# Non-contact video-based vital sign monitoring using ambient light and auto-regressive models

**L Tarassenko**[1]**, M Villarroel**[1]**, A Guazzi**[1]**, J Jorge**[1]**,
D A Clifton**[1] **and C Pugh**[2]

[1] Department of Engineering Science, Institute of Biomedical Engineering,
University of Oxford, UK
[2] Nuffield Department of Medicine, University of Oxford, UK

E-mail: lionel.tarassenko@eng.ox.ac.uk

## Abstract

Remote sensing of the reflectance photoplethysmogram using a video camera typically positioned 1 m away from the patient's face is a promising method for monitoring the vital signs of patients without attaching any electrodes or sensors to them. Most of the papers in the literature on non-contact vital sign monitoring report results on human volunteers in controlled environments. We have been able to obtain estimates of heart rate and respiratory rate and preliminary results on changes in oxygen saturation from double-monitored patients undergoing haemodialysis in the Oxford Kidney Unit. To achieve this, we have devised a novel method of cancelling out aliased frequency components caused by artificial light flicker, using auto-regressive (AR) modelling and pole cancellation. Secondly, we have been able to construct accurate maps of the spatial distribution of heart rate and respiratory rate information from the coefficients of the AR model. In stable sections with minimal patient motion, the mean absolute error between the camera-derived estimate of heart rate and the reference value from a pulse oximeter is similar to the mean absolute error between two pulse oximeter measurements at different sites (finger and earlobe). The activities of daily living affect the respiratory rate, but the camera-derived estimates of this parameter are at least as accurate as those derived from a thoracic expansion sensor (chest belt). During a period of obstructive sleep apnoea, we tracked changes in oxygen saturation using the ratio of normalized

reflectance changes in two colour channels (red and blue), but this required calibration against the reference data from a pulse oximeter.

## 1. Introduction

It has now been established that it is possible to record the blood volume changes associated with the cardiac cycle remotely from facial images of human subjects with a digital video camera up to 2 m away from the subject, using only ambient light as the light source. This ability to record photoplethysmographic signals remotely opens up the possibility of non-contact vital sign monitoring. Conventional patient monitoring using pulse oximetry to measure heart rate and peripheral arterial oxygen saturation ($SpO_2$) requires a probe to be attached to the patient's ear or finger. Long-term monitoring outside the intensive care unit is plagued by motion artefact, leading to frequent false alerts. While non-contact monitoring will also be affected by motion artefact, the use of a camera indicates *when* the subject is moving, whereas conventional monitoring provides no information on the cause of the artefact. Additionally, the probes used in conventional monitoring often cause patient discomfort and increase the risk of spreading infection in hospitals. Hence a method for deriving estimates of heart rate and $SpO_2$ with no electrodes or sensors attached to the patient is an attractive proposition, with the cost of digital video cameras continuing to decrease as the technology becomes more ubiquitous. The same non-contact monitoring technology can also be used to obtain estimates of respiratory rate, another parameter of major clinical significance (Cretikos *et al* 2008).

This paper advances the state-of-the-art in non-contact vital sign monitoring by making a number of significant contributions: firstly, a novel method of cancelling out aliased frequency components caused by strong artificial light flicker, based on auto-regressive (AR) modelling, is presented; secondly, highly accurate maps of the spatial distribution of heart rate and respiratory rate data are constructed from the coefficients of the AR model; and finally vital sign data, including changes in oxygen saturation, are acquired from patients in the clinic who are double-monitored, rather than from healthy volunteers as has invariably been the case in the papers so far published on non-contact vital sign monitoring.

The paper is organized as follows. We start with a review of previous work to establish the novelty of our approach. Our methods for non-contact vital sign monitoring using AR models are then described: we introduce a novel pole-cancellation algorithm for the removal of aliased frequencies; we describe how heart rate, respiratory rate and changes in peripheral oxygen saturation can be estimated from video images of patients undergoing dialysis in the local renal unit. Results from these algorithms, including maps of the spatial distribution of heart rate and respiratory rate for a typical patient, are presented. Finally, the paper ends with a discussion of the applicability of non-contact vital sign monitoring in the clinic.

## 2. Review of previous work

It has been well known since the 1930s—see, for example, the introduction in the paper by Verkruysse *et al* (2008)—that the variations in tissue blood volume in a body segment with each heart beat modulate the transmission or reflection of visible (or infra-red) light through

or from that body segment. All forms of the haemoglobin molecule, the protein responsible for oxygen transport in blood, are highly absorptive in the visible and near-infrared, hence the variations in blood volume during the cardiac cycle affect the transmission or reflection of light in time with the heart beat. The cardiac-synchronous variations in light transmission or reflectance are known as the photoplethysmographic or PPG signal. The pulse rate (or heart rate—the two are used interchangeably) can easily be extracted from the PPG signal by measuring the time interval between two consecutive peaks (or troughs) of the PPG waveform.

In the 1970s, the technique of pulse oximetry was developed to obtain a non-invasive estimate of $SpO_2$ by measuring the PPG signal at two wavelengths (Severinghaus and Honda 1987). Oxygenated haemoglobin ($HbO_2$) and deoxygenated haemoglobin (Hb) have significantly different optical spectra in the wavelength range from 500 to 1000 nm. Hence, by measuring the pulsatile changes in the light transmitted through tissue at two different wavelengths using a simple probe with two light-emitting diodes (LEDs), pulse oximeters determine the oxygen saturation of the arterial blood in the tissue non-invasively.

Most conventional pulse oximeters operate in this transmittance mode (Chan *et al* 2013), with the probe attached to the finger or earlobe. Reflectance-mode pulse oximeters, which have the light sources and detector side by side on the same plane, are also occasionally used. These work because some of the light is back-scattered by the tissues. The main use of this type of pulse oximeter is in forehead pulse oximetry (Agashe *et al* 2006), for clinical applications in which peripheral shut-down occurs, for example during vascular surgery (Wax *et al* 2009) or in severe shock (Nesseler *et al* 2012). During peripheral shut-down, conventional, transmittance-mode, pulse oximetry loses accuracy, and sometimes fails outright, because of the reduction in blood flow in the extremities.

The possibility of measuring back-scattered light remotely using a camera (rather than a reflectance probe in contact with the body) is first discussed in the scientific literature around 2005 (Wieringa *et al* 2005, Humphreys *et al* 2007). The 2005 paper describes how reflectance-mode plethysmographic signals were acquired in seven volunteers using a CMOS camera positioned 0.7 m away from the left inner arm (near the wrist). The arm was exposed to diffuse non-coherent monochromatic light, at three different wavelengths in turn (660, 810 and 940 nm). The camera-derived signals all contained a respiration-correlated pulsatile component, as well as smaller amplitude variations at the cardiac frequency. Despite the paper's title ('A first step toward $SpO_2$ camera technology'), there were no experimental results relating to oxygen saturation.

In the 2007 paper (Humphreys *et al* 2007), a CMOS camera was used to image the light reflected from the inner arm of ten volunteers, adjacent to an area illuminated by an array of LEDs (with wavelengths of 760 and 880 nm). This allowed two multiplexed PPG waveforms to be captured simultaneously at a rate of 16 frames $s^{-1}$. The agreement between estimates of heart rate derived from the camera images and from a conventional pulse oximeter was shown to be excellent. The feasibility of estimating oxygen saturation from the camera images at two different wavelengths was investigated, but no results were presented.

Although the early work concentrated on the feasibility of non-contact oxygen saturation measurement using illuminated tissue, the focus then switched to heart rate and respiratory rate estimation with no illumination other than ambient light. Verkruysse *et al* (2008) showed, for the first time, that PPG signals could be remotely acquired from the human face with a simple, digital, consumer-level camera as the detector more than 1 m away; daylight was used as the illumination source in combination with normal artificial fluorescent light. Regions of interest (ROIs), usually the forehead, were selected in images of the faces of human volunteers. The raw signal, calculated as the average of all pixel values in the ROI, was band-pass filtered

using a fourth-order Butterworth filter (with cut-off frequencies of 0.8 and 6 Hz, corresponding to a heart rate range of 48–360 beats min$^{-1}$).

The authors presented evidence that the reflectance signals were true PPG signals by showing that signals corresponding to movement of facial areas with no exposed skin (edge of the face and hair above the ear) were not predominantly at the heart rate frequency. The green channel was found to provide the strongest plethysmographic signal, corresponding to an absorption peak by oxyhaemoglobin, but the red and blue channels were also shown to contain plethysmographic information. The paper showed how heart rate could be extracted from the frequency content of these images using the fast Fourier transform (FFT) for 10 s windows, and hinted at how respiratory rate might be computed using an ROI which encompasses the entire face.

The first two papers from MIT (Poh *et al* 2010, 2011) describe the recording of videos of facial regions from 12 human volunteers with varying skin colours (Asians, Africans and Caucasians). They used as their detector the standard webcam embedded in a Macbook Pro laptop (Apple Inc.), positioned approximately 0.5 m away from the subject. The studies were all conducted indoors during the day-time, with a varying amount of sunlight as the only source of illumination. The authors' hypothesis was that volumetric changes in the facial blood vessels during the cardiac cycle modify the path length of the incident ambient light. The RGB colour sensors in the webcam amplify a mixture of the reflected plethysmographic signal along with other sources of fluctuations in light due to artefacts caused by motion and changes in ambient light conditions.

In these two papers from MIT, a rectangular bounding box was set up covering most of the skin area of the face, the ROI then being defined using the full height of the box but only 60% of the width. The time series for the three colours were decomposed into three independent source signals using the JADE algorithm for Independent Component Analysis (ICA). An FFT was then applied to the strongest plethysmographic component (over 30 s windows) and the largest spectral peak in the region from 0.75 to 4 Hz was deemed to be the cardiac frequency.

In the second paper (Poh *et al* 2011), the respiratory rate of the volunteers was estimated with a well-known indirect method (Sayers 1973) based on heart rate variability (HRV). The peaks of the PPG waveform (corresponding to the dominant ICA component) were identified to derive a time series of inter-beat intervals. The spectrum of this unevenly-sampled time series was computed using the Lomb periodogram (Moody 1993) with the low-frequency (LF) component and the high-frequency (respiratory) component calculated as the areas under the curve within the 0.04–0.15 Hz and 0.15–0.4 Hz bands, respectively. Scatter plots of heart rate and respiratory rate for the data acquired from the volunteers demonstrated good correlations between the webcam-derived estimates and those from a finger probe (for heart rate) and a chest belt for respiratory rate.

Respiratory rate estimation from HRV works well in healthy young volunteers, but is much less likely to give accurate results in elderly subjects, especially those with chronic diseases, most of which depress autonomic function (Dietrich *et al* 2006). However, breathing is associated with movement of the upper thorax and regions of the face. The changes in the amplitude of the PPG waveform caused by breathing-synchronous motion can be extracted though band-pass filtering and spectral analysis (see section 3.6). In addition, respiratory rate can also be estimated directly through motion-tracking techniques, well known in the computer vision community. A second MIT group has been developing Eulerian video magnification techniques to track and amplify the motion-related changes caused by breathing in the videos of human subjects, including neonates, recorded under normal lighting conditions (Wu *et al* 2012).

