Remote Photoplethysmography with Constrained ICA using Autocorrelation as a periodicity measure

Richard Macwan, Yannick Benezeth, Alamin Mansouri
Le2i UMR6306, CNRS, Arts et Métiers
Univ. Bourgogne Franche-Comté

Abstract

Remote photoplethysmography(rPPG) is being increasingly used to measure heart rate from recorded or live videos. The rhythmic flow of arterial blood, referred to as the blood volume pulse, results in periodic variations in the skin color which are then quantified into a temporal signal for analysis.

We present a novel method for measuring remote photoplethysmography(rPPG) signals using Constrained Independent Component Analysis(cICA). We provide autocorrelation as an a priori information for cICA to extract the most prominent rppg signal. CICA with autocorrelation showed improved performance over traditional ICA in terms of convergence speed and accuracy with two different in-house video databases.

1. Introduction

Photoelectric plethysmography or photoplethysmography(PPG) was first introduced in 1937 by Hertzman where the variations in the light absorption of human skin were measured by a photoelectric cell[7] placed under a finger illuminated by a light source placed above it. Since then PPG has been used widely because of its ease of usage,low cost and non-invasiveness. This non-invasiveness has, however, been superseded by that of remote photoplethysmography, henceforth referred to as rPPG, which aims at measuring the same parameters, but *sans contact*.

Verkrussysse [18] demonstrated the extraction of remote PPG signals using videos from a simple consumer level camera and that the strongest photoplethysmographic signal was manifested in the G channel of the RGB temporal traces. RGB temporal traces are generated by frame wise quantification, e.g. spatial averaging of skin pixels from the face, and concatenating them. Current research focuses on extracting robust rPPG signals from simple web cameras for which Blind Source Separation(BSS) using Independent Component Analysis(ICA) has been used in many different works [4, 14, 15, 16].

ICA is a statistical technique for decomposing a multivariate signal into constituent signals assuming that the input signals are uncorrelated [9]. The problem of rPPG measurements is posed as a signal separation problem where the rhythmic cardiac pulse, appearing as variations in skin color, is assumed to be linearly separated and mixed into the temporal traces of color data from cameras. If the time varying color traces for different channels are represented as $\mathbf{s} = (s_1, s_2, ..., s_n)^T$, which is an instantaneous linear mixture of the original independent signals denoted as $\mathbf{c} = (c_1, c_2, ..., c_m)^T$, then the process of mixing can be formulated as $\mathbf{s} = \mathbf{Ac}$, where the mixing matrix $\mathbf{A}_{n\times m}$ represents the linear memoryless mixing of the channels. The goal of ICA, then, is to estimate the demixing matrix $\mathbf{W_{m\times n}}$ to recover all the ICs from the observed signal with minimal knowledge of $\mathbf{A}$ and $\mathbf{c}$. The recovered signal $\mathbf{y} = (y_1, y_2, ..., y_m)^T$, is given by $\mathbf{y} = \mathbf{Wx}$ [13].

This formulation of ICA is based on linear mixing, and thus suffers from two unavoidable ambiguities [2, 6]. First, the order of the independent components is indeterminable. A different permutation of the columns of $\mathbf{W}$ will give the same independent components. Second, the exact amplitude and sign of the independent components is also indeterminable. In spite of these limitations, ICA is being frequently used for rPPG measurements.

Furthermore, we know that the most periodic signal that might be embedded in the RGB temporal traces must correspond to the blood volume pulse. Additionally, we only require one component, viz., the rPPG pulse from the mixture of the temporal traces. This requirement is quite common in various applications, for example, the On-Off simulation scheme of an fMRI experiments [17]. Consequently, it would be advantageous to provide this a priori knowledge to the component extraction algorithm.

The purpose of Constrained Independent Component Analysis(cICA) is just this, to provide a systematic and flexible method to incorporate more assumptions and prior information into the contrast function to make the ICA problem a better-posed problem. Specifically, cICA avoids the ambiguities of ICA by directly converging to the best independent component. In this paper, we use the periodicity of the rPPG signal as an a priori information to determine the component that is the most periodic one. The optimization is steered in the direction of choosing the component with the highest periodicity which represents the actual blood volume pulse. The cICA algo-

rithm is detailed in section 3.2.

2. Previous Work

One of the first works that used ICA for rPPG measurements comprised of using RGB temporal traces from a simple web camera to extract the cardiac pulse, albeit with limited success under the presence of movement artifacts. Usage of more color channels using a five band camera (RGBCO) with ICA was also investigated [14]. ICA's known caveat of the indeterminacy of the order of the estimated components calls for a heuristic to choose the correct component. Where Poh et. al. [15] simply selected the second obtained component after applying ICA on RGB temporal traces , McDuff et. al. [14] chose the signal with the peak of greatest power of the normalized FFT spectrum between 40 and 180 bpm.

De Haan et. al. [4] introduced chrominance-based methods where two orthogonal chrominance signals were built from the RGB traces in addition to using skin-tone standardization to compensate for illumination variation of different skin colors. They further improved upon the chrominance based methods to show that the different absorption spectra of arterial blood happen along a specific vector in a normalized RGB space, termed as the Blood Volume Pulse vector [5].

In a related work, Wang et. al. [19] attempted to extracted the rPPG signal by constructing pixel based rPPG sensors to estimate a robust pulse signal using motion compensated pixel-to-pixel pulse extraction based on optical flow vectors.

