Performance limits of ICA-based heart rate identification techniques in imaging photoplethysmography

2015 Physiol. Meas. 36 67

(http://iopscience.iop.org/0967-3334/36/1/67)

IOP Publishing | Institute of Physics and Engineering in Medicine Physiological Measurement

Physiol. Meas. **36** (2015) 67–83 doi:10.1088/0967-3334/36/1/67

Performance limits of ICA-based heart rate identification techniques in imaging photoplethysmography

Kavan Mannapperuma[1,2], Benjamin D Holton[1,3], Peter J Lesniewski[1,4] and John C Thomas[1,5,6]

[1] School of Electrical and Information Engineering, University of South Australia, Mawson Lakes, SA 5095, Australia
[2] Power Systems Design Pty Ltd, 16 Williams Ct, Pooraka, SA 5095, Australia
[3] Santos Pty Ltd, 60 Flinders St, Adelaide, SA 5000, Australia
[4] School of Engineering, University of South Australia, SA 5095, Australia
[5] Group Scientific Pty Ltd, PO Box 190, Salisbury South, SA 5106, Australia
[6] School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255049, People's Republic of China

E-mail: John@Group-Scientific.com.au

Received 26 May 2014, revised 29 August 2014
Accepted for publication 16 October 2014
Published 12 December 2014

Abstract

Imaging photoplethysmography is a relatively new technique for extracting biometric information from video images of faces. This is useful in non-invasive monitoring of patients including neonates or the aged, with respect to sudden infant death syndrome, sleep apnoea, pulmonary disease, physical or mental stress and other cardio-vascular conditions. In this paper, we investigate the limits of detection of the heart rate (HR) while reducing the video quality. We compare the performance of three independent component analysis (ICA) methods (JADE, FastICA, RADICAL), autocorrelation with signal conditioning techniques and identify the most robust approach. We discuss sources of increasing error and other limiting conditions in three situations of reduced signal-to-noise ratio: one where the area of the analyzed face is decreased from 100 to 5%, another where the face area is progressively re-sampled down to a single RGB pixel and one where the HR signal is severely reduced with respect to the boundary noise. In most cases, the cardiac pulse rate can be reliably and accurately detected from videos containing only 5% facial area or from a face occupying just 4 pixels or containing only 5% of the facial HR modulation.

Keywords: photoplethysmography, heart rate measurement, ICA, PCA, independent component analysis, principal component analysis

(Some figures may appear in colour only in the online journal)

1. Introduction

Heart rate (HR) monitoring is an effective way to track and determine the state of health of a person. HR can be measured directly with the stethoscope, using electrocardiography (ECG) or using pulse oximeters (Tremper and Barker 1989, Nishime *et al* 2000, Poh *et al* 2008). The latter approach, based on photoplethysmography (PPG), introduced by Hertzman (1937) has become popular due to its simplicity and accuracy. The measurement principle used by PPG relies on detecting periodic changes in the volume of blood vessels and the absorption and reflection of light by hemoglobin associated with the blood flow in the microvascular bed of tissue (Hu *et al* 2008). The popular, low cost contact PPG (CPPG) uses contact probes (such as finger or ear oximeter) or similar equipment that requires continuous contact with the subject's skin (Cheang and Smith 2003). The resulting clinical restrictions, excluding patients with skin burns, or fragile neonates are avoided by non-contact PPG (NCPPG). Imaging PPG (IPPG) is a form of NCPPG which relies on periodic variation of blood vessel volume and differential spectral absorption of light by oxy- (HbO_2) and de-oxy hemoglobin (Hb) (Prahl 1999). This allows monitoring and mapping of the patient's blood vessels and blood flow with 2D photosensitive detectors (Nijboer *et al* 1981, Zijlstra and Buursma 1997) and to determine the HR of a person at a distance using a video of exposed skin. Garbey *et al* (2004) first determined the HR from the frequency of the main peak in the power spectrum of the light intensity signal representing image pixels along the thermally exposed area. Originally thermal imagery and special infrared cameras were used but Hu *et al* (2008) demonstrated that a suitable complementary metal–oxide–semiconductor (CMOS) camera assisted by infrared or red lighting is sufficient to detect the HR. Poh *et al* (2008 2010) and Lewandowska *et al* (2011) produced reliable HR measurements using videos from common web cameras. They used independent component analysis (ICA) and principal component analysis (PCA) respectively to enhance the heart pulse signal by linear mixing of the red–green–blue (RGB) video components using statistical criteria. de Haan and Jeanne (2013) developed less restricted algebraic expressions for chrominance which did not rely on periodicity and were also capable of removing distorting signal artefacts.

Using a large data set we have compared the performance of ICA algorithms, PCA and auto- and cross-correlation in detecting the HR from webcam videos and found that JADE and RADICAL ICA tend to outperform the more standard approach (Holton *et al* 2012 2013). Our recent study (Holton *et al* 2013) indicated that with good signal quality, the HR error is relatively low, irrespective of the signal processing approach. However, we observed that deterioration of quality is manifested by an abrupt rather than gradual increase in the error.

For real-life applications it is important to determine how efficient and robust blind source separation (BSS) techniques are with real, imperfect signals. In this paper we investigate the limits of ICA and other processing techniques to extract the HR from the facial video of a subject. The main objective is to establish the level of reduction of the video data quality causing significant increase in measurement error. We have used several known pre-processing techniques and algorithms that performed well in our earlier investigation (Holton *et al* 2013) and compared their performance in progressively more difficult conditions, reducing relative level or quality of the HR related signal, in four tests, by:

(a) Reducing the analyzed face area (and thus the level of the HR related signal),
(b) Reducing the length of analyzed video segments while using various algorithms,
(c) Reducing the number of 8-bit RGB pixels averaged for analysis (by image down sampling),
(d) Increasing boundary noise within the analyzed image sequence through segmentation.

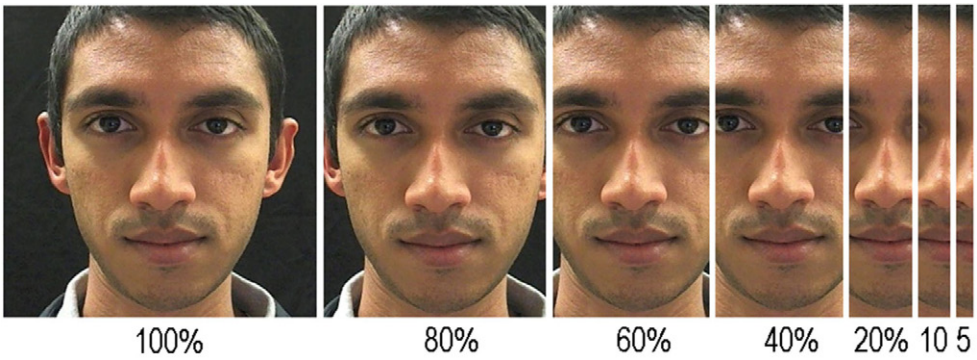

Figure 1. Reduced ROI (width) in performance tests (file 35M).

In the main investigation (a), we modified the video data by reducing the width of the template, i.e. the region of interest (ROI) containing the face, from 100% (156 446 pixels) through 80, 60, 40, 20, 15, 10 to 5% (figure 1). This progressively reduced the number of pixels carrying the HR color modulation while retaining other conditions such as the magnitude of boundary noise (due to head movement). Note that even with the ROI reduced to 5%, the number of pixels averaged was large enough not to visibly impair the HR modulation by the increased quantization intervals.