Interestingly, the latest paper from a third MIT group (Balakrishnan *et al* 2013) hypothesizes that there are also motion-induced changes in the *cardiac*-synchronous changes in the reflected ambient light. These are thought to be caused by the cyclical movement of blood from the heart to the head via the carotid arteries giving rise to periodic head motion at the cardiac frequency. The cardiac-synchronous changes in the ambient light reflected from the face and measured remotely (up to a distance of 1m away) thus appear to be the sum of three components, each of which is caused by the inflow of blood into the head during the cardiac cycle: volume changes in the facial blood vessels (Poh *et al* 2010), changes in the colour of these vessels (Wu *et al* 2012), and ballisto-cardiographic changes (Balakrishnan *et al* 2013).

In parallel with this work in the US, Philips in Europe has developed similar methods based on FFT analysis of the remote PPG signal to estimate heart rate and on motion analysis to estimate breathing rate. This led to the release of a software application for the iOS operating system for the consumer market in 2011. There is also a family of patent applications which describe aspects of non-contact vital sign monitoring such as the identification of people in the image (Cennini and Jeanne 2010) or the automatic selection of the measurement zone (Jeanne *et al* 2010).

Further work from Philips researchers includes the use of multi-spectral illumination and adaptive filtering techniques to minimize the effect of motion artefacts (Cennini *et al* 2010). Related theses from members of the Video and Imaging Processing Group at Philips Research Laboratories in Eindhoven focus on the issues of motion estimation and compensation (Schmitz 2011, Sahindrakar *et al* 2011). The latest paper (de Haan and Jeanne 2013) reports on techniques to deal with the periodic motion arising from exercise. Colour difference ratios are used in an attempt to eliminate the specular reflection component and more generally the effect of motion.

There are many other papers in the literature which describe variations on the theme of remote PPG signal analysis to derive heart rate and occasionally respiratory rate, using for example principal component analysis (Lewandowska *et al* 2011), ICA and time-frequency analysis (Sun *et al* 2011) or the continuous wavelet transform (Bousefsaf *et al* 2013). However, since the early work of Wieringa *et al* (2005) and Humphreys *et al* (2007), with illumination at specific wavelengths, there appears to have been little progress in estimating oxygen saturation with non-contact methods and ambient light. A recent paper (Kong *et al* 2013) uses two CCD cameras, each with a narrow-band filter (at 660 and 520 nm) mounted in front of the camera, but this is a more complicated arrangement than the single camera set-up which is the focus of this paper. Scully *et al* (2012) report on $SpO_2$ measurements made by comparing the red and blue PPG waveforms in light reflected from a flash light but using a contact method, as the finger is placed on the digital camera in a mobile phone.

To summarize, most of the previous studies of non-contact video-based vital sign monitoring have been on healthy volunteers, sometimes in sunlight only, and have usually concentrated on heart rate estimation, with some attempts at deriving respiratory rate. In this paper, we describe novel algorithms which are robust even under strong fluorescent lights, and can be used to estimate the values of the three cardio-respiratory vital signs (heart rate, respiratory rate and oxygen saturation) in patients in a clinical environment.

## 3. Methods

### 3.1. Clinical study

Almost 50% of patients with end stage renal failure (ESRF) in the UK are supported with haemodialysis, a form of renal replacement therapy in which the blood is pumped from the

patient through an extracorporeal filter to remove fluid and waste products and replace some electrolytes. The treatment is usually given in three sessions a week, each session lasting up to 4 h during which two litres of fluid are typically removed.

Patients with ESRF often have multiple co-morbidities such as diabetes and may suffer from associated conditions such as obstructive sleep apnoea. Hence their vital signs (blood pressure, heart rate, respiratory rate and oxygen saturation) will often be highly variable during a dialysis session (Borhani *et al* 2010) and dialysis therefore represents an ideal testing ground for developing clinically relevant non-contact vital sign monitoring techniques.

Our non-contact vital-sign monitoring study took place in the Oxford Kidney Unit, at the Churchill Hospital, which provides care for over 500 patients at any one time. The camera sub-study was part of a clinical study (Research Ethics Committee reference number 11/SC/0207) in which vital signs were recorded with conventional patient monitors, the aim being to understand the changes in physiological variables which occur during dialysis (Borhani *et al* 2010). Patients were given information leaflets approved by the research ethics committee and were recruited and consented by clinical staff.

A non-invasive, multi-parameter recording device (Equivital™ Vital Signs Physiological Monitor EQ-02, Hidalgo, Cambridge, UK) provided the conventional vital sign monitoring. The EQ-02 belt incorporates a thoracic expansion sensor to record the changes in chest circumference during respiration using a conductive material within the belt, the electrical impedance of which increases with length. Software provided by the manufacturer produces a respiratory rate estimate every 5 s, together with a signal quality index based on how sinusoidal the thoracic expansion sensor signal is, giving an indication of the validity of the estimate. A pocket within the belt holds the battery-powered recording device, onto which other monitors, such as a pulse oximeter and an oscillometric blood pressure monitor, are attached. The pulse oximeter (Onyx II Bluetooth Pulse Oximeter Model 9560, Nonin Medical, Plymouth, MN, USA) records a 4 beat average heart rate and $SpO_2$, both at 3 Hz.

The patients in the video camera sub-study were also monitored with a high-quality 5 megapixel camera (Grasshopper2 GigE Point Grey Research, Richmond, Canada), positioned approximately 1 m away from the patient, whose face is then in the focal plane of the camera (see figure 1). Raw uncompressed video data with 8 bits-per-pixel resolution was recorded at a sampling rate of 12 frames $s^{-1}$ (i.e. 36 bytes per pixel per second). A 4 h dialysis session therefore required a maximum of 2.4 terabytes of storage. The physiological and camera data were relayed to a workstation using the Network Time Protocol, with synchronicity up to 1 ms ensured at the start of each recording session. All recordings took place in a brightly-lit ward environment, with several sources of fluorescent light, as well as natural light.

## 3.2. Video pre-processing

An overview of the processing steps involved in deriving estimates for the three vital signs from the video signals is shown in the block diagram of figure 2. The following sections will cover each of these steps in more detail.

The first step towards obtaining one-dimensional physiological signals for each colour channel is to select an area of exposed skin on the subject's face. The process of determining the location of the patient's face is challenging for video recordings in a real-life hospital environment. Patients are usually active during dialysis (for example, talking on the phone) and the clinical staff are constantly interacting with them. Many algorithms for face registration have been reported in the literature, for example (Li and Jain 2011), which are generic in nature and perform best with frontal images. In our implementation, 13 facial features (based on the eyes, nose and mouth) are located using face registration algorithms described in Everingham

**Figure 1.** Patient being monitored during a 4 h dialysis session in the Oxford Kidney Unit. The chest belt, with recording device, is clearly visible. There is also a pulse oximeter probe on the index finger of the right hand and a blood pressure cuff on the left arm. The digital video camera is highlighted with a red circle. In the picture, the patient's face is obscured to preserve anonymity.

*et al* (2006) for frontal and profile faces. The facial features are subsequently tracked over time using the Kanade–Lucas–Tomasi feature tracker (Sivic *et al* 2009).

The algorithms for estimating vital signs described in this paper not only depend on the location of the patient's face, but also on finding a background reference ROI. Once the location of the patient's face is computed for every video frame, algorithms for non-parametric Bayesian image segmentation (Orbanz and Buhmann 2008) are used to segment the image into three clusters: face, upper body and background (see figure 3 in section 4).

### 3.3. Spectral analysis using auto-regressive modelling

As with most of the previous work on the analysis of the non-contact PPG signal, we use spectral analysis to identify the frequencies of interest in the signal. However, rather than using Fourier transforms, we chose AR modelling to capture these frequencies. We have previously applied AR modelling to the problem of identifying frequencies in noisy physiological signals, for example the electroencephalogram (Pardey *et al* 1996) and the cardiotocogram (Cazares *et al* 2001). More recently, we have used this method to extract the respiratory rate information from PPG waveforms recorded with a pulse oximeter and finger probe (Fleming and Tarassenko 2007).

AR modelling looks for regular frequencies in a signal which is deemed to be stationary over the period of analysis (Takalo *et al* 2005). As explained in section 2, the dominant frequency in the reflected PPG waveform is the cardiac frequency, as a result of the changes in colour and volume of superficial vessels with each cardiac cycle (see section 3.5). Breathing-synchronous motion also modulates the amplitude of the PPG waveform, hence there is a
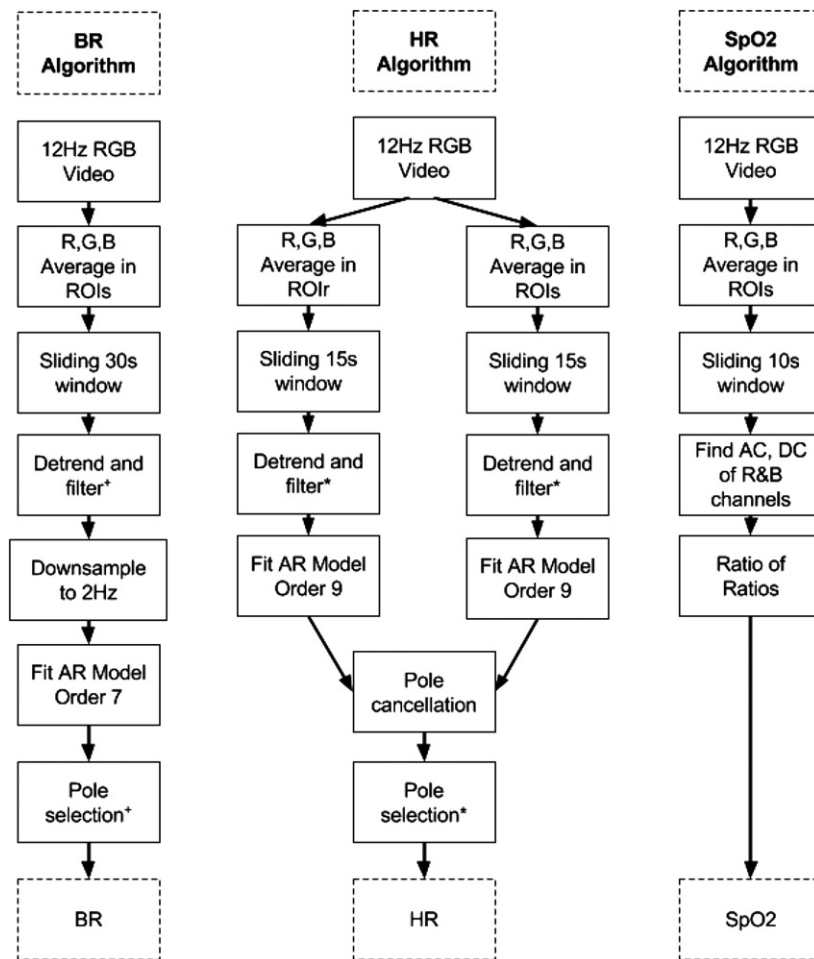
**Figure 2.** Block diagram detailing the three algorithms used to calculate heart rate (HR), breathing rate (BR) and SpO$_2$ from the video sequence. For the details of pole selection and filtering differences between HR($*$) and BR($+$), please refer to table 1.

respiratory peak in the PPG waveform spectrum as well, but of much lower amplitude than the cardiac peak (see section 3.6).

