      |
| Y, Forehead ROI | 0.722      | 0.543      | 0.741      | 1.223      | 1.379      |
| V, Forehead ROI | 0.497      | 0.452      | 0.726      | 2.006      | 0.499      |

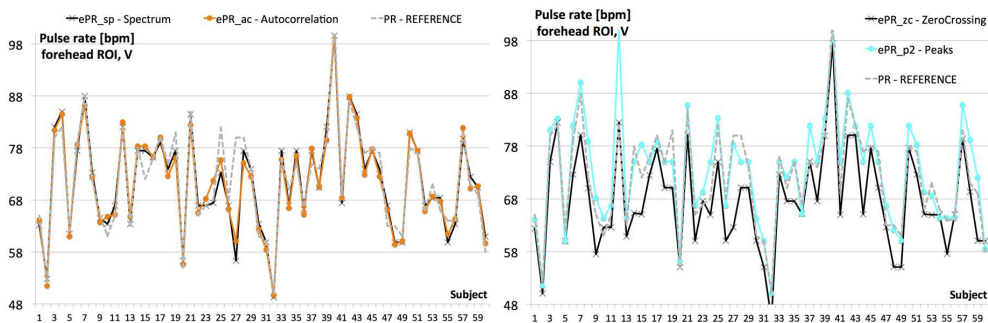


Fig. 4. The results of pulse rate estimations with different methods for the forehead ROI and the V signal.

The best results were obtained for  $V$  signals and for the  $ePR_{ac}$  estimator. Additionally, 5 parameters used to discriminate between the potential PPG signal and other signals were evaluated. The analyzed parameters were ranked using the information gain method. The following results were obtained: complexity (0.772, best), the selective SNR (0.736), the normalized peak energy (0.681), SPI (0.574), and mobility (0.242). The accuracy of 96.67% was achieved using all 5 attributes for training the J48 decision tree (Fig. 5b). The recall value



was 0.975 and precision 0.949. Using only complexity instead of all parameters led to a very small reduction of accuracy (95.56%) and recall (0.956, precision: 0.956). Fig. 5a presents distribution of complexity values calculated for all subjects and ROIs.

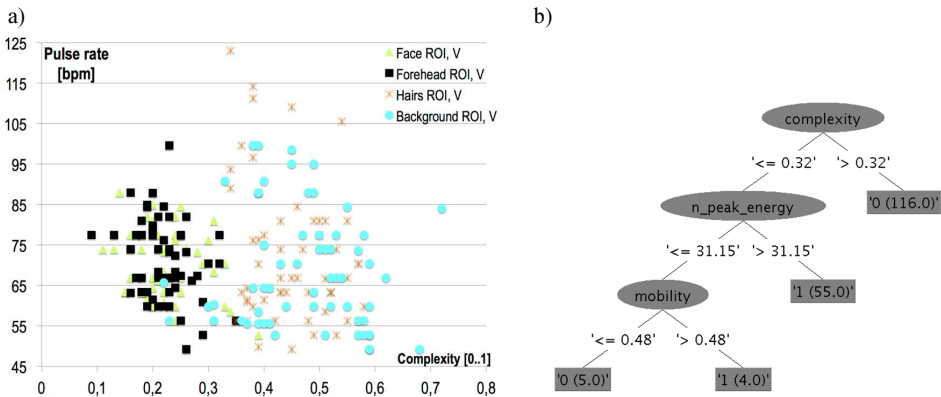


Fig. 5. a) Distribution of “complexity” values calculated for all 4 ROIs for all subjects; b) The decision tree trained using the J48 algorithm and all parameters.

#### 4. Discussion and conclusion

The first objective of this work was to verify the possible use of signals obtained for the YUV color model to estimate the pulse rate. The results of the qualitative and quantitative (spectral coherence) analyses proved that the signals obtained for  $Y$ ,  $U$ ,  $V$  components are spectrally very similar to the results of PCA analysis applied to  $R$ ,  $G$ , and  $B$  signals. The experimental results agree with the nature of relation between RGB and YUV models, which is characterized by a linear combination. The disadvantage of PCA analysis for pulse estimation is the property, that it is not possible to know which principal component will represent the PPG-related signal. The same disadvantage is observed for ICA methods. Four ICA algorithms were evaluated. In some cases the results of PCA were almost equivalent to those of ICA because all correlations in the data were mainly limited to second-order correlations. For some other cases the results of ICA analysis did not produce better results than those obtained with PCA. Concluding, the direct use of YUV model for pulse rate estimation is recommended for processing of uncompressed video sequences recorded from the short distances from subjects.