In our investigation, the RGB signals for 512 frames of each video (or the first 15 s and the first 30 s in the case (b), after low pass (LP) filtering and de-trending were analyzed using the JADE, FastICA and RADICAL ICA algorithms to generate the independent components. Power spectra of all three components were computed using the Fast Fourier transform (FFT) and the component with a clear maximum in the frequency range expected was selected for determining the HR.

The tests (a) and (b) (results in section 3.1) were followed by an analysis of other limiting effects (results in section 3.2) on HR detection which occur in real-life situations. One is the reduction of quantization steps such as in low resolution images or those containing many faces where the number of summed pixels is reduced and distorts the original modulation (c) and the other is the distorting effect of the boundary noise (around moving objects). Both of these effects are essential to fully assess the feasibility of methods used in practical applications.

2. Methods

2.1. HR extraction with different ICA algorithms

Identification of cyclic components in signals is routinely done using either autocorrelation or its equivalent, the power spectrum. This requires sufficient signal quality and signal length. The latter is not an option here as all parameters, including the periodic signal component, vary and the reasonably stationary HR signal is limited to some 30 s. Filtering to increase the signal-to-noise (*S*/*N*) ratio is problematic because the noise spectrum is similar to the spectrum of the signal to be recovered. However, multiple noisy signals (the three RGB signals) can be decomposed into basis source signals using BSS algorithms. In biomedical sciences this approach has been applied to detecting the electroencephalogram (EEG) signals (Walter-Williams and Li 2011) and more recently to extracting the HR modulation from a video of the face Poh *et al* 2010. ICA is a statistical BSS technique useful where the data is a linear

combination of unknown components (sources) which are assumed non-Gaussian and mutually independent (Jutten and Herault1991, Comon 1994, Lewandowska *et al* 2011). In our case the primary signal component of interest is a fine modulation of image intensity representing the cardiac pulse. Ideally, the statistically independent unknown source signals, $s = (s_1,\ldots,s_n)$ are related to the observed mixed data sets, $x = (x_1,\ldots,x_m)$ through a linear, constant mixing matrix A so that,

$$x = As\,. \tag{1}$$

When the number of independent source signals is the same as the number of observed mixed data sets (the case of $m = n$) and no noise or distortion is present, a complete unmixing of source signals is possible. The inverse of the mixing matrix, A, known as the de-mixing matrix, W, is used to find the original independent components through ICA, so that

$$s = Wx\,. \tag{2}$$

ICA algorithms differ from each other by the mathematical measure used to identify the most distinct solutions, i.e. the mixes represented by the weights in the matrix A. FastICA, developed by Hyvärinen and Oja (Harvaruinen 1999) uses fourth order moments (kurtosis) to identify the independent components and is implemented using a fast fixed point iterative algorithm that finds the projections maximizing the non-Gaussian character of the sources.

The ICA joint approximate diagonalization of eigenmatrices (JADE) algorithm or tensorial method, developed by Cardoso (1999) uses fourth order cumulant tensors. After joint diagonalization of the eigenmatrices the off-diagonals of these matrices are to become zero or as close to zero as possible to identify the separate components.

The ICA RADICAL (Robust Accurate Direct Independent Component Analysis aLgorithm), introduced by Learned-Miller and Fisher III (Miller and Fisher 2003) is based on estimated Kullback-Leibler divergence paired with entropy estimators.

2.2. Video data and recording

Facial videos of eighteen adults aged 22–59 years including sixteen males and two females with Caucasian, African and Asian heritage were recorded (University of South Australia Ethics Board approval No 0000026093). All video recordings were made with a standard web camera (Logitech C910 HD Pro Webcam) at a frame rate of 15fps of 24 bit RGB color (8 bits per color) and resolution of 640 × 480 pixels. All videos were recorded and saved in a lossless WMV format. All image and signal processing was implemented in MatLab (The MathWorks, Inc. www.mathworks.com) using the image acquisition toolbox. Videos were recorded at different times of the day with varying amounts of daylight entering through small side windows but the fluorescent ceiling lights (multiple 40 W battens with chokes) were the main light source. The subjects were seated 0.6 m in front of the camera and were asked to remain steady, silent and avoid any excessive facial expressions. Each recording was at least 90 s long but the start was delayed by some thirty seconds in order for the HR to become steady. Nine randomly selected video samples were used in the tests. Reference HR values were obtained from a Rossmax Finger Pulse Oximeter SB220 attached to the index finger of each subject. This was used to obtain independent readouts of the HR (and blood oxygen saturation levels) at 30 s.

2.3. Data preparation and signal processing

HR detection tests were carried on a template or ROI of full or progressively narrowing part of the face (figure 1). To reduce the influence of head movement face-tracking using the

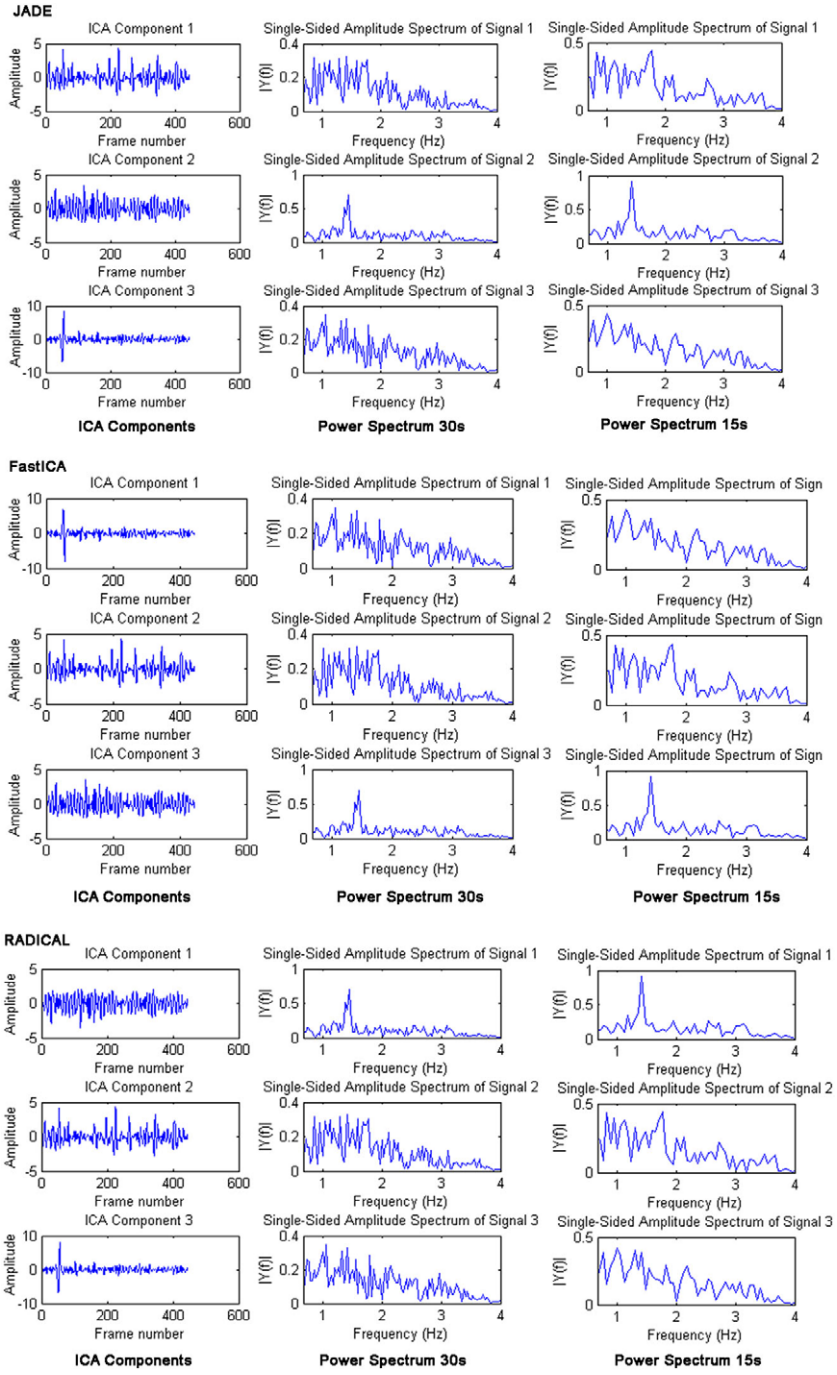

Figure 2. Examples of unmixed signals (LHS column) and their power spectra (centre column: 30 s video clips, RHS: 15 s clips) produced by the three ICA algorithms: JADE (top set), FastICA (middle) and RADICAL (bottom). The HR related spectrum is in the ICA component 2, 3 and 1 respectively.