Incorporating a priori information to guide the optimization process is an interesting approach in signal separation. Lu. et. al. [11] have used an existing reference signal to guide the separation process by using the method of Lagrange multipliers where the distance between the reference signal and the estimated signal is taken as the constraint to be minimized. However, such a reference signal is not always present, or at times is difficult to synthesize. In rPPG measurements, a PPG signal is such an example the synthesis of which depends critically on the required frequency, more so than on the actual shape of the signal.

To overcome this limitation, we use autocorrelation as the a priori information for guiding the cICA separation algorithm which then chooses the most periodic component representing the blood volume pulse. The algorithm with the workflow of the whole experiment is presented in the next section.

3. Methods

The workflow of the experimental procedure as depicted in figure 1 is presented here. Temporal RGB traces, $\mathbf{x} = [x_1\, x_2\, x_3]^T$ where each x_m, $m \in [1...3]$, corresponds to a temporal trace of size N of each channel, generated by spatial averaging of the pixels (either entire image, face-cropped or skin-segmented) were fed to the cICA algorithm.

3.1. Preprocessing for cICA

The initial preprocessing steps required for constrained ICA are the same as those for ICA which are generally recommended for simplifying the calculations for obtaining the independent components. After normalizing the RGB traces, *centering* was performed so that the obtained signal $\mathbf{y}$ in $\mathbf{y} = \mathbf{W}\mathbf{x}$ is zero-mean. After centering, *whitening* was performed to ensure that the components were uncorrelated and their variances equal to unity.

The RGB traces were then detrended using a smoothness priors approach proposed by Karjalainen et. al. [10] to remove low frequency trends in the signal. Finally, a fourth order butterworth bandpass filter was used to filter the signal within limits of human heartrate between .3 and 3 Hz. The RGB traces are now ready to be fed into the constrained ICA algorithm for obtaining the rPPG pulse signal.

3.2. RPPG Pulse Extraction using cICA

Biomedical signals like ECG and PPG signals from finger sensors are typically known to be periodic or quasi-periodic. In this work, instead of using a reference signal typically used with the cICA algorithm [11], we exploit the inherent periodicity property of these biomedical signals, thereby guiding the separation process to choose the component with the highest periodicity. We use autocorrelation as a periodicity measure which, owing to the formulation of lagrange multipliers, requires the calculation of the first and second derivatives of the autocorrelation with respect to the weighting matrix $\mathbf{w}$.

3.2.1. Autocorrelation as a periodicity measure

Autocorrelation is the correlation of a signal with itself at different lag times provided it is sampled at a sufficiently high frequency. For a time series signal $\mathbf{y} = [y_1\, y_2 ... y_N]$ of N elements, it's discrete autocorrelation r_k at lags $k \in [-(N-1), \cdots, N-1]$ is given by

$$r_k = \sum_{j=0}^{N-1} y_j \odot \overset{k}{y_j} \tag{1}$$

where $\overset{k}{y_j}$ is the j^{th} element of the signal $\mathbf{y}$ lagged(or led if $k < 0$) by k units and padded with zeroes to the left (or right if $k < 0$) and $\odot$ is the element-wise multiplication operator. A periodic signal typically has a higher correlation with itself compared to a non-periodic one. This high correlation can be quantified as the mean of the squared autocorrelation of the signal and consequently can be used as a measure of the periodicity of a signal. Figure 2 depicts the high correlation of a periodic sinusoid compared to that of a uniform random signal with the mean of