With an AR model, we assume that the value of the current sample, $x(n)$, is a linear combination of the $p$ previous values of $x(n)$ and the current value of $e(n)$, where $e(n)$ is a white-noise Gaussian distribution. Here $x(n)$ is the sampled value of the PPG time-domain signal, where $n$ is the sample number. Thus:

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + e(n) \tag{1}$$

where the summation is over $k$, from 1 to $p$ (the model order). $x$ is therefore a linear regression on previous values of itself and $e(n)$ is the error of the regression.

The AR model can also be re-formulated in terms of a system with a white-noise input $e(n)$ and an output $x(n)$. In the domain of the $z$-transform, the transfer function $H(z)$ relating the output to the input, can be written as:

$$H(z) = \frac{1}{\sum a_k z^{-k}}. \tag{2}$$

The denominator of $H(z)$ can be factored into the product of p terms:

$$H(z) = \frac{z^p}{(z - z_1)(z - z_2)\ldots(z - z_p)} = \frac{z^p}{P_1 P_2 \ldots P_p}. \tag{3}$$

$P_1, P_2, \ldots, P_p$ are vectors extending from any point $z$ in the complex plane to each of the $p$ poles of $H(z)$, $z_1, z_2, \ldots, z_p$. These poles, each of which is either real or one of a complex-conjugate pair, are the roots of the denominator of $H(z)$. The poles above the real axis correspond, in the frequency domain, to the spectral peaks of the signal being modelled, with higher-magnitude poles corresponding to higher-magnitude peaks. The frequency $f$ of each peak is given by the phase angle $\theta$ of the corresponding pole:

$$\theta = 2\pi f \Delta t \tag{4}$$

where $\Delta t$ is the sampling interval and $\theta$ is expressed in radians.

### 3.4. Pole cancellation

One of the challenges in making non-contact PPG imaging work in real-world settings is the presence of aliased components from artificial light (e.g. fluorescent lights), which can be found at frequencies similar to cardiac frequencies. This occurs because the image is sampled at the camera's frame rate (typically less than 30 Hz), which is much lower than the flicker frequency of the artificial light, 100 Hz in Europe and 120 Hz in the US (Rea 2000).

We are proposing a novel differential technique based on pole cancellation in the $z$-domain to suppress components unrelated to physiological information, i.e. to remove the aliased components in the frequency band of interest. The poles corresponding to the aliased components can be *cancelled* by removing them from the denominator. The new transfer function $H'(z)$, without these cancelled poles in the denominator, will now represent the transfer function for the light intensity reflected from the ROI (from the subject's face, for example) without the aliased components (Tarassenko *et al* 2013). The procedure is now explained in more detail in the next section in the context of heart rate estimation.

### 3.5. Heart rate estimation

(i) We firstly identify a ROI within the subject's face (such as the forehead or the cheek), to be called $ROI_s$.
(ii) We then identify a reference ROI outside the subject's face (such as the background wall), to be called $ROI_r$.
(iii) For each video frame, we calculate the average colour intensity, typically for the green colour, by spatial averaging over the ROI (typically $100 \times 100$ pixels) for $ROI_r$ and $ROI_s$.
(iv) We then construct a de-trended time series of these spatial averages, $x(n)$, for a time window, the duration of which may typically be 15 s for heart rate estimation, for both $ROI_s$ and $ROI_r$. A window length of 15 s corresponds to approximately 20 cardiac cycles, a sufficient number of cycles for accurate estimation, without introducing too long a processing delay.
(v) The windowed PPG waveform is band-pass filtered to enhance the frequency of interest. For heart rate estimation, the cut-off frequencies of the band-pass filter will typically be 0.7 and 5 Hz (corresponding to 42 and 300 beats min$^{-1}$). These cut-off limits represent the range of expected human heart rates.

(vi) We fit an AR model to the time series derived from ROI$_r$. At 12 frames s$^{-1}$ (which is the video sampling rate assumed throughout this paper) and for a window of 15 s, there are 180 samples from which to estimate the coefficients of the AR model in each window. The choice of model order is a compromise between the requirement to identify the dominant cardiac frequency (which favours a low model order), and the need to model the shape of the spectrum between the cardiac frequency and the half-sampling frequency (which favours a high model order). A model order of 9 was found to be a good compromise, as it allows a pole to be fitted to the second harmonic of the cardiac frequency, when the latter has sufficient energy (in sections of high-quality signals) or the noise spectrum can be modelled with higher-frequency poles (in section of low-quality signals). For a more detailed discussion of model-order selection, the reader is referred to Pardey *et al* (1996).

(vii) We then fit a separate AR model to the time series derived from ROI$_s$ in the same way as for ROI$_r$ in step 6.

(viii) We identify the poles in the AR model for ROI$_r$ which are the poles corresponding to the aliased components of the artificial light flicker frequency as these are also present in the AR model for ROI$_s$. The test of identity allows for these poles in ROI$_r$ and ROI$_s$ to be within *m* degrees of each other (*m* = 1 or 2, typically). Pole cancellation in ROI$_s$ gives the new AR model (heart rate information only), ROI$_k$. For example, if the aliased flicker causes two poles to be present at the same frequencies in both ROI$_s$ and ROI$_r$, these poles and their complex conjugates are cancelled in the transfer function for ROI$_s$, which means that ROI$_k$ is now a fifth-order model, given that ROI$_s$ is a ninth-order model.

(ix) The highest-magnitude pole between 0 Hz and the half-sampling frequency in the AR model for ROI$_k$ is the heart rate pole. Its angle corresponds to the heart rate in beats min$^{-1}$, the latter being obtained by multiplying $\theta$ by $60 f_s / 2\pi$, where $\theta$ is the angle in radians and $f_s$ is the sampling frequency in Hz. Note again that the poles below the horizontal axis in the pole-zero plot, which are disregarded in the analysis, are simply the complex conjugates of the poles above the axis (Takalo *et al* 2005).

(x) The radius of that pole (the distance to it from the centre of the pole-zero plot) is an indication of the amplitude of the heart rate component in the green channel for that window.

(xi) We slide the 15 s window by one second and repeat steps 1–10 for the new window. The use of a 1 s offset between consecutive windows allows us to derive heart rate estimates (based on the previous 15 s of data) every second.

### 3.6. Respiratory rate estimation

Low-frequency amplitude variations of the camera reflectance signal $x(n)$ from the subject region of interest ROI$_s$ are mainly caused by breathing-related motion. The size of the region of interest ROI$_s$ for respiratory rate estimation is usually smaller than that for heart rate estimation (see figure 4 later in the paper). The breathing-related amplitude variations are extracted with a separate band-pass filter (or low-pass filter, after de-trending), with an upper cut-off frequency, for normal breathing, of 0.7 Hz (corresponding to 42 breaths min$^{-1}$). The band-pass or low-pass filter requires a narrow transition band so that the cardiac-frequency component (at 1 Hz or above), which is a much stronger component in the camera reflectance signal, is eliminated by the filtering.

Once the low-frequency respiratory component has been obtained from the output of the band-pass filter, a separate AR model is applied to this time series, using a window of 30 s duration, to identify the 'respiratory rate pole'. However, even with the low frame rate of the

**Table 1.** AR model parameters for heart rate and respiratory rate estimation.

| | Heart rate | Respiratory rate |
|---|---|---|
| Frequency for analysis (Hz) | 12 | 2 |
| Cut-off frequencies (Hz) | 0.7–5 | 0.7 |
| Time window size (s) | 15 | 30 |
| AR model order | 9 | 7 |
| Pole cancellation range (°) | 2 | – |
| Criterion for selecting dominant pole | Highest magnitude pole whose angle is within $[0, \pi]$ | Lowest angle within the range equivalent to 0.1–0.7 Hz and with magnitude $> 0.95 \times$ magnitude of highest-magnitude pole |

camera (12 frames $s^{-1}$), downsampling is required. For a respiratory rate of 15 breaths $min^{-1}$ (i.e. 0.25 Hz), the phase angle $\theta$ of the respiratory rate pole will only be 0.3 radian (7.5°). To increase the angular resolution, the band-pass filtered video signal is downsampled to 2 Hz (after low-pass filtering to eliminate the effects of aliasing). This increases the angle of the pole corresponding to a respiratory frequency of 0.25 Hz (15 breaths $min^{-1}$) to $\pi/4$ radians or 45°.

As with the AR model for heart rate estimation, the model order for respiratory rate estimation, which is typically chosen to be 7, is a compromise between the two conflicting requirements: the need to identify the dominant respiratory frequency (which favours a low model order) and the need to model the shape of the spectrum between the respiratory frequency and the half-sampling frequency. We define a sector of interest on the pole-zero plot from 18° to 126°, i.e. 0.1 to 0.7 Hz (corresponding to a range of respiratory rates from 6 to 42 breaths $min^{-1}$). Candidate poles within the sector of interest are those whose magnitude is at least 95% of the magnitude of the highest-magnitude pole and the candidate pole with the lowest angle is selected as the respiratory rate pole. This is the method first described by Fleming and Tarassenko (2007) for estimating the respiratory rate from the PPG waveform recorded with a pulse oximeter and finger probe.

Table 1 is a summary of the parameters used with the heart rate AR model (middle column) and the respiratory rate AR model (right-hand column).

### 3.7. Maps of spatial distribution of heart rate and respiratory rate information

As explained in section 3.1, the angle of the pole in an AR model gives the frequency of that spectral peak; the magnitude (i.e. its radius, which can vary between 0 and 1) indicates the strength of that frequency component. The dominant pole in the cardiac AR model (order 9) and the dominant pole in the respiratory AR model (order 7), selected using the criteria shown in table 1, are the poles corresponding to the cardiac frequency and respiratory frequency, respectively. The magnitude of the dominant pole for each model in small regions of the video image (typically $25 \times 25$ pixels) can be displayed graphically across the entire image (for example, using colour coding) to show the relative strength of that frequency component in different regions of the image (Villarroel *et al* 2013). This will allow us to determine whether the colour and volume changes during the cardiac cycle are uniform across the subject's face and in which areas of the face and neck is the breathing-synchronous motion most prominent.