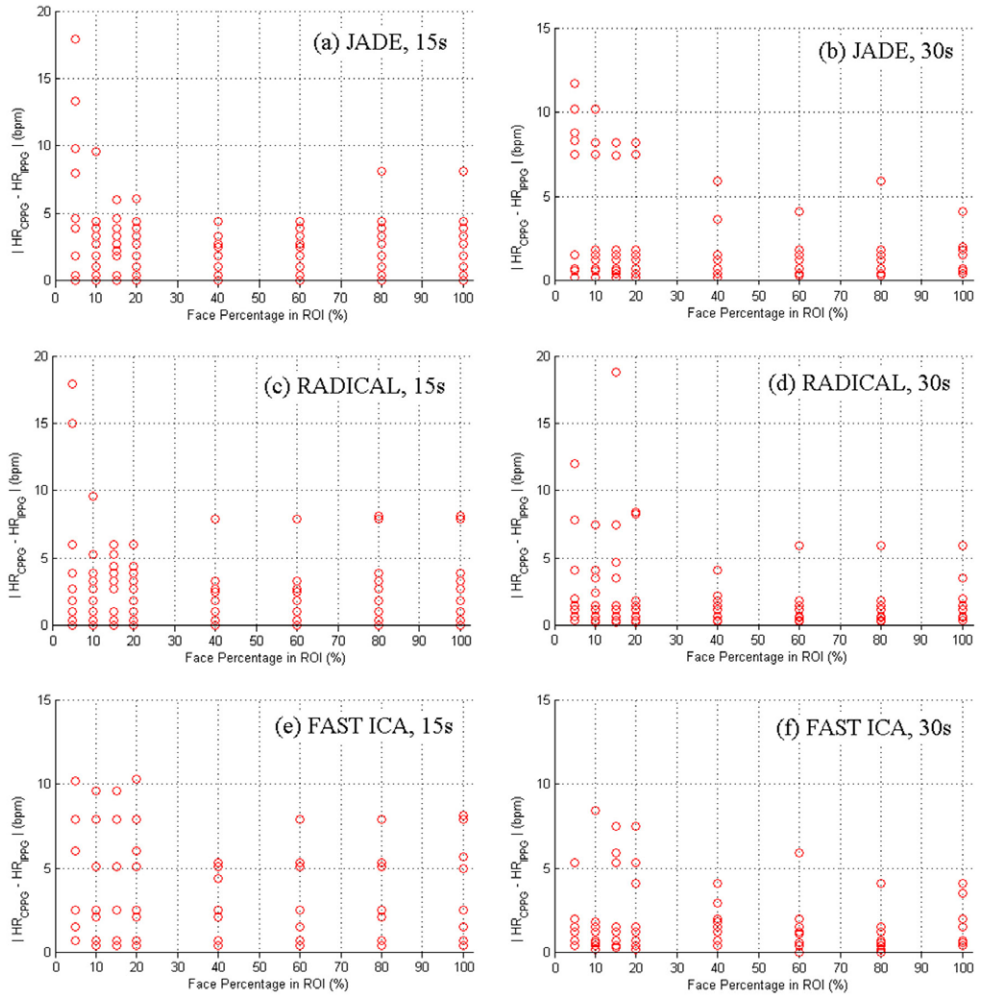

Figure 3. The absolute value of the difference between the reference HR and the imaging PPG measurement plotted against the percentage of face in the ROI for JADE, RADICAL and FastICA; (*a*) JADE 15 s, (*b*) JADE 30 s, (*c*) RADICAL 15 s, (*d*) RADICAL 30 s, (*e*) FastICA 15 s, (*f*) FastICA 30 s.

public domain Open Computer Vision (OpenCV 2.1) library was implemented in MatLab (Bradski 2000).

The OpenCV face detection method is based on the work by Viola and Jones (2004) and Lienhart and Maydt (2002). The frontal face classifier is trained using digital image features called Haar-like features resulting in a cascaded arrangement of detectors producing a sequence of sub-windows of the image. A region is classed as containing a face if a sub-window passes all detector stages Viola and Jones (2004). A face detected in a frame is represented by a box of given size and position but if multiple (false positives) or no faces (false negative) are detected then the previous face coordinates are used.

Three RGB signals (time series) were calculated by averaging all pixels of each color in every frame. A low-pass, finite response filter with a cutoff frequency of 4 Hz and a six-point

Table 1. Performance of three ICA algorithms in HR detection (reference HR values from finger pulse oximeter).

	Percentage of Face in ROI															
	100%		80%		60%		40%		20%		15%		10%		5%	
ICA method	UE	P%	UE	P%	UE	P%	UE	P%	UE	P%	UE	*P%*	UE	P%	UE	P%
JADE 30 s	4.7	100	4.6	89	4.6	100	4.6	89	4.8	78	4.9	78	4.6	67	5.3	67
RADICAL 30 s	4.7	89	4.7	89	4.7	89	4.6	100	4.2	78	4.3	78	5.1	89	9	78
FastICA 30 s	4.7	100	4.6	100	4.7	89	4.5	100	4.8	78	4.8	67	4.8	89	5.2	78
JADE 15 s	4.5	89	4.5	89	4.8	100	4.7	100	5	89	4.8	100	5.2	89	9.2	67
RADICAL 15 s	4.6	78	4.6	78	4.8	89	4.7	89	5	100	5.1	100	5.1	78	5.4	67
FastICA 15 s	4.5	89	4.7	100	4.8	100	4.7	100	4.9	89	5.7	67	5.2	78	6.5	67

Results are shown for 15 s and 30 s video sequence lengths. UE is the unsigned error between the pulse oximeter and IPPG values.
P% is the percentage of the IPPG-ICA results that agree to within ±5 bpm with of the oximeter reading.

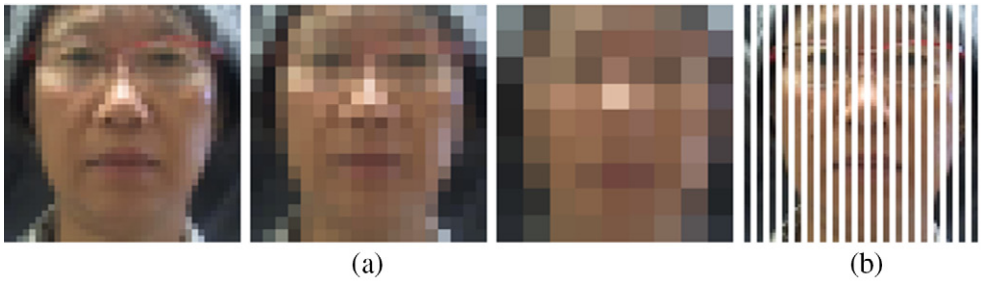

Figure 4. (*a*) Re-sampling steps reducing the number of pixels and quantization levels, from left: 38 × 38, 19 × 19 and 9 × 9 pixels. (*b*) Modification of video frames arranged to increase the ratio of the boundary noise to the part of the signal related to the HR modulation noise (example only—not to scale, the actual image used in our analysis used half of the available 311 columns).

moving average window were used to remove noise and notch out unwanted residues. The signals were then de-trended using a prior smoothness approach (Tarvainen *et al* 2002) with the datum $n = 10$ and normalized, as shown in equation (3).

$$y_i'(t) = \frac{y_i(t) - \mu_i}{\sigma_i}. \tag{3}$$

Here, for $i = 1,2,3,\ldots,n$, μ_i is the mean and σ_i is the standard deviation of the signal $y_i(t)$.