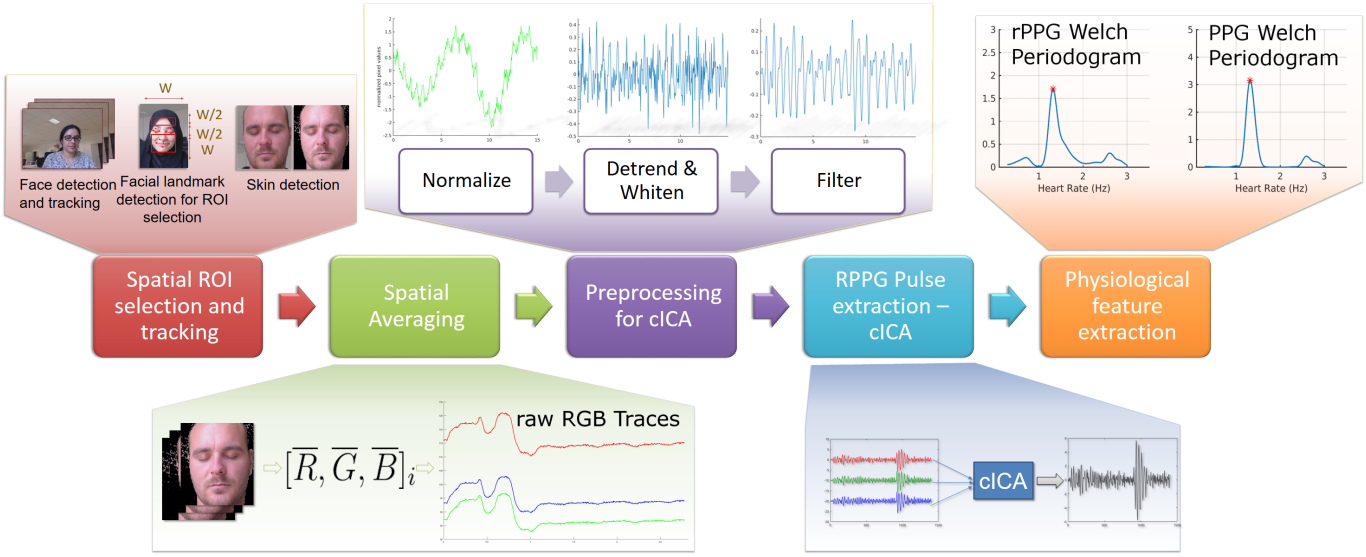
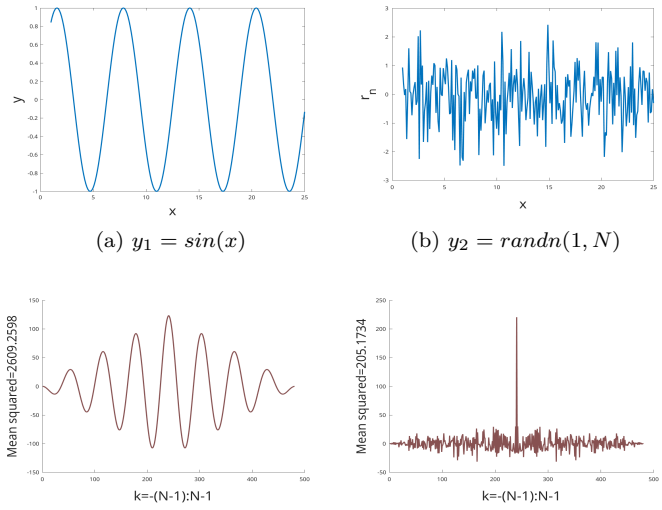


Figure 1. Methodology

the squared autocorrelation is much higher than that of the random signal.



(a) $y_1 = sin(x)$ (b) $y_2 = randn(1, N)$

(c) Autocorrelation of y_1 (d) Autocorrelation of y_2

Figure 2. Autocorrelation of a sinusoid vs a random signal

To aid the use of autocorrelation as a periodicity measure and simplify its computation, two modifications need to be made. First, since the autocorrelation is symmetric, we only compute the correlation for lags $k \in [0, \cdots, N-1]$. Second, since the correlation at lag 0 will always be high, we set the autocorrelation to *0* at lag $k = 0$. Thus, the autocorrelation is given by $\mathbf{r} = [r_1\, r_2 ... \, r_N]$ comprising of N values given by equation 1 and $r_1 = 0$. Keeping in mind that r_k is a scalar, equation 1 can be rewritten in matrix notation as

$$r_k = \mathbf{y}[\overset{k}{\mathbf{y}}]^T = \overset{k}{\mathbf{y}}\mathbf{y}^T \tag{2}$$

where $\overset{k}{\mathbf{y}}$ is again the signal $\mathbf{y}$ lagged by k units. Furthermore, to simplify the derivation of the autocorrelation, $\overset{k}{\mathbf{y}}$ can be rewritten as yT_k where T_k is a toeplitz-like matrix that incorporates the lagging at lag k and padding with zeroes of the signal and is given by

$$T_k = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} 0_{N-k,k} & I_{N-k} \\ 0_{k,k} & 0_{k,N-k} \end{bmatrix} \tag{3}$$

T_k is an $N \times N$ matrix composed of the first $N-k$ rows made up of $(N-k) \times k$ zeroes and an identity matrix of size $N-k$. Thus, r_k becomes

$$r_k = \mathbf{y}T_k\mathbf{y}^T \tag{4}$$

making its differential with respect to $\mathbf{y}$ easier to calculate. This autocorrelation is then used to guide the optimization process of constrained ICA to converge to the component with the highest periodicity, which is presented in the next subsection.

3.2.2. Constrained ICA

A generic contrast function for ICA as defined by [9], is the negentropy function given by $J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y})$ where $H(.)$ is the differential entropy and $\mathbf{y}_{gauss}$is a random variable with a variance equal to that of the output signal $\mathbf{y}$. In FastICA, an approximation of the negentropy was introduced for more reliability and flexibility given by

$$J(\mathbf{y}) \approx \rho[E\{G(\mathbf{y})\} - E\{G(v)\}]^2 \tag{5}$$

where ρ is a positive constant, v is a zero mean, unit variance Gaussian and G(.) can be any non-quadratic function. As suggested by [8] a good general purpose function is given by

$$G(y) = \frac{\log\cos(a_1 y)}{a_1} \tag{6}$$

with $1 < a_1 < 2$.

Constrained ICA aims to alleviate the issues of ICA with the help of Lagrange multiplier methods [11]. Lagrange multiplier methods [1] are a tool for performing constrained optimization problems following the general form

$$\text{minimize } f(\mathbf{X}), \text{ subject to } g(\mathbf{X}) \leq 0,\ h(\mathbf{X}) = 0 \quad (7)$$

where $f(\mathbf{X})$ is the objective function, $g(\mathbf{X})$ is a set of inequality constraints and $h(\mathbf{X})$ is a set of equality constraints.