The second goal of the research was to examine what is the best pulse estimation method for short video sequences. After analysis of 240 signals ( $2 \times 60$  for the face ROI and  $2 \times 60$  for the Forehead ROI) the best results were obtained for the method based on the use of autocorrelation method. The value of calculated MSE was best for all analyzed ROIs and all processed color components ( $Y$  and  $V$ ). Additionally, it was proved that the best results were obtained for analysis of  $V$  signals, recorded from the Forehead ROI. It was observed that  $Y$  signals were more sensitive to changes of the scene illumination (many reflections from moving objects in the camera field of view). Additionally, it was observed that changing a scene content (e.g. moving people near the subject) could lead to automatic corrections of camera exposure, which introduces additional artifacts. Other artifacts were observed for face ROIs due to eye blinks. Therefore, the results for the Forehead ROI were slightly better than those for the Face ROI. However, sometimes it is not possible to record a signal from a subject’s forehead (e.g. covered

by hairs) and in such cases the face ROI could be used. It is also important to underline that measurements of the reference pulse were performed using a pulse oximeter, which always calculates an average value of the pulse rate. Typically, the number of samples taken between two heartbeats is calculated (e.g.  $T_s = 1\mu\text{s}$ ) producing an instantaneous beat-to-beat pulse rate value. Then, after  $K$  measurements the average value is calculated and presented to the user. Typically,  $K$  is equal to 5 [22], 6 or 12 [23] measurements or seconds. Cardiologists, in their routine practice, also analyze ECG using 6 s or 12 s wide windows to calculate the heart rate (the heart rate – the number of QRS complexes within the time window multiplied accordingly by 10 or 5). Since the pulse/heart rate values are calculated using averages for different and short time windows, the measured values can slightly vary in time and between devices. For example, in [23] it was shown that the standard error of the mean for pulse rate measurements using pulse oximeters was 2.1–2.5 bpm. The obtained MSE values for the presented study were lower than 2, but in some individual cases the calculated PR values were exceptionally higher than for the reference (e.g. about 17 bpm for  $ePR_{p2}$  method, for subject 12). In our experiments we used 12s time windows (last observed values) for data analysis and the last value of the pulse rate presented by the reference pulse oximeter. It can be concluded that short time signal windows can be used to reliably estimate the pulse rate values from  $V$  signals recorded from the Forehead ROI.

The last objective of the study was to analyze how potential PPG-signals can be distinguished from other (e.g. background) signals. Since the ideal method does not exist we based on prior knowledge about PPG signals: the VPG signal is recorded for a face, the filtered VPG signal is periodical (similar to the sine function). Since the first property is relatively simple to verify (e.g. using observation, face detection algorithms, etc.) we focused on the second property. We examined 5 different parameters characterizing the recorded true (face, forehead) and false (hairs, background) VPG-signals. From the set of true signals we eliminated signals obtained for subject 27, since the VPG-signal was not observed in the face and forehead ROIs (high reflectance artifacts, which led to camera exposure fluctuations). The results of analysis proved that the *complexity* parameter proposed by Hijorth, enabled to distinguish true and false highly periodical signals (assumed to be VPG signals). Such methods are important, since even for *false* VPG signals (e.g. hairs) the pulse rate values can be wrongly calculated (Fig. 5a). An attractive property of the used parameters is that it is possible to efficiently compute them in the time domain. Low computational requirements are especially important for implementation of algorithms in mobile platforms (e.g. smart glasses). Additionally, the value of complexity measure tends to 0 if a signal is more similar to the sine function. Similarly, the SPI value indicates how well the signal is described by a single frequency (how *pure* is the spectrum). For the sine function the SPI value is equal to 1. The presented parameters have been previously found very useful for analysis of EEG and ECG signals (e.g. [24, 25]). In the future it would be also interesting to verify DFT-based methods that can improve the estimation accuracy in the case of a low signal-to-noise ratio (e.g. [26]) and to apply the Kalman filter [27].

Finally, it can be concluded that using native YUV video formats and analyzing  $V$  signals recorded for the Forehead ROI in the time domain, it is possible to accurately and efficiently measure the pulse rates. It is also possible to efficiently verify the reliability (or reduce the uncertainty) of a pulse-wave signal. Effectiveness of the procedure is important for the future use of this methodology for smart glasses or other portable devices that can be used for screening purposes and remote diagnostics. For example, the pulse rate play an important role in the *Emergency Severity Index* (ESI) triage and a person with a heart rate above 100 bpm is considered to be in the “danger zone” [28]. In this context, the pulse rate estimation error about 2 bpm is likely to be acceptable.



## Acknowledgements

This work has been partially supported by NCBiR, FWF, SNSF, ANR and FNR within the framework of the ERA-NET CHIST-ERA II, European project *eGLASSES – The interactive eyeglasses for mobile, perceptual computing* and by Statutory Funds of Electronics, Telecommunications and Informatics Faculty, Gdansk University of Technology.