3. Effects of processing algorithm and compromised image quality

3.1 . Reduction of the analyzed face area and signal length

Examples of the three unmixed ICA components and the FFT spectra for 15 s and 30 s video clips of the '80% face image' of JADE, RADICAL and FastICA algorithms are given in figure 2.

The 80% face image (figure 1) was expected to give the best signal quality as the image embraces most of the face with a minimum border. The highest spectral peak close to the reference HR was used to identify the unmixed component representing the detected HR. The

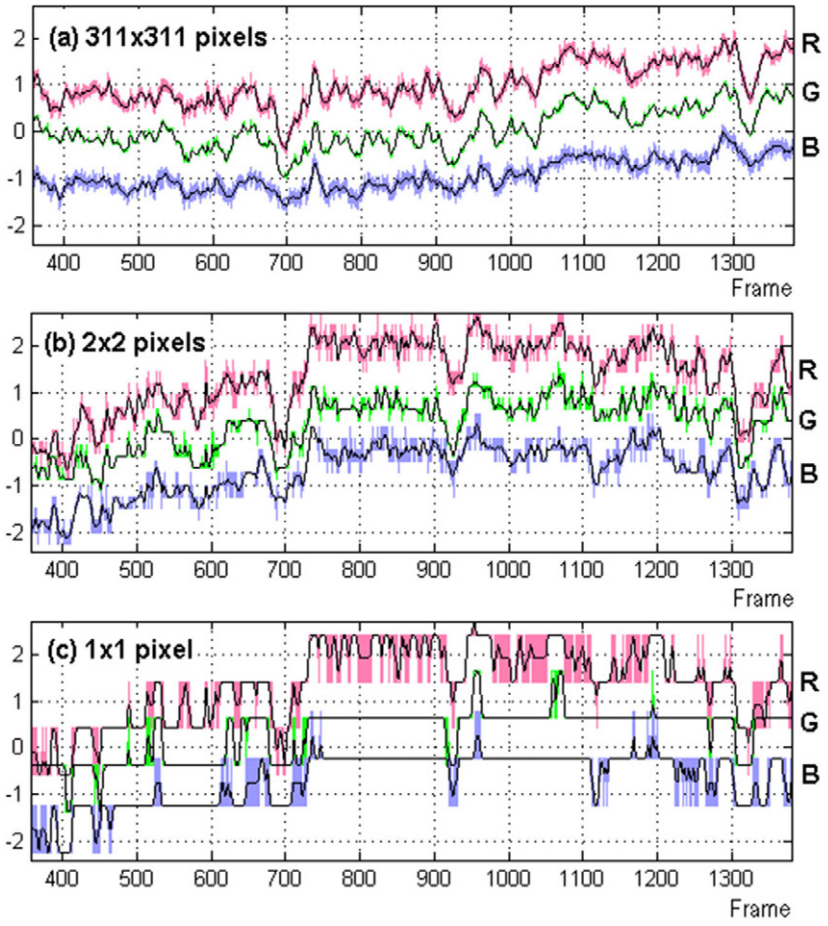

Figure 5. Variation of intensity per pixel about the mean of RGB signals (red/green/blue traces) for different re-sampling steps: top box—original resolution of 96 721 pixels, middle box—4 pixels, bottom box—1 pixel. Black, solid traces—LP filtered (4 Hz); shadow/color traces—original signals. R and B traces are offset for legibility bringing R traces up and B traces down in each plot.

degree of agreement between the ICA values and those from the finger oximeter is shown in figure 3 as absolute error for all data with individual data shown as circles. It is clear that all algorithms perform similarly but in all cases reduction of the ROI as well as the reduction of signal length have different effects on the error.

In other words, some data are more prone to error under compromised conditions while others exhibit low error in all cases. A more conclusive assessment of error requires an overall measure to represent all data points. Table 1 lists the overall performance for all the underlying data (as the mean square error and the percentage of the HRs detected by the ICA within 5 bpm of the oximeter readings) of the three ICA algorithms as the percentage of ROI and the length of the video signal are varied. In our previous study (Holton *et al* 2013), JADE and RADICAL algorithms outperformed FAST ICA based on root-mean-square (RMS) error and Pearson's correlation coefficient. However, for the current results, with reduced ROI, it is difficult to clearly identify the best overall approach. RADICAL and FASTICA algorithms seem

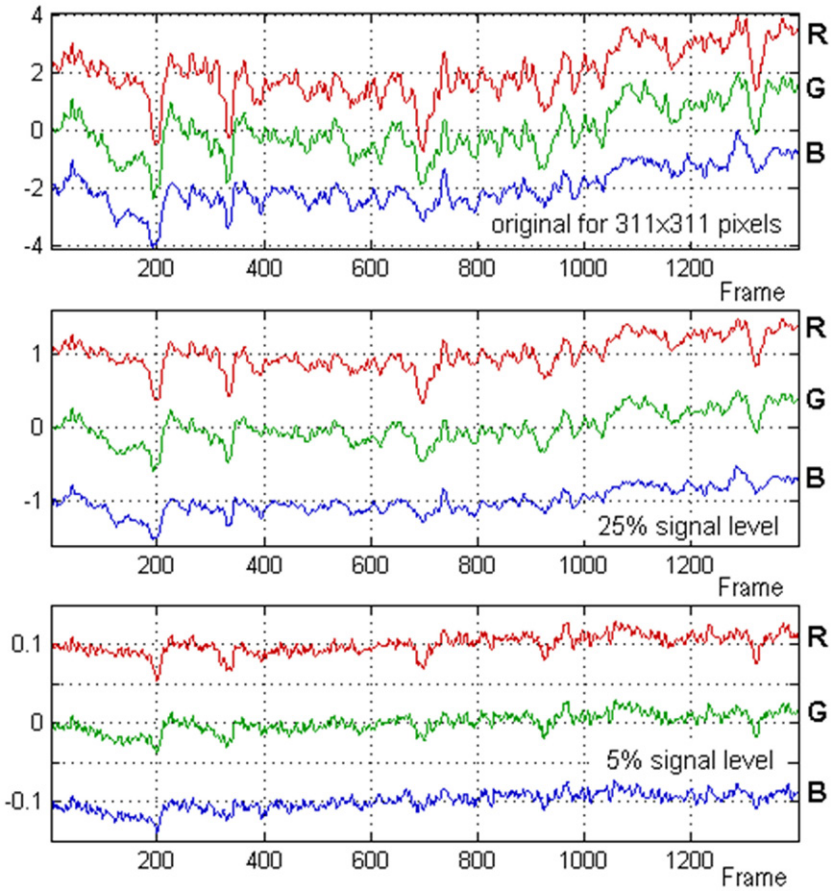

Figure 6. RGB signals (red/green/blue, LP filtered) as intensities per pixel, resulting from the reduction of the HR modulation ratio to the boundary noise. Top box: original face 311 × 311pixels, middle box: HR data volume reduced to 25%, bottom box: HR data reduced to 5% of the original. R and B traces are offset for legibility bringing R traces up and B traces down in each box.