The objective of obtaining the most periodic component using cICA can be fulfilled with the help of the inequality constraint

$$g(\mathbf{w}) = \epsilon(\mathbf{w}) - \zeta \leq 0 \quad (8)$$

where $\mathbf{w}$ represents a single demixing weight vector of size equal to the number of input channels. The optimum $\mathbf{w}$ then extracts the most periodic component using $\mathbf{y} = \mathbf{w}^{\mathrm{T}}\mathbf{x}$. $\epsilon(\mathbf{w})$ represents the closeness measure. Using average of squared autocorrelation as a closeness measure gives $g(\mathbf{w})$ as

$$g(\mathbf{w}) = \zeta - E\{\mathbf{r}^2\} \leq 0 \quad (9)$$

where ζ now denotes the threshold for the lower bound of the optimum autocorrelation. The details of the use of this constraint are presented in the next subsection.

The Augmented Lagrangian for cICA The general cICA problem is defined as [11]

$$\begin{aligned} \text{Maximize}: \quad J(\mathbf{y}) &= \rho[E\{G(\mathbf{w}^{\mathrm{T}}\mathbf{x})\} - E\{G(v)\}]^2, \\ \text{Subject to}: \quad g(\mathbf{w}) &\leq 0,\ h(\mathbf{w}) = E\{\mathbf{y^2}\} - 1 = 0 \end{aligned} \quad (10)$$

where $J(\mathbf{y})$ is the one-unit contrast function as defined in equation 5, $g(\mathbf{w})$ is the closeness constraint and $h(\mathbf{w})$ constrains the output $\mathbf{y}$ to have unit variance.

Finally, the augmented Lagrangian method was used owing to its wider applicability and improved stability [1]. The augmented Lagrangian for our formulation as adapted from [11] is given by

$$\begin{aligned} \mathcal{L}_1(\mathbf{w}, \mu, \lambda) &= J(\mathbf{y}) - \frac{1}{2\gamma}\left[\{[max\{0, \bar{g}(\mathbf{w})\}]^2 - \mu_i^2\}\right] \\ &\quad -\lambda h(\mathbf{w}) + \frac{1}{2}\gamma_1 \left\|h(\mathbf{w})\right\|^2 \end{aligned} \quad (11)$$

where $\bar{g}(\mathbf{w}) = \mu + \gamma g(\mathbf{w})$, μ and λ are the lagrange multipliers corresponding to $g(\mathbf{w})$ and $h(\mathbf{w})$ respectively.$\|.\|$ denotes the Euclidean norm and the term $\frac{1}{2}\gamma \left\|.\right\|^2$ is the penalty term that makes sure that the optimization problem is held at the condition of local convexity assumption:$\nabla^2_{xx}\mathcal{L} > 0$.

To find the maximum of $\mathcal{L}_1$ in equation 11 a Newton-like learning method can be used to iteratively adapt $\mathbf{w}$

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \eta(\mathcal{L}^{''}_{1_{\mathbf{w}_k^2}})^{-1}\mathcal{L}^{'}_{1_{\mathbf{w}_k}} \quad (12)$$

where k is the iteration index, η is the positive learning rate to avoid uncertainty in convergence and $\mathcal{L}^{'}_{1w}$ is the first derivative of $\mathcal{L}$ w.r.t $\mathbf{w}$ given by

$$\mathcal{L}^{'}_{1w} = \bar{\rho}E\{\mathbf{x}G^{'}_y(\mathbf{y})\} - \frac{1}{2}\mu E\{g^{'}(\mathbf{w})\} - \lambda E\{\mathbf{xy}\} \quad (13)$$

where $\bar{\rho} = \pm\rho$ depending on the sign of $E\{G(\mathbf{y})\} - E\{G(v)\}$, $G^{'}_y(\mathbf{y})$ and $g^{'}(\mathbf{w})$ are the first derivatives of $G(\mathbf{y})$ and $g(\mathbf{w})$ w.r.t $\mathbf{y}$ and $\mathbf{w}$ respectively. The Hessian $\mathcal{L}^{''}_{1_{\mathbf{w}_k^2}}$ in equation 12, is calculated as

$$\mathcal{L}^{''}_{1_{\mathbf{w}_k^2}} = \bar{\rho}\mathbf{R_{xx}}E\{G^{''}_y(\mathbf{y})\} - \frac{1}{2}\mu E\{g^{''}(\mathbf{w})\} - \lambda \quad (14)$$

the inversion of which is not problematic because $\mathbf{R_{xx}}$ being the covariance matrix of the whitened and centered signal $\mathbf{x}$ is an identity matrix. $G^{''}_{y^2}(\mathbf{y})$ and $g^{''}(\mathbf{w})$ are second order derivatives and $\mathcal{L}_{1_{w_k^2}}$ is of size $m \times m$. The first and second derivatives of $g(\mathbf{w})$ are not trivial and are presented in the appendix. Finally, the approximate Newton learning step is given by

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \eta\mathcal{L}^{'}_{1_{\mathbf{w}_k}}/\mathcal{L}^{''}_{1_{\mathbf{w}_k^2}} \quad (15)$$

The optimum multipliers μ_i^* and λ^* are obtained iteratively based on a gradient-ascent method [12]:

$$\mu_{k+1} = max\{0, \mu_k + \gamma g(\mathbf{w}_k)\}, \quad (16)$$

$$\lambda_{k+1} = \lambda_k + \gamma h(\mathbf{w}_k) \quad (17)$$

Following the above equations, the optimization procedure converges to the optimum point defined by the triplet (w^*, μ^*, λ^*) representing the tuned parameters and final weighting matrix $\mathbf{w}*$ which is then used to obtain the final rPPG signal.

