Non-contact vital sign monitoring in the clinic

Mauricio Villarroel[1], João Jorge[1], Chris Pugh[2] and Lionel Tarassenko[1]
[1] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK
[2] Nuffield Department of Medicine, University of Oxford, UK

***Abstract*— Current monitoring systems available to track changes in the vital signs of patients (such as heart rate, respiratory rate or peripheral oxygen saturation) require contact with the subject. Most patients requiring regular monitoring find the probes difficult to wear over prolonged periods of time. Research in non-contact vital sign monitoring has recently expanded through the use of off-the-shelf video cameras; nevertheless, most of the current work in video-based non-contact vital sign monitoring has so far been performed over short time periods (typically up to a couple of minutes), under tightly controlled conditions with relatively still and healthy volunteer subjects.**

Using an off-the-shelf camera, we have been able to compute estimates of heart rate and respiratory rate, and also detect changes in peripheral oxygen saturation in a real hospital scenario, without interfering with regular patient care. Videos were recorded for 369.1 hours from 40 patients undergoing haemodialysis treatment in the Renal Unit of the Churchill Hospital in Oxford, UK. The mean absolute error between the heart rate estimates from the camera and the average from two reference pulse oximeters (positioned at the finger and earlobe respectively) was 2.8 beats per minute for over 65% of the time, which was comparable to the error between the two reference pulse oximeters. The mean absolute error between the respiratory rate estimates from the camera and the reference values (computed from the Electrocardiogram and a thoracic expansion sensor - chest belt) was 2.1 breaths per minute for over 69% of the time for which the reference signals were valid. By calibrating the camera data with the reference pulse oximeters, changes in peripheral oxygen saturation could also be tracked during time periods with minimal patient motion.

I. INTRODUCTION

The measurement of the standard vital signs such as heart rate (HR), respiratory rate (RR), peripheral oxygen saturation (SpO_2), blood pressure and temperature is a core component of the physical assessment of most patients [1]. Of these vital signs, heart rate, respiratory rate and SpO_2 are of main interest for this paper.

Heart rate is a measure of the rate at which the heart beats. An appropriate monitoring of the heart's pumping mechanism is of vital importance as with each heart beat, blood is sent throughout the body carrying gases, nutrients, hormones and other substances used in metabolic processes by the cells [2].

MV was supported by the Oxford Centre of Excellence in Medical Engineering funded by the Wellcome Trust and EPSRC under grant number WT 88877/Z/09/Z. The clinical study in the Oxford Kidney Unit was funded by the NIHR Biomedical Research Centre Programme, Oxford. JJ was supported by the RCUK Digital Economy Programme grant number EP/G036861/1 (Oxford Centre for Doctoral Training in Healthcare Innovation). We would like to thank all the patients from the Oxford Kidney Unit who agreed to take part in the clinical study, as well as Dr David Meredith and Ms Sheera Sutherland who carried out the study.

Respiratory rate is recognised as an important vital sign since it has been found to be predictive of lower respiratory tract infections [3], the evaluation of the severity of pneumonia [4], a risk factor for unplanned hospital readmissions [5] and mortality risk assessment for paediatric patients in intensive care units [6].

Oxygen is a chemical substance essential to the functioning of each cell in the human body and, therefore, necessary to sustain life. It is important to monitor if organs are receiving a sufficient supply of oxygen as it is being delivered to all the body parts. The measurement of blood oxygenation, also known as oxygen saturation, is an important indicator of a patient's health. A prolonged lack of oxygen can rapidly cause permanent damage to cell tissue, leaving patients with devastating neurological handicaps and has the potential to be life-threatening, if cells having high metabolic rate in organs such as the brain, heart or the central nervous system are damaged [2], [7].

Conventional patient monitoring systems require a probe to be attached to the patient, such as the finger or ear in a pulse oximeter or on the chest in an Electrocardiogram (ECG) monitor. These have the potential to cause skin irritation, increasing the risk of infection and increasing the costs of implementing and maintaining new technology. The ideal technology to estimate vital signs would involve sensors with no direct contact with the patient ("non-contact sensing"). This paper proposes algorithms for the remote monitoring of heart rate, respiratory rate and identifying changes in SpO_2 using a standard colour camera and ambient light.

The paper is organised as follows. A summary of the clinical study is presented in the next section, followed by the description of the proposed methods for non-contact vital sign monitoring. Subsequently, the evaluation of these algorithms against the reference signals are presented. Finally, the paper ends with a discussion of the applicability of non-contact vital sign monitoring in the clinic.

II. CLINICAL STUDY

The dataset for this paper was recorded from patients undergoing haemodialysis treatment in the Renal Unit of the Churchill Hospital in Oxford. The research application was submitted to the Oxford University Clinical Trials and Research Governance (CTRG) (reference number 11/SC/0207).