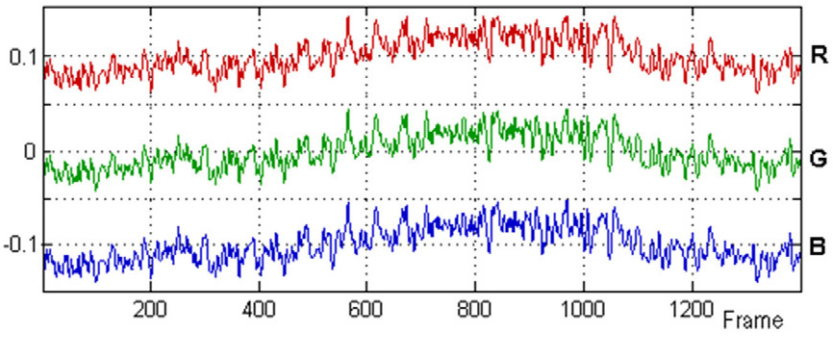

Figure 7. RGB signals (red/green/blue, LP filtered, intensities per pixel) as in figure 6 but now after full removal of the HR modulation exposing the boundary noise. R and B traces are offset for legibility bringing R traces up and B traces down.

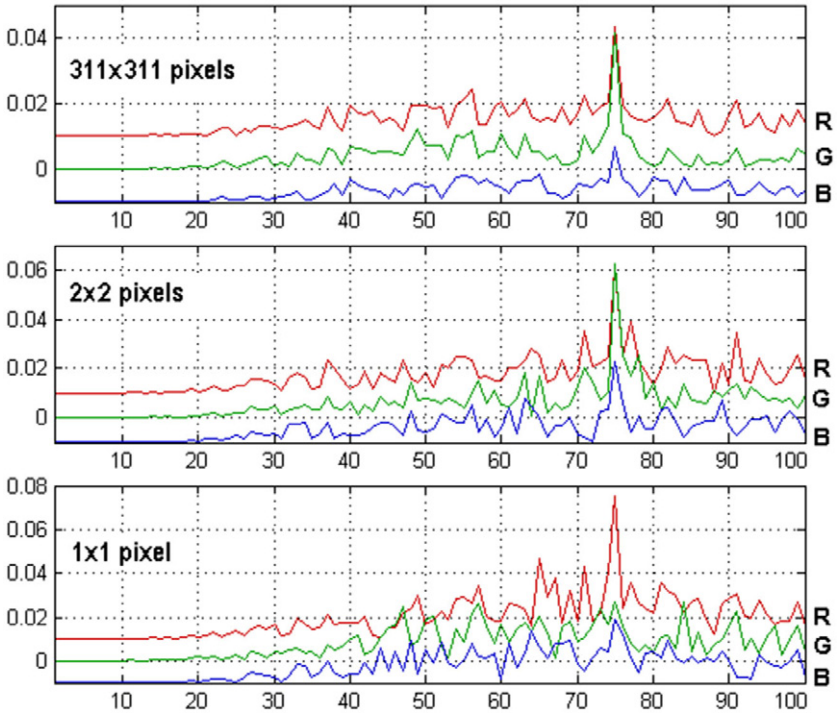

Figure 8. Spectra of RGB signals (1024 frames) from re-sampled videos (as in figure 5), before applying ICA. Frequency unit is 15/1024 = 0.0146 Hz. Oximeter reading 66 bpm =1.1 Hz (75.1 units)

to perform better when the ROI is reduced to the minimum. On the other hand, JADE ICA gives smaller errors when the ROI is larger than 10% but then its performance drops more rapidly than that of the other algorithms.

3.2. Effects of low signal quality: down sampling and boundary noise

The good results, shown by the low error rates obtained even when the ROI was reduced to only 5% of the full face area, presented in table 1, indicated that the signal processing chosen is robust and reliable. However, to gain more insight into the process of detection breakdown it was necessary to identify and separate the two types of signal distortion associated with the reduction of the object area.

The two types of signal distortion are

- decrease in signal quantization steps down to 256 (8bit) per color if one pixel used,
- increase in the boundary noise (due to boundary movement).

The first effect is illustrated in figure 4(*a*) and leads to strong nonlinear distortion of small intensity modulation typically due to the HR. This may eventually show up as extended periods of signal at one quantization level without visible modulation as seen in figure 5. This is important when a video includes many faces to be monitored, each occupying just a few pixels.

The second effect occurs when the face moves against the background leading to intensity modulation at its boundary but it also occurs if the moving object is perfectly tracked due to the background moving ie varying. Other sources of such distortion are lip, teeth and eye

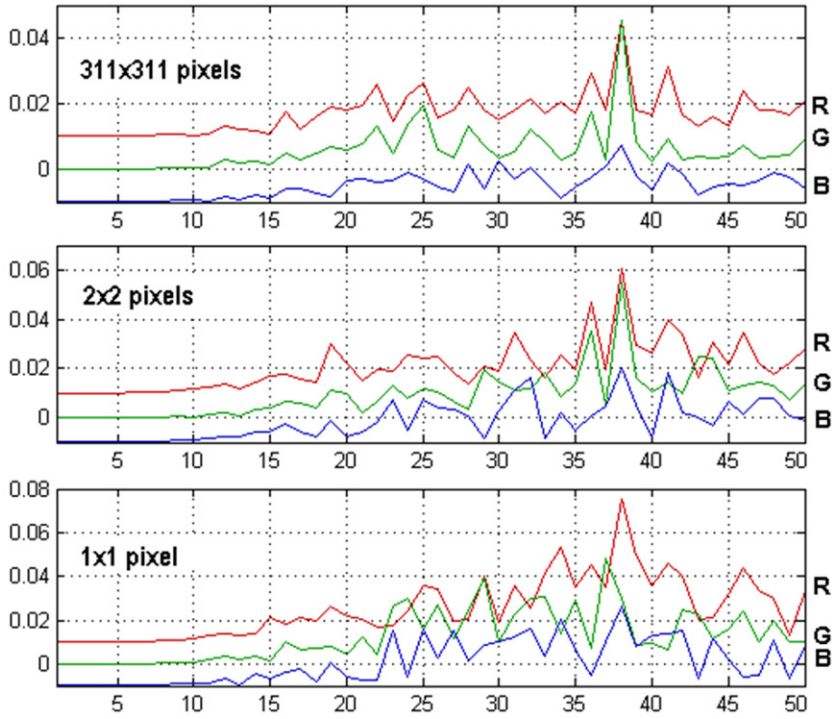

Figure 9. Spectra of RGB signals from re-sampled videos (512 frames), before applying ICA. Frequency unit is 15/512 = 0.0293 Hz. Oximeter reading 66 bpm =1.1 Hz (37.6 units).

movement. Since the signal of interest is in the low frequency range, even slow motion can generate distortion.

In the case of the normal reduction of the ROI to only a fragment of the face (results in section 3.1) there is no boundary noise other than from lip and eye movement.

Rather than produce a large quantity of special videos for each type of distortion, we have generated progressive controlled distortion by modifying just a few videos. Those selected were good quality videos with a large number of pixels.

To increase the boundary noise without affecting the HR modulation, for each video frame we generated two complementary images A and B containing half of the available columns each as shown in figure 4(*b*) so that summing them up would produce the full undistorted image.

Therefore, summing the two signals (ie average RGB intensities per pixel)) representing each image is equal to the signal representing the full undistorted image.