3.3. Physiological feature extraction

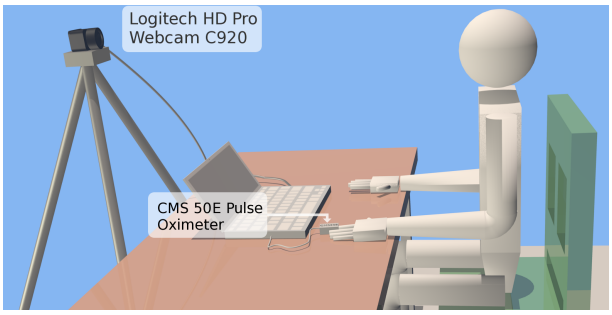
After the rPPG signal was obtained, the per window heart rate was calculated from the highest peak of the periodogram obtained using Welch's method. over a 15 second moving window using a step size of 1 second. ICA (and consequently cICA) is known to work much better with a signal of longer duration. However, to emulate a live scenario as closely as possible, steps 3.1 to 3.3 were performed for each window, using the weighting matrix $\mathbf{w}_k$ at step k as an initial estimate for calculation of $\mathbf{w}_{k+1}$ at the next step. Kalman filtering was used for predicting HRs for windows with $\delta = |HR_{rPPG} - HR_{PPG}| > 10\,bpm$. We present the results of the experiments of windowed and entire signal analysis in the next section.

4. Results and Discussion

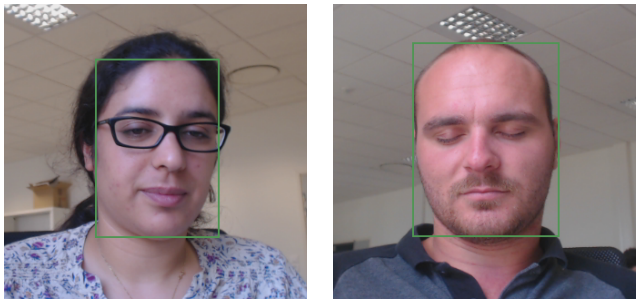
An in-house dataset comprising of 9 videos of around 90 seconds each was used to test the performance of the cICA method where the subjects were asked to sit in a relaxed pose in front of the camera. The video frames were obtained with a custom C++ application using a Logitech C920 web camera placed at a

distance of about 1m from the subject with a resolution of 640x480 in 8-bit uncompressed RGB format at 30 frames per second. A CMS50E transmissive pulse oximeter was used to obtain the ground truth PPG data. The experimental setup with images from two different videos showing the light conditions is depicted in figure 3.

To generate the temporal RGB traces, face detection was first performed using the Viola-Jones implementation provided by the computer vision toolbox of MATLAB. Corner detection in the detected face was performed for tracking to select the facial landmarks. Skin detection as formulated by Conaire et. al. [3] was then performed to select the candidate pixels which were then spatially averaged to obtain a triplet of RGB values per frame which were then concatenated to obtain the final RGB temporal traces.



(a) Experimental seup



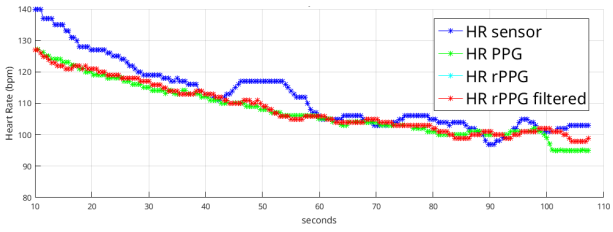
(b) Images from two videos

Figure 3. Experimental Setup

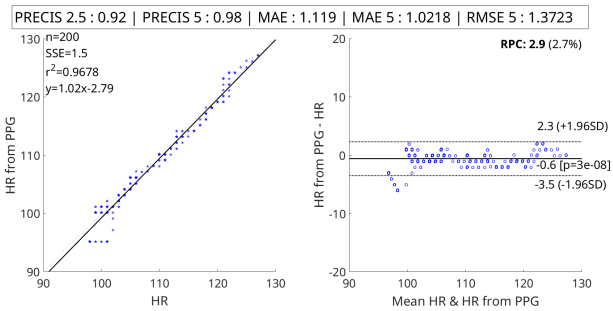
Figure 5c shows a typical rPPG signal and HR_{rPPG} vs HR_{PPG}. The slight phase shift in between the peaks of rPPG and PPG can be attributed to the pulse transit time: the time difference between the cardiac pulse wave reaching different peripheral organs. Figure 5a shows the Bland Altman (BA) analysis performed for the same video on each window. The window-wise difference between HR_rPPG and HR_PPG varies between +1.7 to -2 bpm and has a mean absolute error, $MAE = .63$.

One of the videos in the dataset was recorded after performing physical exercise to verify the drop in HR over time. The cICA method performs well in this case. The temporal comparison of HR from rPPG vs HR from PPG and its corresponding BA analysis is presented in figure 4.