### 3.8. Oxygen saturation

With conventional (transmission-mode) pulse oximetry, in which light is shone through a body segment such as a finger or earlobe, only that part of the signal directly related to the inflow of arterial blood into the body segment, the cardiac-synchronous pulsatile component, is used for the calculation of oxygen saturation. It is then assumed that the increase in attenuation of light is caused only by the inflow of arterial blood into the body segment, and the oxygen saturation of the arterial blood is then estimated from the relative amplitudes of the cardiac-synchronous pulsatile component at the two wavelengths. The approach, which is based on the Beer–Lambert law, is sometimes known as the 'ratio of ratios' method. The equation for oxygen saturation is given by:

$$\mathrm{SpO_2} = A - B\frac{(I_{ac}/I_{dc})_{\lambda_1}}{(I_{ac}/I_{dc})_{\lambda_2}} \tag{5}$$

where $A$ and $B$ are empirically−determined coefficients, $I_{ac}$ and $I_{dc}$ are respectively the amplitudes of the pulsatile (ac) and dc components of the transmitted (or reflected) light at wavelengths $\lambda_1$ and $\lambda_2$. In conventional pulse oximetry, the two wavelengths $\lambda_1$ and $\lambda_2$ are usually chosen to be 660 nm (red) and 940 nm (near infra-red). Data from calibration experiments are used to derive population-based estimates for $A$ and $B$.

The ambient light used in our non-contact vital sign monitoring studies has most of its energy in the visible part of the spectrum, and visible light only has limited penetration in skin tissue. However, forehead vascularization is supplied from the supra-orbital artery and forehead reflectance pulse oximetry has demonstrated better correlation with *arterial* blood gas determination of oxygen saturation than standard finger-probe transmittance pulse oximetry in a number of studies (see for example Yönt *et al* (2011) and Nesseler *et al* (2012)). We therefore assume that the ac values in equation (5) will still depend on *arterial* oxygen saturation, even in our measurements of these quantities with the video camera 1 m away from the patient. A final limitation to note is that the three colour sensors in RGB cameras have overlapping broad-band, rather than narrow-band, spectral responses.

Despite these limitations, we analysed the sections in which the camera-derived estimates of heart rate, computed as described in section 3.5 above, agreed with the reference heart rate values to estimate the ac and dc quantities of equation (5) for the red and blue wavelengths. The dc value was obtained by calculating a 10 s moving average and the ac value was estimated by averaging the peak-to-trough heights in each 10 s window. As before, we processed the ROIs from the patient's faces during periods in the dialysis sessions for which patient motion was minimal. In each case, we estimated the $A$ and $B$ coefficients for each subject in equation (5) by obtaining 'ground-truth' $\mathrm{SpO_2}$ values from the pulse oximeter attached to the finger, and then finding the best-fit linear equation.

## 4. Results

### 4.1. Clinical study

A total of 46 patients had their vital signs double-monitored (conventional monitoring and camera recording) during 133 dialysis sessions in the Oxford Kidney Unit. The demographics and physical characteristics of these patients are given in table 2.

The main goal of our study was to capture the changes in vital signs which occur during a typical 4 h dialysis session without affecting patient care or behaviour. Patients were mostly awake throughout the dialysis sessions, although there were occasional periods of sleep. During wakefulness, there was no restriction on patient activity, which included
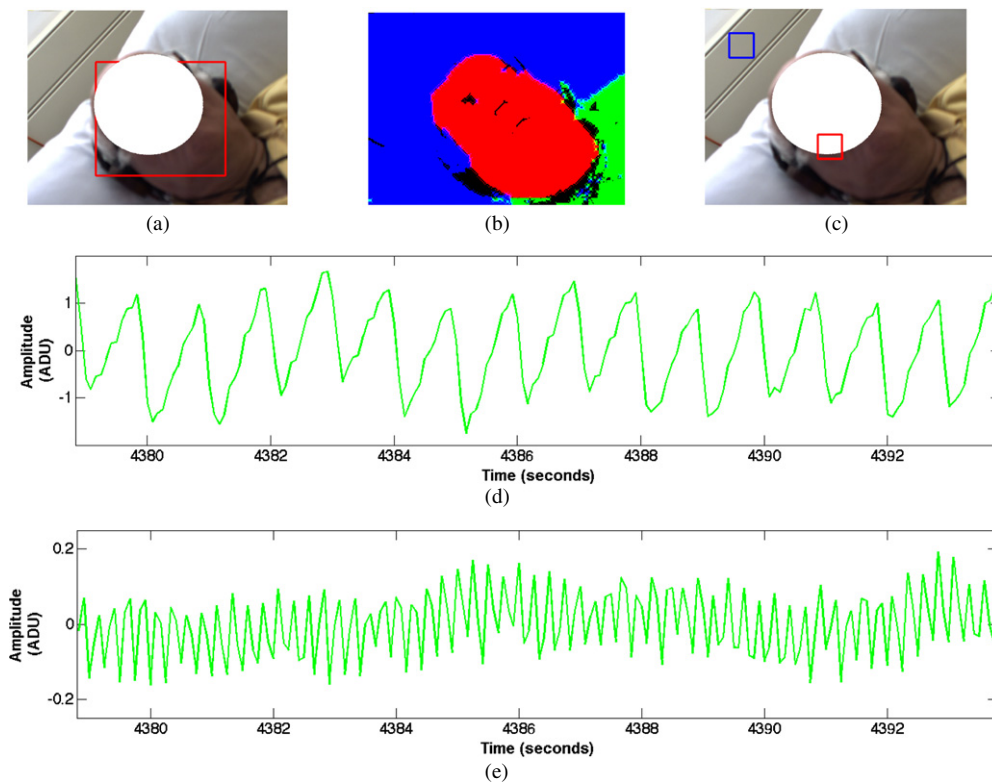
**Figure 3.** Region of interest selection: (a) original image frame, with the red square representing face registration; (b) computed image segments, red: face; blue: background; green: upper body; (c) regions of interest selected, red square: subject (ROI$_s$); blue square: background (ROI$_r$); (d) 15 s time series extracted for the green channel colour intensity from ROI$_s$; (e) 15 s time series extracted for the green channel colour intensity from ROI$_r$.

**Table 2.** Characteristics of dialysis patients in clinical study of non-contact vital sign monitoring. Where relevant, $\pm$ one standard deviation is given in brackets after the mean value for that variable.

| | |
|---|---|
| Age (yrs) | 64.7 ($\pm$15.3) |
| Gender (males) | 36 (78.3%) |
| Pre-dialysis weight (kg) | 78.8 ($\pm$17.1) |
| Post-dialysis weight (kg) | 77.0 ($\pm$16.8) |
| Height (cm) | 171.4 ($\pm$8.9) |
| Body Mass Index | 26.5 ($\pm$5.4) |
| Length on dialysis (months) | 30.2 ($\pm$23.3) |

talking (sometimes on mobile phones), listening to music or watching television. As a result, there were regular occurrences of patient motion, including changes of posture, which caused major discontinuities in the reflectance signals recorded by the video camera.

As outlined in section 3.2, two types of ROIs (see figure 3) were identified for every video frame.

- A ROI$_s$ on the patient's skin, preferably the face, for which the heart rate was estimated by analysing the varying intensity of the reflected light.
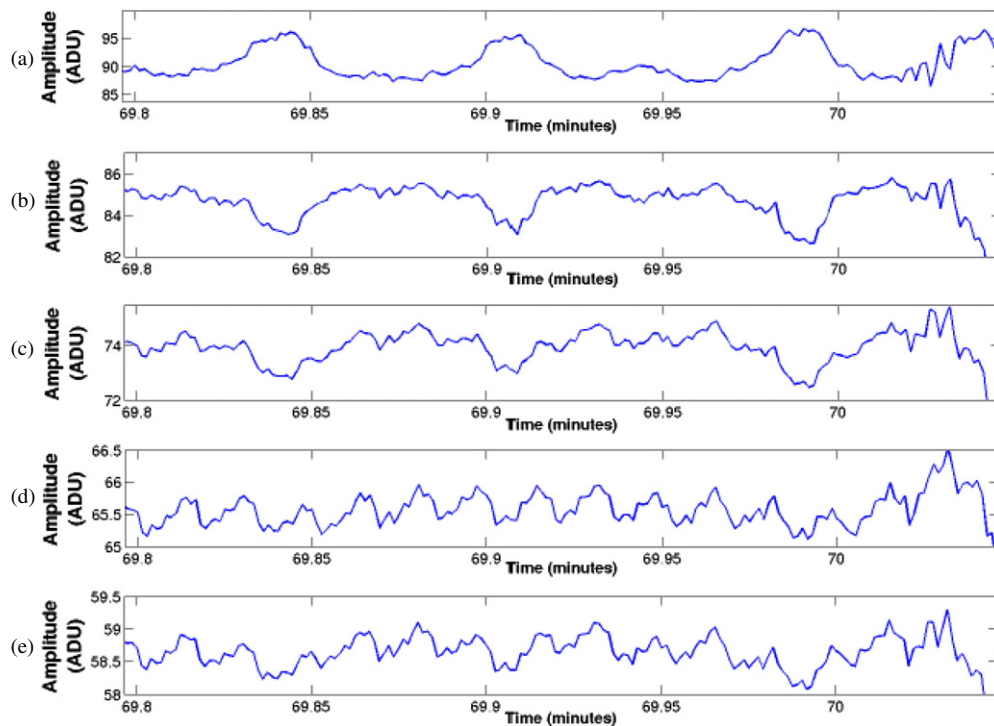
**Figure 4.** 15 s time series for the green channel colour intensity (recorded without any ambient light flicker) averaged from five regions of interest (ROI) with increasing pixel sizes: (a) $10 \times 10$, (b) $25 \times 25$, (c) $50 \times 50$, (d) $100 \times 100$ and (e) $150 \times 150$ pixels. The time on the horizontal axis represents the time since the start of dialysis. Heart rate can be extracted from the larger regions of interest, whereas changes in the reflected light intensity as a result of breathing are more prominent in the smaller regions of interest.

- A background region of interest ($ROI_r$), used to minimize the effects of external lighting sources such as fluorescent lights (see section 3.4).

The size of the ROI to be selected is dependent on the physiological information we wish to extract. Figure 4 shows the mean colour intensity for the green channel for different ROI sizes, from $10 \times 10$ to $150 \times 150$ pixels. Heart rate can be extracted from the bigger $ROI_s$, whereas changes in intensity due to breathing are more prominent in smaller $ROI_s$.