In the extreme case, subtracting those two signals would produce a signal with boundary noise doubled but with virtually no HR modulation (ie with the *S*/*N* ratio reduced to zero). To obtain an intermediate *S*/*N* reduction throughout the image, we mix the two signals representing frames A and B by adding the variable portion of signal B to signal A with the weight ranging from 1 to −1.

The resulting RGB signals as intensities per pixel are given in figure 6 and show that even at 5% of the original HR modulation the original signals can be identified. Only full subtraction of the two complementary signals fully masks the HR modulation by the boundary noise (figure 8).

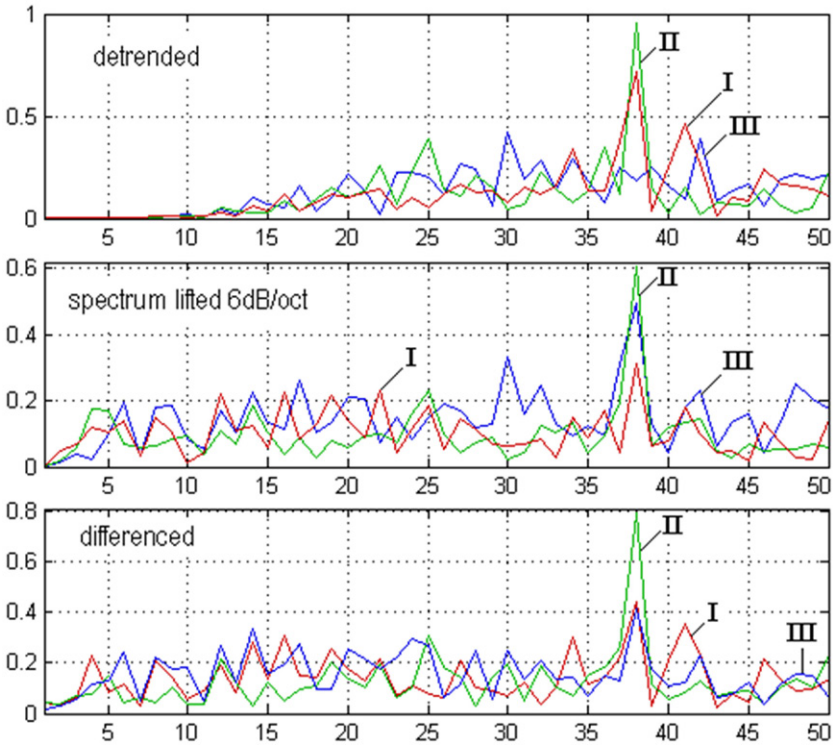

Figure 10. Spectra of unmixed components from JADE ICA (512 frames, 311 × 311 pixels). Frequency unit is 15/512 = 0.0293 Hz. Oximeter reading 66 bpm =1.1 Hz (37.6 units). Green traces (II): the HR relevant ICA component 2. Effects of different pre-processing, after LP filtering: top box—detrended, middle box—spectrum boost 6 dB/oct, bottom box—signal differenced.

R and B traces are offset for legibility bringing R traces up and B traces down.

Spectra of the signals distorted by re-sampling show that even at a single pixel size, the HR frequency peak remains unchanged and is clearly visible for video segments of 1024 frames (figure 8) and marginally for 512 frames (figure 9) but not for 256 frames.

When using the JADE ICA, the quality of the spectra for a signal length of 512 frames show a clearer HR peak (figure 11) but not so for 256 frames (not shown). On the other hand, if the signal quality is high (figure 8–12), other independent components also show a dominant HR frequency peak.

The effects of different pre-processing, after LP filtering of the RGB signals on the ICA performance is shown in figures 10 and 11 indicating great tolerance of the ICA approach to signal quality. All that is needed for the FFT spectrum to extract the HR peak in its usual frequency range is to attenuate the low frequency range. This can be done using the *prior smoothness approach* de-trending algorithm or, more simply, by differencing the data which gives a similar effect under normal circumstances. Alternatively, the differencing can be replaced by computationally simpler and faster scaling (lifting at 6 dB/oct) of the final frequency spectrum. However, if the videos have variable lighting with low frequency pulsations then de-trending tends to be better due to its sharp high pass (HP) filtering ability and short response time.

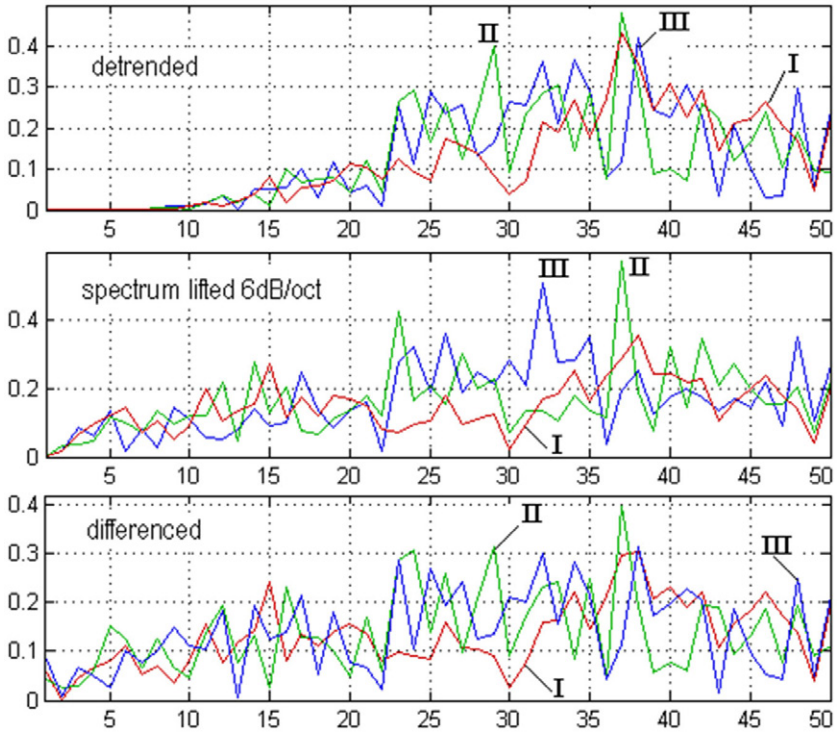

Figure 11. Spectra of unmixed components from JADE ICA (512 frames, 1 × 1 pixel). Frequency unit is 15/512 = 0.0293 Hz. Oximeter reading 66 bpm = 1.1 Hz (37.6 units). Green traces (II)—the HR relevant ICA component 2. Effects of different preprocessing, after LP filtering: top box—detrended, middle box—spectrum lift by 6 dB/oct, bottom box—signal differenced.

The autocorrelation function of a signal is the inverse Fourier transform of its power spectrum but the autocorrelation is not as useful in identifying a dominant frequency in the spectrum. Figure 12 presents autocorrelation functions computed for the JADE ICA component 2 under the same signal conditions as in figures 10 and 11. It can be seen that only for fair quality de-trended signals is the autocorrelation maximum at 15 frames lag (1 Hz) similar to that found from the FFT results (1.09 Hz). Results for signals with the HR modulation progressively reduced with respect to the boundary noise are shown in figures 13–15. In all cases of the HR signal deterioration, the HR spectral peak was dominant, even in grayscale images when the ICA was used.

Traces are offset for legibility.

Frequency unit is 15/512 = 0.0293 Hz. Oximeter reading 66 bpm =1.1 Hz (37.6 units).