## References

- [1] Humphreys, K., Markham, Ch., Ward, T.E. (2005). A CMOS camera-based system for clinical photoplethysmographic applications. *Proc. SPIE*, 5823, 88–95.
- [2] Verkruysse, W., Svaasand, L.O., Nelson, J.S. (2008). Remote plethysmographic imaging using ambient light. *Optics Express*, 16(26), 21434–21445.
- [3] 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, 10762–10774.
- [4] Christinaki, E., Giannakakis, G., Chiarugi, F., Pediaditis, M., Iatraki, G., Manousos, D., Marias, K., Tsiknakis, M. (2014). Comparison of blind source separation algorithms for optical heart rate monitoring. *Proc. of Mobihealth Conference*, 339–342.
- [5] Mannapperuma, K., Holton, B.D., Lesniewski, P.J., Thomas, J.C. (2015). Performance limits of ICA-based heart rate identification techniques in imaging photoplethysmography. *Physiol Meas.*, 36(1), 67–83.
- [6] Lewandowska, M., Rumiński, J., Kocejko, T., Nowak, J. (2011). Measuring pulse rate with a webcam – A non-contact method for evaluating cardiac activity. *Proc. of FedCSIS Conference, IEEE Xplore*, 405–410.
- [7] Smart Glasses Portal. (2015). <http://www.smartglassesnews.org>
- [8] Evena Medical. (2014). <http://evenamed.com/~even5672/products/glasses>
- [9] Rumiński, J., Bujnowski, A., Wtorek, J., Andrushevich, A., Biallas, M., Kistler, R. (2014). Interactions with recognized objects. *Proc. of HSI 2014 Conference, IEEE Xplore*, 101–105.
- [10] Rumiński, J., Smiatacz, M., Bujnowski, A., Andrushevich, A., Biallas, M., Kistler, R. (2015). Interactions with recognized patients using smart glasses. *Proc. of HSI 2015 Conference, IEEE Xplore*, 187–194.
- [11] Czuszyński, K., Rumiński, J. (2014). Interaction with medical data using QR-codes. *Proc. of 7th HSI 2014 Conference, IEEE Xplore*, 182–187.
- [12] Kwasniewska, A., Klimiuk-Myszk, J., Rumiński, J., Forrier, J., Martin, B., Pecci, I. (2015). Quality of graphical markers for the needs of eyewear devices. *Proc. of 8th HSI 2015 Conference*, 388–395.
- [13] Fuller, W.A. (1996). *Introduction to Statistical Time Series*. UK: John Wiley & Sons, Inc.
- [14] Gävert, H., Hurri, J., Särelä, J., Hyvärinen, A. (2005). FastICA, <http://research.ics.aalto.fi/ica/fastica/about.shtml>
- [15] Cardoso, J.F. (2005). Jade 1.8, <http://perso.telecom-paristech.fr/~cardoso/guidesepsou.html>
- [16] Bach, F. (2002). Kernel ICA 1.2, <http://www.di.ens.fr/~fbach/kernel-ica/>
- [17] Learned-Miller, E.G. (2004). RADICAL 1.2, <https://people.cs.umass.edu/~elm/ICA/>
- [18] Kay, S.M. (1988). *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall.
- [19] Augustyniak, P. (2013). Coherence-based measure of instantaneous ECG noise. *Computing in Cardiology Conference (CinC)*, 787–790.
- [20] Clifford, G.D. (2006). ECG statistics, noise, artefacts and missing data. Chapter 3 in: Clifford G.D., Azuaje, F., McSharry, P.E. *Advanced Methods and Tools for ECG Analysis*, Artech House Publishing.
- [21] Sörnmo, L., Laguna, P. (2005). *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Academic Press.
- [22] Lopez, S. (2012). Pulse Oximeter Fundamentals and Design, Freescale Semiconductor, Application Note, Document Number: AN4327 Rev. 2.

- [23] Iyriboz, Y., Powers, S., Morrow, J., Ayers, D., Landry G. (1991). Accuracy of pulse oximeters in estimating heart rate at rest and during exercise. *British Journal of Sports Medicine*, 25(3), 162–164.
- [24] Oh, S.H., Lee, Y.R., Kim, H.N. (2014). A Novel EEG Feature Extraction Method Using Hjorth Parameter. *International Journal of Electronics and Electrical Engineering*, 2(2), 106–110.
- [25] Deshmane, A.V. (2009). *False arrhythmia alarm suppression using ECG, ABP, and photoplethysmogram*. MS Thesis, MIT, USA.
- [26] Xin, L., Yongfeng, R., Chengqun, C., Wei, W.F. (2015). Accurate Frequency Estimation Based On Three-Parameter Sine-Fitting With Three FFT Samples. *Metrol. Meas. Syst.*, 22(3), 403–416.
- [27] Smieszek, M., Dobrzanska, M. (2015). Application Of Kalman Filter In Navigation Process Of Automated Guided Vehicles. *Metrol. Meas. Syst.*, 22(3), 443–454.
- [28] Wuerz, R.C., Travers, D., Gilboy, N., Eitel, D.R., Rosenau, A., Yazhari, R. (2001). Implementation and refinement of the emergency severity index. *Acad. Emerg. Med.*, 8(2), 170–176.