Table 1 shows the accuracy comparisons between ICA and cICA using the mean absolute error (MAE) between HR_{rPPG} and HR_{PPG}. The windowed



(a) RPPG



(b) Bland Altman Analysis

Figure 4. cICA for after-exercise video

	Full	Face	Skin	Full	Face	Skin
	Entire signal			Windowed		
ICA	5.34	6.53	2.1	14.81	13.25	14.12
cICA	0.82	0.79	0.81	8.91	8.09	5.62

Table 1. Mean Absolute Error of ICA vs cICA

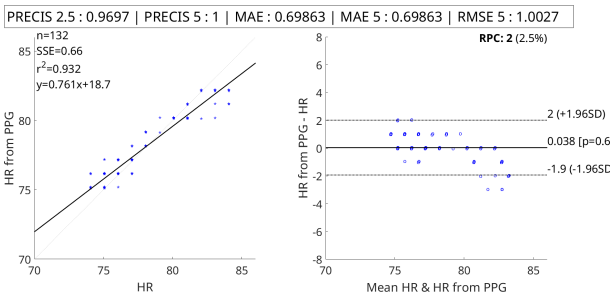
method is computationally more taxing but is more realistic and at understandably less accurate since we are operating only on a 15 second window of the signal which makes the component separation more difficult.

A global Bland Altman analysis was also performed using window-wise calculations from all the videos in each dataset for three different configurations: the entire image, face-cropped image and skin-segmented image. Figure 5b shows the BA analysis for the entire dataset using the entire image. The metrics PRECIS 2.5 and PRECIS 5 show the percentage of windows where $\delta = (HR_{rPPG} - HR_{PPG}) < 2.5$ and 5 bpm respectively. Correspondingly, the MAE 5 and the RMSE 5 metrics measure the mean absolute difference and the root mean squared difference respectively for windows with $\delta < 5$ bpm.

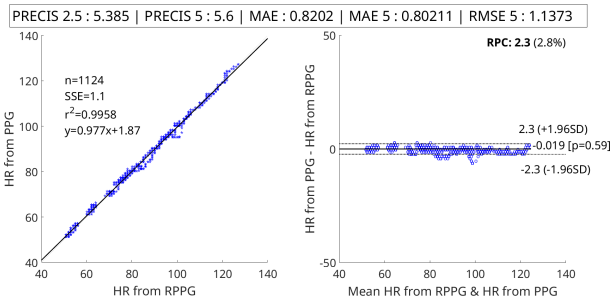
5. Conclusions and future work

In this paper we presented a novel semi blind source separation method for the application of rPPG measurements using autocorrelation as the constraint to guide the ICA separation process. The cICA using autocorrelation provides better result than simple ICA while removing the extra step for choosing the best component. The periodogram of the extracted signals was also constintently closer to that of the PPG.

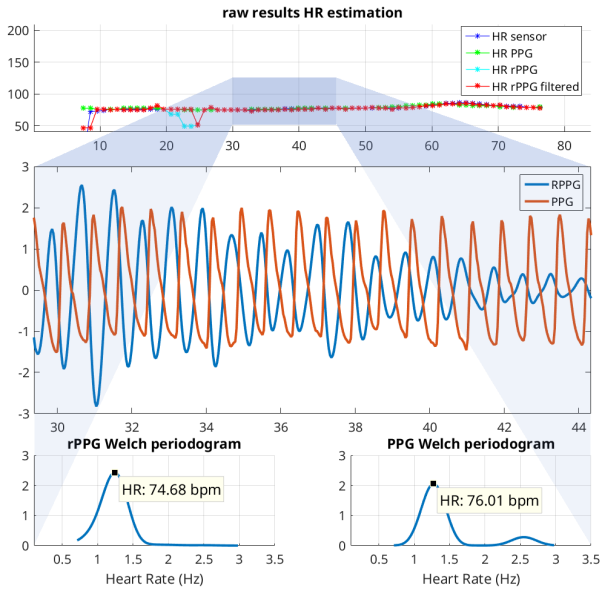
For improving accuracy, better face and skin detectors and trackers can be investigated. Also, the assumption that the most periodic component is the cardiac pulse signal does not hold in scenarios with periodic motion e.g. in fitness. The method can thus



(a) Bland Altman for video 5



(b) Bland Altman for dataset SIMPLE



(c) Extracted rPPG signal for a 15 sec window of video 5. HR_sensor is HR from the finger sensor. HR_rPPG_filtered is HR_rPPG after Kalman filtering

Figure 5. RPPG using cICA

benefit with motion compensation which itself is another subject for research. Finally, we average the entire image (or cropped or skin segmented version of it) to obtain a single value and thus loose any spatial information. Higher order analysis which preserves the spatial relationships between pixel neighborhoods can be an important avenue to look into.