4. Discussion

The reduction of the ROI area has little effect on the HR error. From 100% through to 40% reduction in the ROI, the percentage of data representing HRs within 5 bpm error band with respect to the reference value is close to 100% for JADE, RADICAL and FastICA. When the ROI percentage drops to 20% and further, the error exceeds the 5 bpm limit but slightly. Only

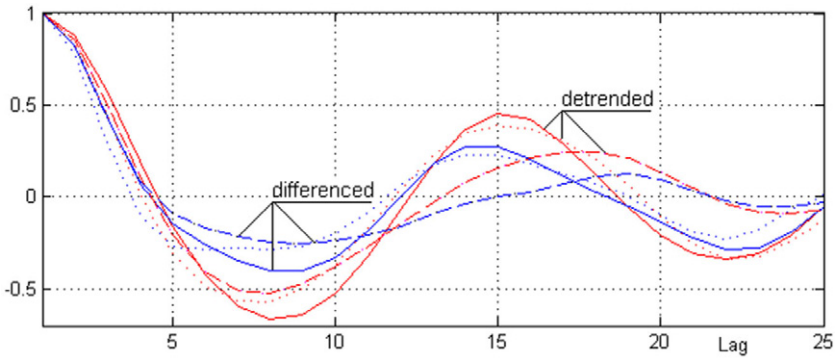

Figure 12. Autocorrelation of the JADE ICA component 2 traces: red/blue—de-trended/differenced RGB signals. Data was down sampled to: solid trace—311 × 311pixels, dotted—2 × 2 pixels, dashed—1 × 1 pixel.

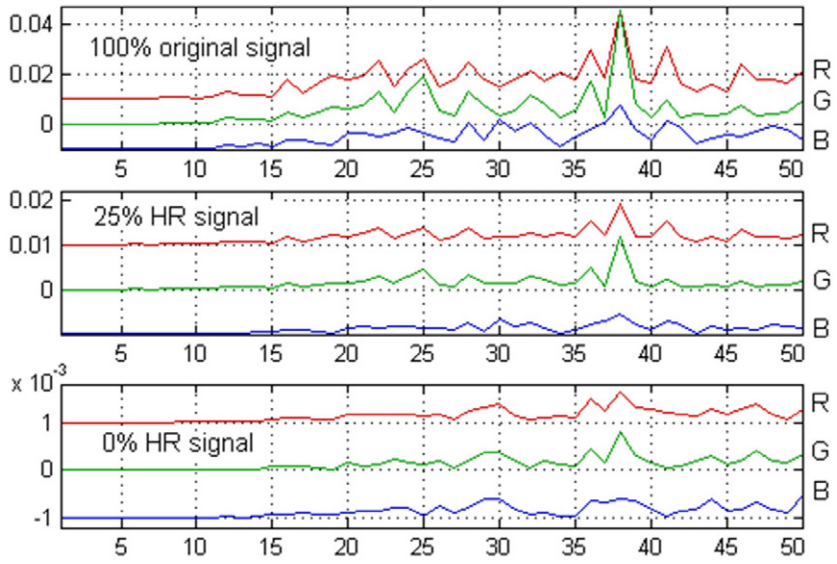

Figure 13. Spectra of RGB signals (512 frames) with reduced HR energy and increased boundary noise. Top-down: 100% (original HR signal), 25%, 0% (signal as in figure 7). R and B traces are stacked for legibility bringing R traces up and B traces down. Frequency unit is 15/512 = 0.0293 Hz. Oximeter reading 66 bpm =1.1 Hz (37.6 units).

when the ROI is reduced to 5% does the error increase significantly, approaching the expected knee on the plot. This could result from a reduction of the PPG *S*/*N* level caused by specular reflection on the nose in many videos.

The results for the 15 s data length showed similar trends to the 30 s segments for all ICA algorithms used. Only the overall RMS error for all data indicated that the RADICAL and FastICA algorithms performed similarly and better than JADE. Possibly, different RGB signals had their statistics matched better to one algorithm than another. Also, many videos included white glare on the nose, forehead or glasses, and blinking, lip and lateral head movements

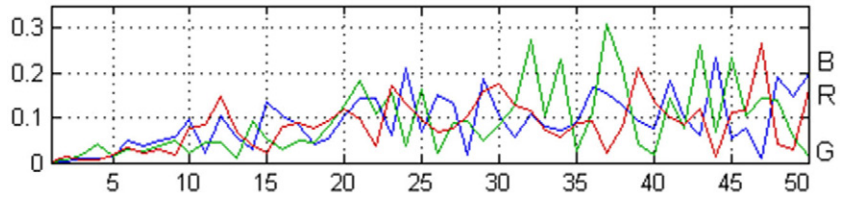

Figure 14. Spectra of unmixed JADE ICA components for signals (512 frames) with reduced HR energy and increased boundary noise—the case of 0% HR energy (signal as in figure 7). Signal not detrended but spectrum lifted 6 dB/oct. Frequency unit is 15/512 = 0.0293 Hz. Oximeter reading 66 bpm = 1.1 Hz (37.6 units).

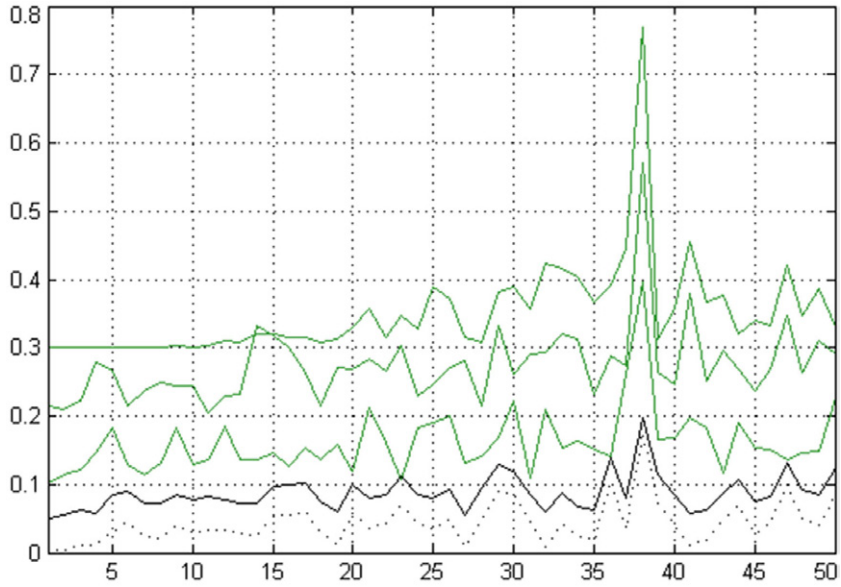

Figure 15. Offset spectra of unmixed JADE ICA component 2 or grayscale signal (representing 512 frames) for signals with reduced HR energy and increased boundary noise—case of 5% HR energy. Top-down: detrended signal, differenced, spectrum lifted 6 dB/oct, grayscale de-trended, grayscale differenced (dotted).

which reduce the HR modulation signal or produce large spikes that have similar masking effects. That type of irregular distortion is very difficult to assess with respect to performance.

Remarkably, on selected (good and fair quality) videos, the HR could be correctly and easily measured at 5% of the ROI or when its energy was reduced to 5% or when the whole face was contained only within 4 pixels.

Interestingly, the removal of the HR related energy from signals had a disproportionately smaller effect on the reduction of the HR peak in the spectrum (figure 13). Surprisingly, even for the signal denoted 0%, a remnant spectral peak at the expected HR point was still present. A possible explanation is that the HR modulation is not simultaneous across the face so that the various parts contribute less than an algebraic sum; in other words, there is a phase difference between the HR modulation in different parts of the face. Another reason for the ability to detect the HR spectral component even from a single pixel image and with the quantization

as coarse as 8-bit was the presence of higher frequency noise and intermodulation increasing the overall signal modulation above the quantization interval. After LP filtering, the noise is removed while the HR modulation reappears—a linearization effect similar to that in analog magnetic recording with a high frequency bias.