*Pole cancellation results.* Figure 5 shows the results of applying the pole-cancellation algorithm to a 15 s section of green-channel camera PPG data in which there is a strong aliased component at around 4 Hz as a result of artificial light flicker. The top two plots in the figure show the 15 s windowed time series for the $100 \times 100$ pixel ROI ($ROI_s$) recorded from the face of one of the subjects in the clinical study (green channel) and, just below it, the windowed time series for the equivalent ROI from the background behind the subject, the 'reference region' ($ROI_r$). In both cases, the time series are shown after de-trending, band-pass filtering (between 0.7 and 5 Hz) to enhance the cardiac information and windowing prior to spectral estimation. It is clear from inspection of the two time series that a common component dominates the time-domain signal; this component has 20 peaks between $t = 6125$ and $6130$ s, i.e. 4 peaks $s^{-1}$ or 4 Hz.

We fit ninth-order AR models to both $ROI_s$ and $ROI_r$. The pole-zero plots are shown on the right-hand side of the figure and the corresponding AR spectra on the left-hand side. In
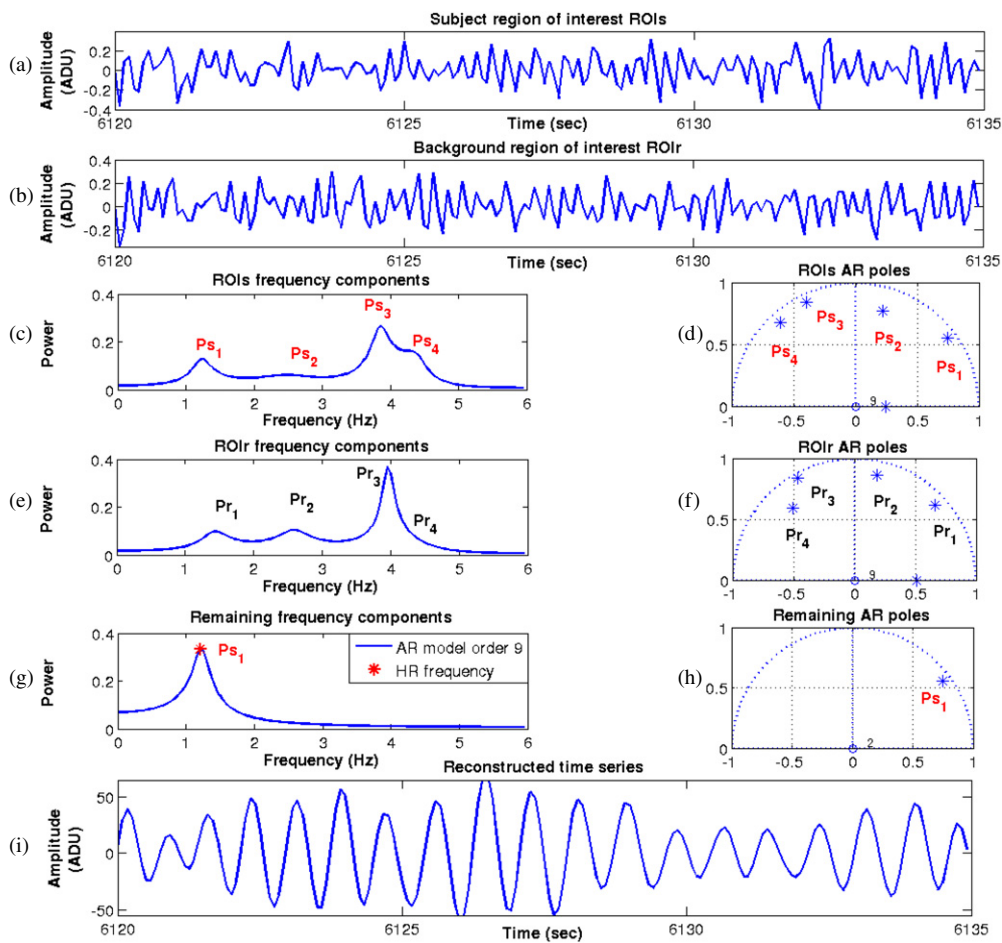
**Figure 5.** From the top: (a) 15 s time series for the green channel colour intensity reflected from the subject's face and (b) the equivalent time series for the light reflected from the background behind the subject. In each case, the time series has been de-trended, band-pass filtered and windowed. The figure then splits into two: on the right-hand side (from the top) are shown (d) the pole-zero plot for the time-series data from the subject's region of interest, (f) the pole-zero plot for the time-series data from the background region of interest and (h) the pole-zero plot (for the subject) after pole cancellation. On the left-hand side are shown, in (c), (e) and (g), the equivalent AR spectral plots (model order 9). The lowest plot (i) is the reconstructed waveform (at the cardiac frequency) using a finite white-noise sequence to drive the AR model after pole cancellation (model order 2).

both cases, there is a zero at (0.0), a pole on the real axis and four poles, corresponding to the four spectral peaks (the other four poles being the mirror images of these poles below the real axis). The strong aliased component has split into two poles, $p_{r_3}$ and $p_{r_4}$, in the reference AR model, and these are also found, at approximately the same angle, in the subject AR model ($p_{s_3}$ and $p_{s_4}$). There is another, lower-amplitude, aliased component at approximately 2.7 Hz in both models ($p_{r_2}$ and $p_{s_2}$). This only leaves the lowest-frequency pole, $p_{r_1}$ in the reference model, which occurs at approximately half the frequency of $p_{r_2}$ (sub-harmonic at 1.4 Hz), and $p_{s_1}$ in the subject model, at 1.2 Hz.
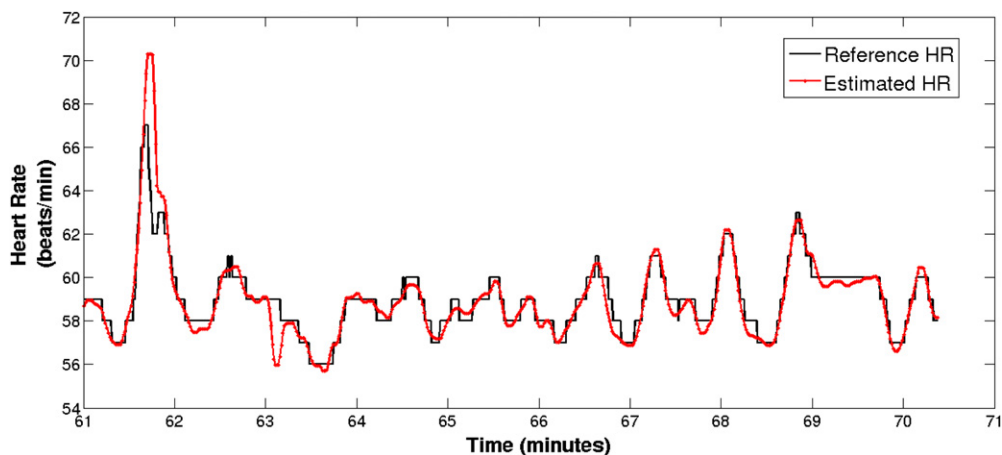
**Figure 6.** Heart rate estimates from a 10 min section of data recorded during a dialysis session. The second-by-second estimates derived from the video camera PPG data (green channel) as explained in the text are shown in red; the synchronized 4 beat averages obtained from a standard finger probe and pulse oximeter (reference values) are shown in black.

Hence $p_{s_4}$, $p_{s_3}$ and $p_{s_2}$ can be cancelled in the subject model, leaving only $p_{s_1}$ as the 'cardiac pole', corresponding to a frequency of 1.2 Hz, i.e. a heart rate of 72 beats min$^{-1}$. Finally, the lowest plot in the figure shows the time-domain waveform reconstructed from the *z*-domain transfer function for the AR model of order 2 corresponding to the original transfer function after pole cancellation in the denominator. It is clear that this is a signal at the cardiac frequency, which could not be seen in the original time-domain signal at the top of the figure, even though the latter is shown after band-pass filtering between 0.7 and 5 Hz. The signal is not a pure sinewave because the AR model is driven by a finite-length Gaussian noise sequence.

*Heart rate results.* The angle of the dominant pole in figure 5 ($p_{s_1}$) gives the cardiac phase angle for that 15 s section of video camera data. This angle is computed from the AR model for each consecutive window (with each window being offset from the previous one by 1 s).

Figure 6 shows, in red, the heart rate estimates derived second-by-second for a 10 min section of reflectance PPG data (green channel) for a typical patient in our clinical study. These estimates can be visually compared to the 4 beat average (reference) heart rate (in black) obtained from the pulse oximeter attached to the patient's right index finger. For sections such as this one, during which there is little patient movement, the correlation between the two heart rate values is very high. In fact, the mean absolute error (MAE) between the two sets of estimates (approximately 3 beats min$^{-1}$) is similar to the MAE between the heart rate values derived from the pulse oximeter/finger probe combination and from the pulse oximeter/earlobe probe combination (Villarroel 2014).

The lower plot in figure 7 shows the camera-derived estimate of heart rate (in red) superimposed on the values obtained with the pulse oximeter (in black) for the entire dialysis session for that patient (just under 4 h). The upper plot shows the intensity values for the 100 × 100 pixel ROI (green channel) from the patient's ROI. The lower plot provides visual evidence of good agreement between camera-derived estimates and reference values, whenever the upper plot indicates a steady-state section with little or no patient movement, as between $t =$ 61 and 71 min (the 10 min section shown in figure 6). In contrast, when there is a period of
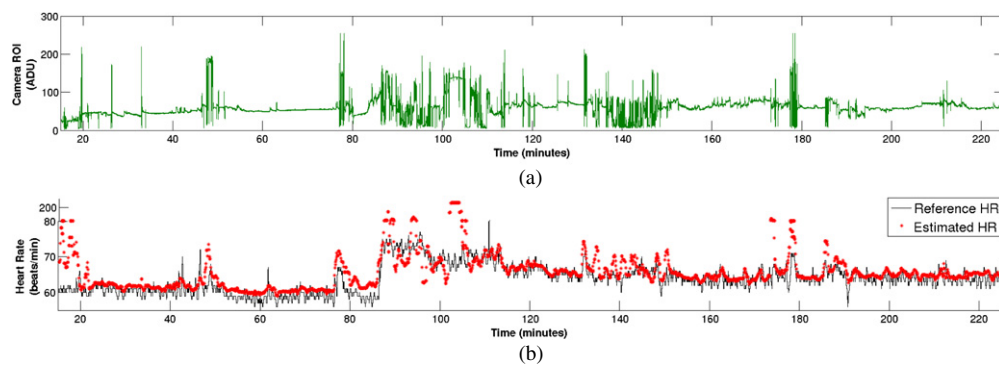
**Figure 7.** Heart rate estimates for an entire dialysis session. Top: intensity values in analogue-to-digital units (ADU) for the reflected PPG signal from the region of interest in the patient's face. Bottom: in red, second-by-second estimates derived from the camera PPG data (green channel); in black, the 4 beat averages obtained from a standard finger probe pulse oximeter (reference values).

high patient activity, such as occurs for the half-hour period between $t = 85$ and 115 min, the camera-derived estimates are no longer accurate.