Dialysis is a process by which blood is removed from the patient, filtered, and then replaced back into the body. Haemodialysis, together with kidney transplant

Fig. 1. A typical dialysis data collection set-up with the red circle showing the location of the video camera.

and peritoneal dialysis, are collectively known as renal replacement therapy (RRT) methods. These methods are commonly used in the United Kingdom (UK) for the treatment of end stage renal disease (ESRD), the last stage of chronic kidney disease occurring when the kidneys can no longer meet the daily demands to remove waste products and water from the body [8].

According to the last report from the UK Renal Registry [9], there were 888 patients per million population receiving renal replacement therapy in 2013, a 69% increase from 2000. Although kidney transplant is the most common treatment (52%), haemodialysis accounts for 41.6% of the cases and peritoneal dialysis for the rest. ESRD is a predominantly adult disease, the median age of patients receiving RRT being 58.4 years (haemodialysis 66.9 years, peritoneal dialysis 63.7 years and transplant 52.8 years) [8].

One of the advantages of recording data from patients undergoing haemodialysis treatment is that, in a relatively short amount of time (a typical dialysis session lasts 4 hours), these patients experience a wide range of physiological values. This therefore helps to validate vital sign estimation algorithms.

Figure 1 shows the typical recording set-up during a dialysis session. Using technical information supplied by the manufacturer, custom software was developed for the real-time acquisition of video from a high-quality 5 megapixel camera (Grasshopper2 GigE Point Grey Research, Richmond, Canada), positioned approximately 1m away from the patient. Raw uncompressed video data with 8-bits-per-pixel resolution was recorded at a sampling rate of 12 frames per second.

Conventional monitoring was provided by two devices: a Bluetooth pulse oximeter (Model 4100, Nonin Medical, Plymouth, MN, USA) and Hidalgo's Equivital EQ02 LifeMonitor (Equivital^{TM}, Hidalgo, Cambridge, UK). These reference devices were chosen as the primary means to correlate the results of the analysis with the data extracted from the video recordings. The Nonin pulse oximeter, a device that is often used in patient care, was attached to the patient's finger tip recording a 4-beat average heart rate and SpO_2, both at 3Hz.

The Equivital EQ02 LifeMonitor is a FDA 510(k) certified ambulatory multi-parameter vital signs telemetry device intended for the monitoring of adults in hospital care facilities, the home, workplace, and alternate care settings [10]. It consists of a chest belt harness, ECG electrodes, a thoracic expansion sensor, and a separate pulse oximetry module which connects to the patient's ear lobe. It records a two-channel ECG at 256 Hz, a respiration signal at 25 Hz, and reports heart rate and SpO_2 estimates every 5 seconds.

As shown in table I, a total of 104 dialysis sessions were recorded from 40 patients. The total length of video for all sessions is 369.1 hours, with the average session lasting 3.5 hours. In dialysis studies, the patient population is usually comprised of elderly patients [11], [12], hence the average age for a patient in this study was 64.7 years. The majority of patients were males (78 %) with a mean body mass index (BMI) of 26.5. At the time of writing, out of the 40 patients, 18 have died (45%), 18 continue to receive haemodialysis treatment in the hospital (45%) and 4 patients have received a kidney transplant (10%).

III. Methods

A. Reference signals

As discussed in the previous section, the reference heart rate estimates were provided by two transmission-mode pulse oximeters, one located on the finger and the other on the ear lobe. Most of the studies reported in the literature use only one pulse oximeter as a reference device, recording the photoplethysmographic (PPG) waveform and heart rate estimates from a single body site. There are limited studies of multi-site PPG recordings. Allen et al [13] suggested that, by studying pulses obtained simultaneously from different sites, important information about the peripheral circulation can be analysed.

Using the estimates from more than one pulse oximeter introduces physiological and non-physiological factors to be considered that can affect the timing and values of the two recorded reference values. Consequently, a direct

TABLE I

Population characteristics summary for the study.

Item	Value
Total number of sessions	104
Total number of patients	40
Total video length	369.1 hours
Average video length of a session[1]	3.5 (±0.8) hours
Deceased patients[2] [3]	18 (45%)
Receiving haemodialysis treatment[2]	18 (45%)
Patients with kidney transplant[2]	4 (10%)
Age (yrs)[1]	64.7 (±15.3)
Gender (males)[2]	36 (78.3%)
Height(cm)[1]	171.4 (±8.9)
Dry Weight(kg)[1]	77.1 (±15.3)
Body Mass Index[1]	26.5 (±5.4)

[1] mean (± std)
[2] N (percentage from total number of patients)
[3] Number of patients who died in the curse of the study: 11

comparison (on a sample-by-sample basis) of the heart rate estimates between the camera and each of the two pulse oximeters can potentially be affected by factors not only caused by physiology but also by the recording set-up and each of the manufacturer's processing context, leading to incorrect analysis. Furthermore, the measurement of a physiological process implies some degree of error; when two sensing devices exist, neither provide an absolute correct measurement. Since the true value is not known, the mean of the two measurements is usually taken as a representative value [14].