In spite of the high performance of the ICA based approach, there are obvious drawbacks. Firstly, the unmixing of components is 'blind' which means that it is designed to separate distinct components rather than search for a specific one. Also, the presence of three other signals which were stronger than the cardiac signals, would lead to their unmixing and may not yield the cardiac pulse. Secondly, the HR enhancing mix is not necessarily present in the same unmixed component every time. This occurs as a result of the ICA iterative operation. The ability to determine the power spectrum order at which the HR will appear, and thereby to automate the system is therefore hindered by the need to compute and analyze all three components every time.

Spectral analysis with FFT has its well known limitation linking frequency resolution with the length of the analyzed signal. In our situation the limit was 512 samples which gives frequency resolution of 15/512 Hz = 1.75 bpm. Analyzing half length (256 frame) clips would halve the resolution increased to 3.5 bpm which is too coarse. This limits measurement of high variability HR.

The pre-trained frontal face classifier provided with OpenCV could not be regarded as an ideal face detector in this application as there were a number of false positives observed in some cases. Although this was remedied in the algorithm developed, in an unsupervised scenario with subject moving to a greater extent, this could be an issue. As a long term solution, a classifier could be given extensive training on a particular scenario or application to achieve user defined targets (Munder and Gavrila 2006). The face detector tends to work well for sideways tilted faces to an angle of approximately ±15° and about ±45° out of plane (toward a profile view) (Viola and Jones 2004). No false negatives were observed in our videos.A thorough application-specific training carried on a classifier could lead to improving the rejection rates of negative detections. A faster and simpler approach is object tracking using adaptive template which is even capable of tracking rotating or deformed objects (Nicoud *et al* 2011). Otherwise, a different stable solution for the problems of false positives and false negatives would be the implementation of a skin or chroma discriminator to pass undistorted signal (Heisele *et al* 2006, Chen *et al* 2007, de Haan and Jeanne 2013).

5. Conclusion

It has been demonstrated that the cardiac pulse rate of a person can be determined with good accuracy (~90%) using a minimum facial area percentage of 40% taken from the centre of the face. This performance can be achieved using a video segment exposing 15 s of face. All ICA algorithms produced similar levels of accuracy. This could be investigated further by the use of a large number of facial videos. If the power spectrum in which the cardiac pulse appears could be determined the system can be automated.

References

Bingham E and Hyvarinen A 2000 A fast fixed-point algorithm for independent component analysis of complex valued signals *Int. J. Neural Syst.* **10** 1–8

Bradski G 2000 The OpenCV library *Dr. Dobb's J. Software Tools* (ftp://ftp.math.utah.edu/pub/tex/bib/toc/dr-dobbs-2000.html)

Cardoso J F 1999 High-order contrasts for independent component analysis *Neural Coput.* **11** 157–92
Cheang P Y S and Smith P R 2003 An overview of non-contact photoplethysmography *Electron. Syst. Control Div. Res.* 57–9
Chen H-Y, Huang C-L and Fu C-M 2007 Hybrid-boost learning for multi-pose face detection and facial expression recognition *Pattern Recognit.* **41** 1173–85
Comon P 1994 Independent component analysis, a new concept? *Signal Process.* **36** 287–314
de Haan G and Jeanne V 2013 Robust pulse-rate from chrominance-based rPPG *IEEE Trans. Biomed. Eng.* **60** 2878–86
Garbey M, Sun N, Merla A and Pavlidis I 2004 Contact-free measurement of cardiac pulse based on the analysis of thermal imagery *IEEE Trans. Biomed. Eng.* **54** 1418–26
Harvaruinen A 1999 Fast and robust fixed-point algorithms for independent component analysis *IEEE Trans. Neural Networks* **10** 626–34
Heisele B, Serre T and Poggio T 2007 A component-based framework for face detection and identification *Int. J. Comput. Vision* **74** 167–81
Hertzman A B 1937 Photoelectric plethysmography of the fingers and toes in man *Exp. Biol. Med.* **37** 290–2
Holton B, Mannapperuma K, Lesniewski P J and Thomas J C 2012 Data analysis techniques in imaging photoplethysmography *Proc. 20th Australian Institute of Physics Congress* (Sydney, Australia, December 2012)
Holton B D, Mannapperuma K, Lesniewski P J and Thomas J C 2013 Signal recovery in imaging photoplethysmography *Physiol. Meas.* **34** 1499–511
Hu S, Zheng J, Chouliaras V and Summers R 2008 Feasibility of imaging photoplethysmography *(BMEI2008) Int. Conf. BioMedical Engineering and Informatics (Sanya, China)* pp 72–5
Jutten C and Herault J 1991 Blind separation of sources: I. An adaptive algorithm based on neuromimetic architecture *Signal Process.* **24** 1–10
Lewandowska M, Ruminski J, Kocejko T and Nowak J 2011 Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity *(IEEE) Proc. Fed. Conf. Computer Science and Information Systems (FedCSIS2011 Sept 2011, Szczecin Poland)* pp 405–10
Lienhart R and Maydt J 2002 An extended set of Haar-like features for rapid object detection *Proc. Int. Conf. on Imaging Processing (Sept 2002, Rochester NY)* **1** 900–3
Miller E G L and Fisher J W 2003 ICA using spacings estimates of entropy *J.Mach. Learn. Res.* **4** 1271–95
Munder S and Gavrila D M 2006 An experimental study on pedestrian classification *IEEE Trans. Pattern Anal. Mach. Intell.* **28** 1–6
Nicoud F, Castellazzi G, Lesniewski P and Thomas J C 2011 Fast template tracking in video sequences *Rev. Sci. Instrum.* **82** 105110
Nijboer J A, Dorlas J C and Mahieu H F 1981 Photoelectric plethysmography-some fundamental aspects of the reflection and transmission methods *Clin. Phys. Physiol. Meas.* **2** 205–15
Nishime E O, Cole C R, Blackstone E H, Pashkow F J and Lauer M S 2000 Heart rate recovery and treadmill exercise score as predictors of mortality in patients referred for exercise *Am. Med. Assoc.* **284** 1392–8
Poh M-Z, McDuff D J and Picard R W 2010 Advancements in noncontact, multiparameter physiological measurements using a webcam *IEEE Trans. Biomed. Eng.* **58**
Poh M-Z, McDuff D J and Picard R W 2010 Non-contact, automated cardiac pulse measurements using video imaging and blind source separation *Opt. Express* **18** 10762–74
Prahl S 1999 Optical absorption of hemoglobin (http://omlc.ogi.edu/spectra/hemoglobin/index.html) 22 August 2011
Tarvainen M P, Ranta-aho P O and Karjalainen P A 2002 An advanced detrending method with application to HRV analysis *IEEE Trans. Biomed. Eng.* **49** 172–5
Tremper K K and Barker S J 1989 Pulse oximetry *Anesthesiology* **70** 98–108
Viola P and Jones M J 2004 Robust real-time face detection *Int. J. Comput. Vision* **57** 137–54
Walter-Williams J and Li Y 2011 Performance comparison of known ICA algorithms to a wavelet ICA merger *Signal Process.* **5** 80–92
Zijlstra W G and Buursma A 1997 Spectrophotometry of hemoglobin: absorption spectra of bovine oxyhemoglobin, deoxyhemoglobin, carboxyhemoglobin, and methemoglobin *Comparative Biochem. Physiol.* B **118** 743–9