References

[1] D. Bertsekas. Constrained optimization and Lagrange multiplier methods, 1982. 4

[2] P. Comon. Independent component analysis, A new concept? *Signal Processing*, 36(3):287–314, 1994. 1

[3] C. Ó. Conaire, N. E. O'Connor, and A. F. Smeaton. Detector adaptation by maximising agreement between independent data sources. *CVPR*, 2007. 5

[4] G. De Haan and V. Jeanne. Robust pulse rate from chrominance-based rPPG. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013. 1, 2

[5] G. de Haan and a. van Leest. Improved motion robustness of remote-PPG by using the blood volume pulse signature. *Physiological measurement*, 35(9):1913–1926, 2014. 2

[6] D. Djuwari, D. Kant Kumar, and M. Palaniswami. Limitations of ICA for Artefact Removal. *IEEE Medicine and Biology Society*, 5:4685–4688, 2005. 1

[7] A. B. Hertzman. Photoelectric Plethysmography of the Fingers and Toes in Man. *Experimental Biology and Medicine*, 37(3):529–534, 1937. 1

[8] A. Hyvärinen and E. Oja. A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation*, 9(7):1483–1492, 1997. 3

[9] A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4-5):411–430, 2000. 1, 3

[10] P. a. Karjalainen. An advanced detrending method with application to HRV analysis Mika P. Tarvainen, Perttu O. Ranta-aho, and Pasi A. Karjalainen. pages 1–4. 2

[11] W. Lu and J. C. Rajapakse. ICA with Reference. 2, 4

[12] W. Lu and J. C. Rajapakse. Rajapakse, âĂIJConstrained independent component analysis. 10:570–576, 2000. 4

[13] W. Lu and J. C. Rajapakse. Approach and applications of constrained ICA. *IEEE Transactions on Neural Networks*, 16(1):203–212, 2005. 1

[14] D. McDuff, S. Gontarek, and R. W. Picard. Improvements in Remote Cardio-Pulmonary Measurement Using a Five Band Digital Camera. *IEEE transactions on bio-medical engineering*, 9294(10):1–8, 2014. 1, 2

[15] M.-Z. Poh, D. J. McDuff, and R. W. Picard. Noncontact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18(10):10762–10774, 2010. 1, 2

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

[17] J. C. Rajapakse, F. Kruggel, J. M. Maisog, and D. Y. von Cramon. Modeling hemodynamic response for analysis of functional MRI time-series. *Human brain mapping*, 6(4):283–300, 1998. 1

[18] W. Verkruysse, L. O. Svaasand, and J. S. Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445, 2008. 1

[19] W. Wang, S. Stuijk, and G. D. Haan. Exploiting Spatial Redundancy of Image Sensor for Motion Robust rPPG. 62(2):415–425, 2015. 2

Appendix A. Derivatives of $g(w)$

Here we present the first and second derivatives of $g(w)$ needed by the lagrange multipliers method. We follow the convention that the derivative of a scalar w.r.t a column vector is a column vector of the same size as that of the vector. The first derivative of $g(w)$ in equation 9 can be obtained as follows considering squared autocorrelation as $\mathbf{r}^2 = [r_1^2\, r_2^2 \ldots r_N^2]$.

$$g'(\mathbf{w}) = -E\{\frac{\partial}{\partial \mathbf{w}}(\begin{bmatrix} r_1^2 & r_2^2 & \cdots & r_N^2 \end{bmatrix})\} \tag{18}$$

where the derivative of the squared autocorrelation $\mathbf{r}^2$ is then obtained using the chain rule of derivatives. Also, we know that $\mathbf{y} = \mathbf{w}^T\mathbf{x}$ giving $\frac{\partial \mathbf{y}}{\partial \mathbf{w}} = \mathbf{x}$.

$$\frac{\partial(\mathbf{r}^2)}{\partial \mathbf{w}} = \mathbf{x}\frac{\partial(\mathbf{r}^2)}{\partial \mathbf{y}} \tag{19}$$

$$= \mathbf{x}\begin{bmatrix} \frac{\partial(r_1^2)}{\partial y_1} & \frac{\partial(r_2^2)}{\partial y_1} & \cdots & \frac{\partial(r_N^2)}{\partial y_1} \\ \frac{\partial(r_1^2)}{\partial y_2} & \frac{\partial(r_2^2)}{\partial y_2} & \cdots & \frac{\partial(r_N^2)}{\partial y_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial(r_1^2)}{\partial y_N} & \frac{\partial(r_2^2)}{\partial y_N} & \cdots & \frac{\partial(r_N^2)}{\partial y_N} \end{bmatrix} \tag{20}$$

$$= \mathbf{x}\begin{bmatrix} 2r_1\frac{\partial r_1}{\partial y_1} & \cdots & 2r_N\frac{\partial r_N}{\partial y_1} \\ \vdots & \ddots & \vdots \\ 2r_1\frac{\partial r_1}{\partial y_N} & \cdots & 2r_N\frac{\partial r_N}{\partial y_N} \end{bmatrix} \tag{21}$$

$$= 2\mathbf{x}\begin{bmatrix} r_1\frac{\partial r_1}{\partial y_1} & \cdots & r_N\frac{\partial r_N}{\partial y_1} \\ \vdots & \ddots & \vdots \\ r_1\frac{\partial r_1}{\partial y_N} & \cdots & r_N\frac{\partial r_N}{\partial y_N} \end{bmatrix} \tag{22}$$