*Respiratory rate results.* The pre-processing carried out to enhance the breathing-synchronous changes in the camera reflectance signal is similar to that applied to extract the heart rate: de-trending, band-pass filtering and windowing, except that the pass-band of the filter extends from 0.1 to 0.7 Hz and the window is twice the length of that used for heart rate estimation (i.e. 30 s). In addition, the pre-processed green reflectance signal is downsampled to 2 Hz to increase the angular resolution in the pole-zero plot.

For each 30 s window, we identify candidate poles whose magnitude is at least 95% of the magnitude of the highest-magnitude pole. The candidate pole with the lowest angle is selected as the respiratory rate pole (see section 3.6). The angle of this pole (in radians) is then converted to a respiratory rate in breaths min$^{-1}$ by multiplying it by $60 f_s / 2\pi$. The respiratory rate pole is thus estimated from the seventh-order AR model for each consecutive 30 s window (with each window being offset from the previous one by 5 s).

Figure 8 shows, in red, the respiratory rate estimates derived every 5 s from the breathing-synchronous changes in amplitude of the reflectance PPG signal (green channel), for the same 10 min section as in figure 7. These estimates, in red, follow broadly the respiratory rate estimates derived from the chest belt (in black). At this point, the patient is asleep and experiencing an episode of periodic hypoventilation, as indicated by the estimate from the chest belt decreasing from about 14 to 9–10 breaths min$^{-1}$ on two occasions, at approximately 62 and 63 min. In fact, it can be argued that the camera-based estimates are more accurate than those from the belt as they continue to track the cyclical pattern associated with periodic hypoventilation, albeit at a reduced amplitude, throughout the 10 min section.

This is confirmed by figure 9 which shows a time plot of both sets of respiratory estimates (again in red and black—lower plot) for the entire dialysis session, with the green-channel reflectance PPG signal at the top and the heart rate estimates in the middle plot for comparison. Outliers in the camera-derived estimates of respiratory rate are removed whenever they are more than two standard deviations outside a 10 min running mean of the valid respiratory rate estimates; similarly outliers in the belt-derived estimates are not shown if they are more than
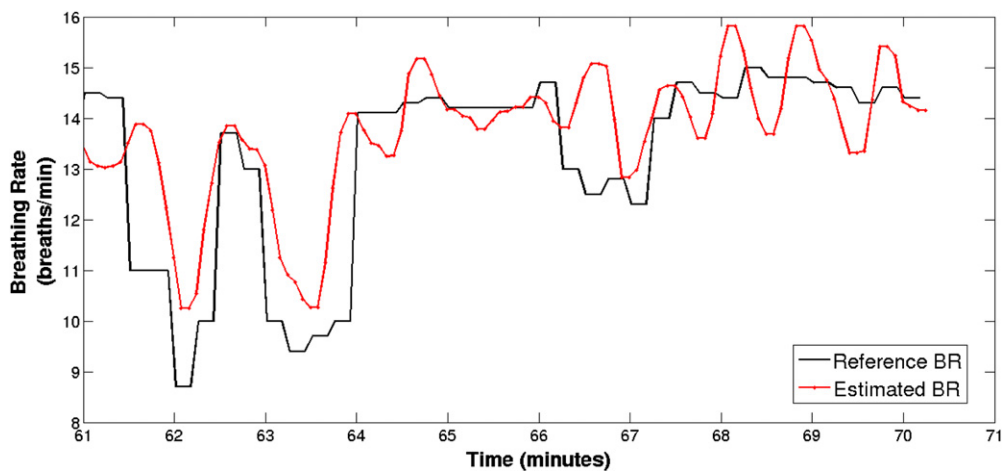
**Figure 8.** Respiratory rate estimates from a 10 min section of data recorded during a dialysis session. The estimates derived from the breathing-synchronous changes in the amplitude of the reflected PPG signal (green channel from the RGB camera) are shown in red; the respiratory rate estimates derived from the thoracic expansion sensor in the chest belt are shown in black.
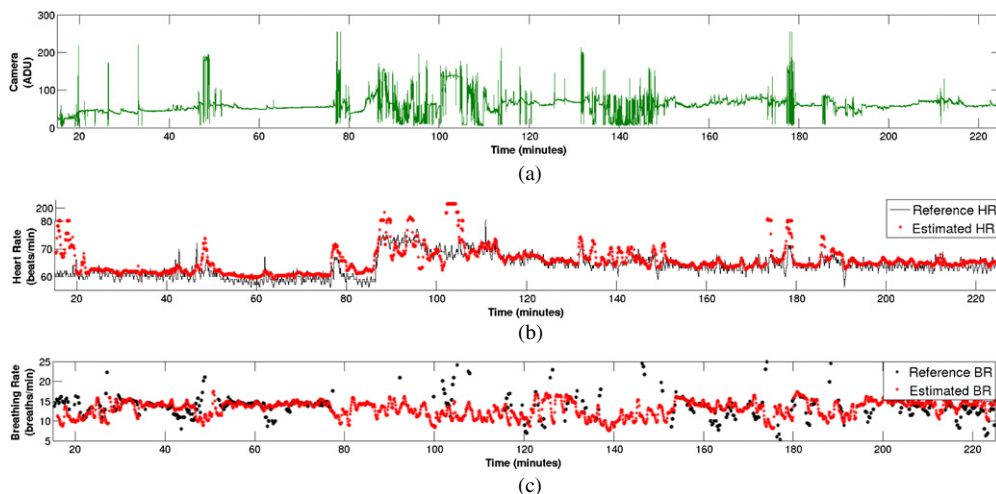


**Figure 9.** Respiratory estimates for an entire dialysis session. (a) Intensity values for the reflected green signal from the region of interest in the patient's forehead, (b) heart rate estimates, (c) in red: estimates derived from the breathing-synchronous changes in the amplitude of the reflected PPG signal (green channel); in black: the estimates derived from the chest belt.

two standard deviations outside a 10 min running mean of valid estimates or the belt signal quality index provided by the manufacturer is below threshold (also set by the manufacturer).

Inspection of the lower plot of figure 9 reveals that the camera-derived estimates of respiratory rate are much more consistent that those derived from the thoracic sensor in the belt. This is further demonstrated by the lowest plot in figure 10, for which respiratory rate estimates are only 'trusted' if the reference values are within two standard deviations of their 10 min running mean and the belt signal quality index is above threshold. This shows that
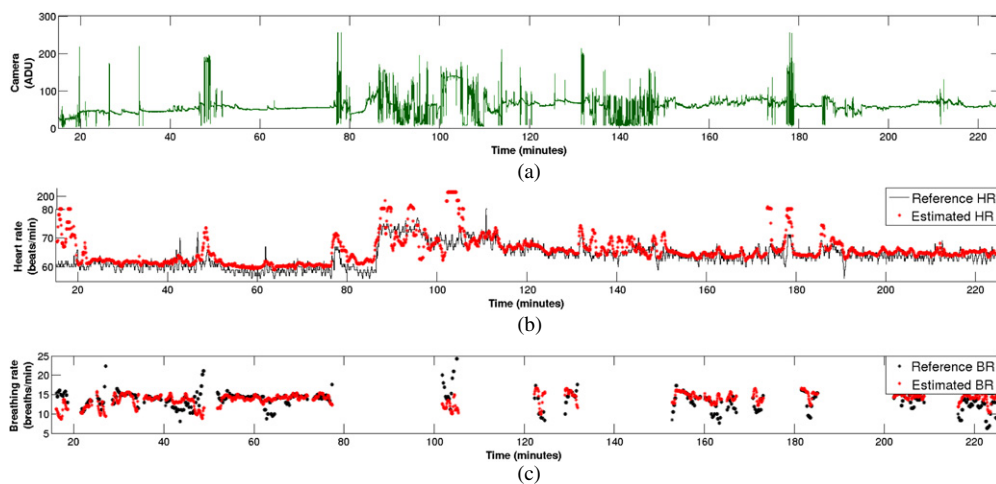
**Figure 10.** Same as figure 9, except that the respiratory rate estimates are only shown if the 'reference values' from the chest belt are within two standard deviations of their 10 min running mean and the manufacturer's signal quality index for the belt is above threshold. This leads to significant gaps in the record.
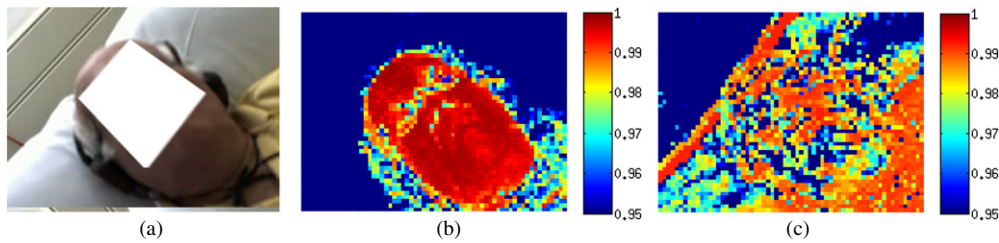


**Figure 11.** (a) Patient resting while listening to music through headphones during dialysis session; (b) image colour-coded according to strength of cardiac-frequency information (radius of cardiac AR pole); (c) image colour-coded according to strength of respiratory-frequency information (radius of respiratory AR pole).

there are multiple sections during which the belt-derived estimate cannot be considered to be a valid reference. This is not entirely surprising as the chest belt is affected by the patient's frequent changes in posture (from lying horizontally on one side to the other, to lying in the supine position or sitting upright). We would therefore argue that the respiratory rate estimates derived every five seconds from the breathing-synchronous changes in the amplitude of the non-contact PPG signal are more accurate than the so-called reference values.

*Heart rate and respiratory rate maps.* The maps representing the strength of the heart rate and respiratory rate frequency components for a typical patient in our clinical study are shown in figure 11. These maps have been constructed by identifying the dominant pole in the cardiac and respiratory AR models for all the $25 \times 25$ regions in the image and colour-coding the radius of the dominant pole as shown in the figure. Since these are dominant poles, the colour range only occupies a limited range of values, with pole radii lying between 0.95 and 1.0.

As expected, the cardiac-synchronous changes, as given by the spatial distribution of the highest-magnitude pole within the 0.7 to 5 Hz frequency band, are equally strong over the entire face, except around the eyes (the patient is wearing glasses) and nostrils. The

highest-magnitude respiratory pole, within the 0.1 to 0.7 Hz frequency band, can be found in the upper thorax region (as well as on the edge of the pillow—as a result of the patient's head movement with breathing). However, there are also high-strength poles in the patient's forehead and in regions close to the nose, indicating where the breathing information is present in the image of the patient's face.