The size of $\frac{\partial(\mathbf{r}^2)}{\partial \mathbf{w}}$ is then $3 \times N$ from the product of $\mathbf{x}_{3\times N}$ with the jacobian of size $N \times N$. Consequently, its expectation ends up having a size of 3×1 since it is nothing but a temporal mean over N samples. The jacobian in equation 22 can be concisely expressed as $\begin{bmatrix} r_1\frac{\partial r_1}{\partial \mathbf{y}} & r_2\frac{\partial r_2}{\partial \mathbf{y}} & \cdots & r_N\frac{\partial r_N}{\partial \mathbf{y}} \end{bmatrix}$ where each column is the product of the derivative $\frac{\partial r_1}{\partial \mathbf{y}}$ and the scalar r_k and is of size $N \times 1$. Deriving equation 4, $r_k = \mathbf{y}T_k\mathbf{y}^T$, listed here for convenience, w.r.t $\mathbf{y}$ using the product rule of differentiation,[1]

$$\begin{aligned} \frac{\partial r_k}{\partial \mathbf{y}} &= \mathbf{y}\frac{\partial}{\partial \mathbf{y}}(T_k\mathbf{y}^T) + \mathbf{y}T_k\frac{\partial}{\partial \mathbf{y}}(\mathbf{y}^T) \\ &= \mathbf{y}\frac{\partial}{\partial \mathbf{y}}(\mathbf{y}T_k^T) + \mathbf{y}T_k \\ &= \mathbf{y}T_k^T + \mathbf{y}T_k = \mathbf{y}(T_k^T + T_k) \end{aligned} \tag{23}$$

where $\frac{\partial}{\partial \mathbf{y}}(T_k\mathbf{y}^T) = T_k^T$ comes from the fact that the differential of $T_k\mathbf{y}^T$, a vector, will remain the same even when it is transposed and the derivative is computed element-wise. For conciseness, we will represent the sum $T_k + T_k^T$ as $\mathbf{T}_k$. Finally to be consistent with our convention, using the same argument of the differential being immutable to transpositions, the row vector $\frac{\partial r_k}{\partial \mathbf{y}}$ can be transposed into a column vector and the matrix $\frac{\partial \mathbf{r}}{\partial \mathbf{y}}$ can be built as

$$\frac{\partial \mathbf{r}}{\partial \mathbf{y}} = \begin{bmatrix} r_1\mathbf{T}_1\mathbf{y}^T & \cdots & r_N\mathbf{T}_N\mathbf{y}^T \end{bmatrix} \tag{24}$$

[1] This result is owing to the fact that T_k is not symmetric. If it were symmetric, then the result would have been $2\mathbf{y}T_k$.

giving $g'(\mathbf{w})$ in equation 18 as

$$g'(\mathbf{w}) = -2\mathbf{x}E\left\{\begin{bmatrix} r_1\mathbf{T}_1\mathbf{y}^T & \cdots & r_N\mathbf{T}_N\mathbf{y}^T \end{bmatrix}\right\}$$

which can be further simplified to

$$g'(\mathbf{w}) = -2\mathbf{x}\begin{bmatrix} \mathbf{T}_1\mathbf{y}^T & \cdots & \mathbf{T}_N\mathbf{y}^T \end{bmatrix}\mathbf{r}^T/N \tag{25}$$

Since the expectation is a temporal mean the element-wise multiplication with r_k can be replaced by multiplication with the vector $\mathbf{r}^T$ which also simplifies the computation.

Next, to simplify the calculation of the second derivative of $g(\mathbf{w})$, we perform column-wise matrix multiplication in equation 25, omitting the scalar multiplication and division, to obtain

$$g'(\mathbf{w}) = -\mathbf{x}\begin{bmatrix} \mathbf{T}_1 r_1\mathbf{y}^T + & \cdots & +\mathbf{T}_N r_N\mathbf{y}^T \end{bmatrix} \tag{26}$$

$$= -\mathbf{x}\sum_{k=0}^{N}\mathbf{T}_k r_k\mathbf{y}^T \tag{27}$$

And since differentiation and summation are interchangeable based on the sum rule, $g''(w)$ can be obtained by

$$g''(\mathbf{w}) = -\mathbf{x}\sum_{k=0}^{N}\frac{\partial(\mathbf{T}_k r_k\mathbf{y}^T)}{\partial \mathbf{w}} \tag{28}$$

$$= -\mathbf{x}\sum_{k=0}^{N}\frac{\partial(\mathbf{T}_k r_k\mathbf{y}^T)}{\partial \mathbf{y}}\frac{\partial \mathbf{y}}{\partial \mathbf{w}} \tag{29}$$

$$= -\mathbf{x}\left(\sum_{k=0}^{\mathbf{N}}\frac{\partial(\mathbf{T_k r_k y^T})}{\partial \mathbf{y}}\right)\mathbf{x}^T \tag{30}$$

The derivative of $\mathbf{T}_k r_k\mathbf{y}^T$ w.r.t $\mathbf{y}$ is then obtained by the product rule of differentiation.

$$\frac{\partial(\mathbf{T}_k r_k\mathbf{y}^T)}{\partial \mathbf{y}} = \mathbf{T}_k\frac{\partial r_k}{\partial \mathbf{y}}\mathbf{y}^T + \mathbf{T}_k r_k \tag{31}$$

$$= \mathbf{T}_k\left(\frac{\partial r_k}{\partial \mathbf{y}}\mathbf{y}^T + r_k\right) \tag{32}$$

which is of size $N \times N$. Consequently, the size of $g''(\mathbf{w})$ turns out to be 3×3 since the sum of $\frac{\partial(\mathbf{T}_k r_k\mathbf{y}^T)}{\partial \mathbf{y}}$ over N samples is also of size $N \times N$. $g''(\mathbf{w})$ is further used in the lagrange multipliers method for cICA , implemented as an augmented lagrangian method,presented in the next section.