*Oxygen saturation results.*    Oxygen saturation ($SpO_2$) is estimated by measuring the amplitude of the ac and dc components of the reflectance signal for the red and blue colour channels, as required by equation (5) (the ratio of ratios method). Building on the signal processing for heart rate estimation, we band-pass filter (from 0.7 to 5 Hz) the red and blue signals acquired from the subject's ROI. It is even more important for oxygen saturation estimation to find sections of camera data during which there is minimal patient movement. In addition, automatic camera parameters such as the white balance, the gain, the brightness and the camera shutter speed should ideally remain approximately constant as changes in these parameters may affect the intensity of the two colour channels differently over time. Finally, it is necessary to find sections in which there are changes in $SpO_2$, which means, in practice, identifying occurrences of hypoventilation in our dialysis patient population.

Figure 12 shows a stable 100 s section of data, centred on $t = 800$ s and indicated by the vertical red bars, during which a dialysis patient (different from the one whose data are shown in figures 6–10) has two desaturations below 90% either side of a lesser desaturation.

The panels in figure 13 indicate the results of processing the band-pass filtered signals at the red and blue wavelengths. The dc values are simply the average intensities of the raw signals for the two colour channels during each 10 s window. The ac values (averages of the peak-to-trough heights for each window and shown in the top two panels) are normalized by dividing them by the corresponding dc value for that colour channel, enabling the ratio of ratios of equation (5) to be computed. The *A* and *B* coefficients in that equation are estimated by obtaining the 'ground-truth' $SpO_2$ values from the pulse oximeter and then finding the best-fit linear equation for that 100 s section (lower left panel). The ratio of ratios over the 87%–95% $SpO_2$ range of that section is noisy (coefficient of determination, $r^2 = 0.64$), but its time evolution is clearly correlated with the changes in oxygen saturation during the 100 s. The close correspondence between the estimated values and the reference $SpO_2$ values from the pulse oximeter is shown in the bottom right-hand panel, where the values of the ratio of ratios over time are given by the magenta crosses and the grey waveform represents the $SpO_2$ values measured at the same time by the finger-attached pulse oximeter.

## 5. Discussion

Most of the previous work on non-contact video-based vital sign monitoring has tended to rely on normal ambient lighting (occasionally sunlight) rather than strong fluorescent lights as in clinical environments. All the work has also been carried out on healthy human volunteers under controlled conditions, over narrow ranges of heart rates and respiratory rates.

We have found in our clinical studies that the 100 Hz flicker frequency component from strong artificial lighting can be aliased down to (varying) frequencies which may be close to the heart rate. We have shown in this paper how the use of AR methods to model both a ROI (for example, the forehead) and the background can eliminate these unwanted components. A further advantage of AR models is that they are not affected by quantization as are discrete Fourier transforms; thus accurate estimates of heart rate and respiratory rate can be obtained from the models, unlike FFT-based methods which give significant quantization errors at
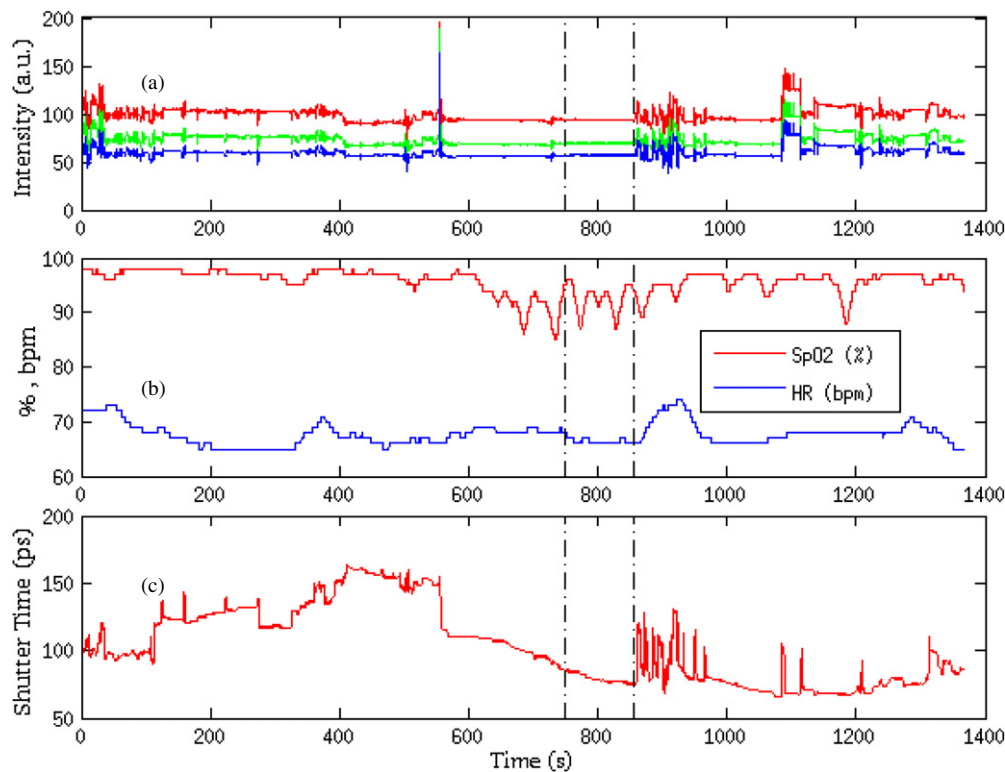
**Figure 12.** The top plot (a) shows the RGB reflectance data stream from a 100 × 100 pixel region of interest recorded during one of the dialysis sessions. The vertical black lines either side of $t = 800$ s indicate a 100 s stable section during which the patient experiences three cyclical desaturations; see $SpO_2$ values from the pulse oximeter attached to the finger, in the middle plot (b). The lower plot (c) also indicates that the shutter time is relatively constant during these 100 s, and hence there is no real change of exposure during that time. The baseline value for the shutter time was 2 147 486 ns. The offset from this baseline value is shown on the vertical axis.

typical frame rates. It should be remembered, however, that both AR models and FFT-based methods rely on the use of sliding windows. Hence there is an in-built latency in the estimation of vital sign values: a delay of 7.5 s for the estimation of heart rate as a result of using 15 s windows, and a delay of 15 s for the estimation of respiratory rate with a 30 s window. In some clinical scenarios, for example the detection of apnoea, the effect of this in-built delay will need to be taken into account.

Our analysis of the reflectance signals, mainly from the green channel, leads us to agree with the latest hypotheses that the camera PPG waveform, which is predominantly a cardiac signal, arises from cardiac-synchronous colour changes combined with blood volume pulsations. We have also observed, however, that a small proportion of the cardiac-synchronous signal is due to the motion of face landmarks in time with the heart beat (not shown).

Most of the authors in the literature have used HRV to estimate respiratory rate in healthy young adult volunteers. This is not likely to translate well to typical hospital or chronic disease patients, who are elderly and have multiple co-morbidities. We therefore focused on extracting the breathing-synchronous changes in PPG amplitude, again in (smaller) ROIs of the green reflectance signal. We showed that this approach led to respiratory rate estimates which were
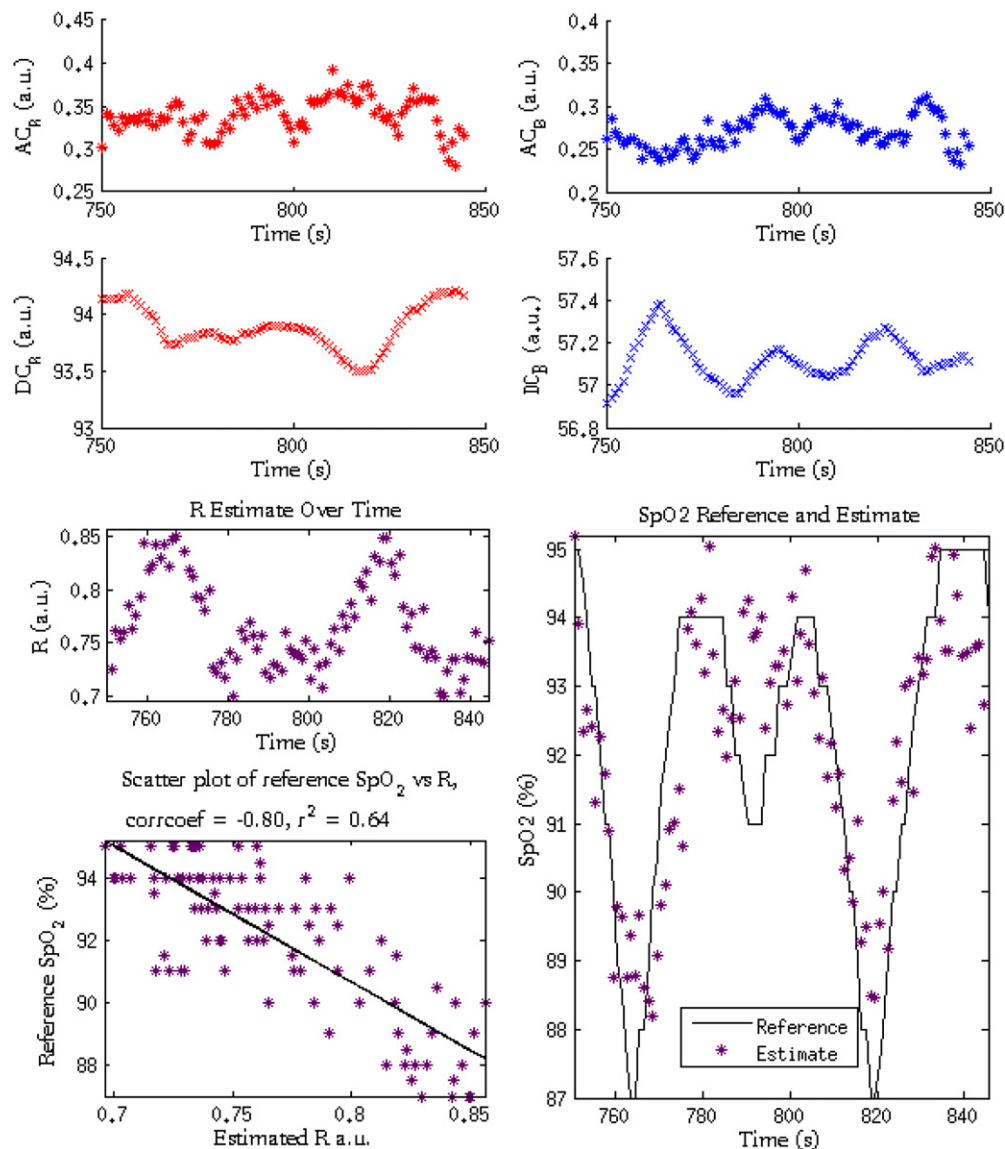
**Figure 13.** The various steps required to estimate oxygen saturation from the changes in reflectance at the red and blue wavelengths, for the 100 s section of camera data indicated by the vertical red lines in figure 12. The top two panels show the ac values (computed over sliding 10 s windows) for the two wavelengths, with the two panels below showing the corresponding dc values. From these the ratio of ratios of equation (5) can be derived and correlated, for that 100 s section, with ground truth from the pulse oximeter. The bottom-right panel shows that the calibrated ratio of ratios is able to track the changes in SpO$_2$ over time.

more accurate than the so-called reference values (derived from a thoracic expansion sensor in a chest belt) in our clinical studies.

There is a further factor which needs to be considered here: the activities of daily living modify the very parameter (respiratory rate) which is to be estimated. For example, the

activities of engaging in phone conversation, moving or and interacting with the clinical staff can all change the value of the patient's respiratory rate, for variable amounts of time (for example, because of the influence of exertion on the need for gas exchange). The advantage of having camera data for estimating respiratory rate is that it provides knowledge of the patient's behaviour (and hence of the possible effect on respiratory rate), unlike the traditional methods of vital sign monitoring using wearable sensors.

The respiratory rate maps tended to confirm the hypothesis that motion is the main cause of the breathing-synchronous changes in the reflectance PPG signal recorded by the remote camera: the highest-magnitude respiratory pole can be found in the upper thorax region (as well as on the edge of the pillow). However, there are also high-strength poles in the patient's forehead and in regions close to the nose. It may also be the case that there is some influence by the phenomena which give rise to the amplitude modulation of the contact PPG waveform (Meredith *et al* 2012). Firstly, inspiration results in a momentary reduction in stroke volume and hence a corresponding reduction in cardiac output, which in turn will reduce the amplitude of the PPG waveform. Secondly, there will be tissue blood volume changes during the respiratory cycle as a result of the changes in thoracic pressure: a reduction in intra-thoracic pressure during inspiration is transmitted through the venous system, effectively siphoning blood from the vascular bed within the tissue.

The validation of any new method of deriving oxygen saturation is hampered by the difficulties associated with trying to obtain $SpO_2$ values over the most clinically relevant range, from 80% to 95%. The significant minority of male dialysis patients with intermittent hypoventilation provided us with desaturation data which enabled us to circumvent this problem. We showed that adaptation of the method of ratio of ratios does have promise for the estimation of $SpO_2$ from camera-derived signals using two colour channels but the ac component amplitude is only marginally greater than the camera noise level. This has led us to investigate the possibility of using higher-resolution camera (12 bits) for our next clinical studies.

The issue of $SpO_2$ calibration remains a challenge fully to be solved. Angle changes caused by subject motion, as well as changes in the spectrum of the incident illumination, will affect the amplitude of the ac signal for each of the two colour channels differently, and it is not clear yet how this issue may be resolved. There may be alternative methods of processing the reflectance signals from each channel, and most clinical scenarios only require the detection of *changes* in $SpO_2$ values rather than an accurate knowledge of the absolute value of oxygen saturation at any one time.

Throughout this study, the red and blue channels have been used to detect changes in $SpO_2$, as the difference between the average absorptions of oxy- and deoxy-haemoglobin is greatest within the range of wavelengths that characterize the spectral responses of these channels. Nonetheless, future work in this field is likely to benefit from the additional use of the dc and ac values of the green channel, which could potentially be used as a normalizing factor.

## 6. Conclusion

In this paper we have described the results from the use of our novel methods for non-contact vital sign monitoring in a clinical environment. There are specific challenges which we have had to overcome in order to acquire non-contact vital sign data from patients in the clinic who are double-monitored, rather than from healthy volunteers. Full quantitative results for heart rate and respiratory rate for all the patients in our dialysis study, against reference values, will be presented elsewhere (Villarroel 2014).

We are now extending the use of our novel non-contact sensing algorithms to other clinical studies where there is an advantage in non-contact estimation of vital signs, for example in the monitoring of the vital signs of premature infants in neonatal intensive or high-dependence care. In addition, cost-effective solutions for non-contact vital sign monitoring, cheaper than conventional contact monitoring, may be highly appropriate for patients being cared for in single hospital rooms or at home. Finally, non-contact vital sign monitoring could also be applied to other inter-disciplinary fields, such as human–machine interaction analysis.

## Acknowledgments

## References

Agashe G S, Coakley J and Mannheimer P D 2006 Forehead pulse oximetry: headband use helps alleviate false low readings likely related to venous pulsation artifact *Anesthesiology* **105** 1111–6

Balakrishnan G, Durand F and Guttag J 2013 Detecting pulse from head motions in video *IEEE Conf. on Computer Vision and Pattern Recognition (Portland, OR)* pp 3430–7

Borhani Y, Fleming S, Clifton D A, Sutherland S, Hills L, Meredith D, Pugh C W and Tarassenko L 2010 Towards a data fusion model for predicting deterioration in dialysis patients *Computing in Cardiology (26–29 September 2010)* pp 967–70

Bousefsaf F, Maaoui C and Pruski A 2013 Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate *Biomed. Signal Process. Control* **8** 568–74

Cazares S, Moulden M, Redman W G and Tarassenko L 2001 Tracking poles with an autoregressive model: a confidence index for the analysis of the intrapartum cardiotocogram *Med. Eng. Phys.* **23** 603–14

Cennini G, Arguel J, Akit K and van Leest A 2010 Heart rate monitoring via remote photoplethysmography with motion artifacts reduction *Opt. Express* **18** 4867–75

Cennini G and Jeanne V 2010 Method of controlling a function of a device and system for detecting the presence of a living being *WO Patent Application* 2010100593

Chan E D, Chan M M and Chan M M 2013 Pulse oximetry: understanding its basic principles facilitates appreciation of its limitations *Respiratory Med.* **107** 789–99

Cretikos M A, Bellomo R, Hillman K, Chen J, Finfer S and Flabouris A 2008 Respiratory rate: the neglected vital sign *Med. J. Aust.* **188** 657–9

de Haan G and Jeanne V 2013 Robust pulse rate from chrominance-based rPPG *IEEE Trans. Biomed. Eng.* **60** 2878–86

Dietrich D *et al* 2006 Heart rate variability in an ageing population and its association with lifestyle and cardiovascular risk factors: results of the Sapaldia study *Europace* **8** 521–9

Everingham M, Sivic J and Zisserman A 2006 'Hello! My name is... Buffy'—automatic naming of characters in TV video *Proc. British Machine Vision Conf.* pp 889–908

Fleming S G and Tarassenko L 2007 A comparison of signal processing techniques for the extraction of breathing rate from the photoplethysmogram *Int. J. Biomed. Sci.* **2** 232–6

Humphreys K, Ward T and Markham C 2007 Noncontact simultaneous dual wavelength photoplethysmography: a further step toward noncontact pulse oximetry *Rev. Sci. Instrum.* **78** 044304

Jeanne V *et al* 2010 Processing images of at least one living being *WO Patent Application* 2010100594

Kong L, Zhao Y, Dong L, Jian Y, Jin X, Li B and Feng Y 2013 Non-contact detection of oxygen saturation based on visible light imaging device using ambient light *Opt. Express* **21** 179–83

Lewandowska M, Ruminski J, Kocejko T and Nowak J 2011 Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity *FedCSIS: Federated Conf. on Computer Science and Information Systems* pp 405–10

Li S Z and Jain A K 2011 *Handbook of Face Recognition* (Berlin: Springer)

Meredith D J, Pugh C W, Clifton D, Charlton P, Brooks J and Tarassenko L 2012 Photoplethysmographic derivation of respiratory rate: a review of relevant physiology *J. Med. Eng. Technol.* **36** 1–7

Moody G 1993 Spectral analysis of heart rate without resampling *Proc. Computers in Cardiology Conf.* pp 715–8

Nesseler N, Frénel J-V, Mallédant Y, Morcet J, Seguin P and Launey Y 2012 Pulse oximetry and high-dose vasopressors: a comparison between forehead reflectance and finger transmission sensors *Intensive Care Med.* **38** 1718–22

Orbanz P and Buhmann J 2008 Nonparametric Bayesian image segmentation *Int. J. Comput. Vis.* **77** 25–45

Pardey J, Roberts S and Tarassenko L 1996 A review of parametric modelling techniques for EEG analysis *Med. Eng. Phys.* **18** 2–11

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

Poh M-Z, McDuff D J and Picard R W 2011 Advancements in noncontact, multiparameter physiological measurements using a webcam *IEEE Trans. Biomed. Eng.* **58** 7–11

Rea M S 2000 *The IESNA Lighting Handbook: Reference and Application* (New York: Illuminating Engineering Society)

Sahindrakar P, de Haan G and Kirenko I 2011 Improving motion robustness of contact-less monitoring of heart rate using video analysis *Master's Thesis* Department of Mathematics and Computer Science, Eindhoven University of Technology

Sayers B 1973 Analysis of heart rate variability *Ergonomics* **16** 17–32

Schmitz G 2011 Video camera based photoplethysmography using video camera based photoplethysmography using ambient light *Master's Thesis* Department of Mathematics and Computer Science, Eindhoven University of Technology

Scully C G, Member S, Lee J, Meyer J, Gorbach A M, Granquist-Fraser D, Mendelson Y and Chon K H 2012 Physiological parameter monitoring from optical recordings with a mobile phone *IEEE Trans. Biomed. Eng.* **59** 2010–3

Severinghaus J W and Honda Y 1987 History of blood gas analysis: VII. Pulse oximetry *J. Clin. Monit.* **3** 135–8

Sivic J, Everingham M and Zisserman A 2009 Who are you?—learning person specific classifiers from video *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 1145–52

Sun Y, Hu S, Azorin-Peris V, Greenwald S, Chambers J and Zhu Y 2011 Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise *J. Biomed. Opt.* **16** 077010

Takalo R, Hytti H and Ihalainen H 2005 Tutorial on univariate autoregressive spectral analysis *J. Clin. Monit. Comput.* **19** 401–10

Tarassenko L, Clifton D A and Villarroel M 2013 Remote monitoring of vital signs *WO Patent Application* 2013027027

Verkruysse W, Svaasand L O and Nelson J S 2008 Remote plethysmographic imaging using ambient light *Opt. Express* **16** 21434–45

Villarroel M 2014 Non-contact monitoring of vital signs in the clinic *PhD Thesis* Department of Engineering Science, Oxford University

Villarroel M, Clifton D A and Tarassenko L 2013 Method and system for signal analysis *Patent Application GB* 1302451.8

Wax D B, Rubin P and Neustein S 2009 A comparison of transmittance and reflectance pulse oximetry during vascular surgery *Anesthesia Analgesia* **109** 1847–9

Wieringa F P, Mastik F and van der Steen A F W 2005 Contactless multiple wavelength photoplethysmographic imaging: a first step toward $SpO_2$ camera technology *Ann. Biomed. Eng.* **33** 1034–41

Wu H-Y, Rubinstein M, Shih E, Guttag J, Durand F and Freeman W T 2012 Eulerian video magnification for revealing subtle changes in the world *Proc. SIGGRAPH; ACM Trans. Graph.* **31** 65

Yönt G H, Korhan E A and Khorshid L 2011 Comparison of oxygen saturation values and measurement times by pulse oximetry in various parts of the body *Appl. Nursing Res.* **24** e39–e43