The heart rate estimates from the two pulse oximeters were found to be within 2 beats per minute on average. Both time series have a high positive correlation and are in good agreement. Therefore, for the analysis in this paper, a reference heart rate time series is computed from the average of the reported values from the two pulse oximeters. This "ground truth" reference heart rate is used to compare the estimates derived from the camera.

Two simultaneous respiratory rate estimates are produced by the Hidalgo kit, the first one computed from the chest belt, and the second using the ECG. These estimates were found to have large errors unrelated to physiology and were not suitable as reference values to be compared with the respiratory rate estimates computed from the camera. Therefore, new respiratory rate estimates were computed from the ECG and chest belt using current state-of-the-art algorithms. Following published literature [15], [16], the "ground truth" reference respiratory rate was taken as the average of all the estimates from these algorithms that do not differ for more than 2 breaths per minute.

Similarly to the heart rate estimates, a "ground truth" reference $SpO2$ was computed from the average of the reported values from the two pulse oximeters.

B. Video analysis

Most of the previous work in non-contact vital sign monitoring has been performed over short time periods, under tightly-controlled conditions with healthy volunteers. While studies in controlled environments can potentially produce usable data, the robustness of algorithms for estimating the values of vital signs is challenged when processing video data recorded from subjects under real-life conditions.

Figure 2 shows some examples of typical patient behaviour during the recordings. One of the goals of our study was not to interfere with regular patient care. As a result, the video recordings were affected by several external sources of distortion, such as:

- Regular interaction between the patient and the clinical staff
- Camera angle changed to allow for patient care (see figure 2c)
- Patient changing position during the 4-hour recording (a regular occurrence as in figure 2b)
- Patient torso or head moving out of camera frame (figure 2d)

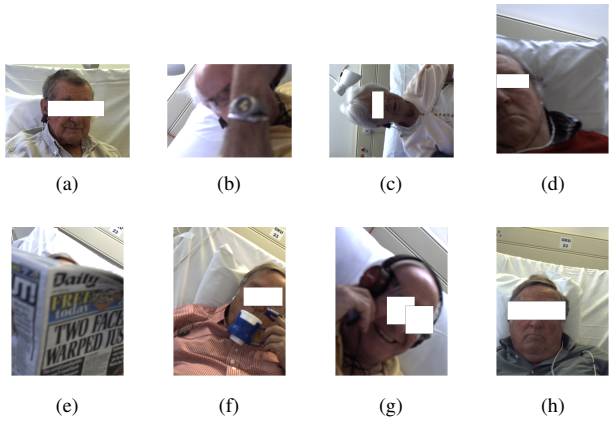

Fig. 2. Examples from typical dialysis sessions where the patient is: (a) correctly in frame, (b) moving freely with limbs obstructing the frame, (c) being filmed with the camera at an angle, (d) sleeping with torso half frame, (e) reading the newspaper and obstructing the frame, (f) having tea, (g) listening to music and speaking on a mobile phone, (h) sleeping and correctly in frame.

- The use of electronic devices such as mobile phones, music players or tablets (figure 2g)
- Regular consumption of beverages such as water or tea (figure 2f)
- Objects obstructing the patient's face such as newspapers (figure 2e), magazines or limbs (figure 2b)

To estimate physiological signals, suitable time periods need to be identified from the videos during which the patient is stable and in the frame. Examples of such periods are when the patient's torso is frontal to the camera (figure 2a), the patient is sleeping quietly (figure 2h) or when minimal motion occurs (figure 2c if the image is rotated to portrait mode).

The video analysis process starts with the task of detecting and tracking the patient's face. Subsequently, periods of high activity or motion are found by tracking the movements of the patient. The results of these tasks are combined to identify time periods within the video for which the location of the face is known and the patient is relatively still.

Several algorithms for face detection are reported in the literature using cues such as skin colour, facial or head shape, facial appearance, or a combination of more than one technique [17]. Although the problems have received a lot of attention, face detection and facial feature extraction are still challenging, especially when illumination, subject expression and object occlusion vary considerably [18].

Face detection and tracking was performed using the method described by Zisserman et al [19], [18]. A combination of frontal, left-profile and right-profile cascade classifiers were computed from a custom training data set extracted from the dialysis videos.

The centroid of the detected face is tracked as the patient moves within the video frame and the Euclidean distance between centroids is quantified for every successive frame. When the movement of the face centroid crosses a given threshold, the frame is considered as active and assigned

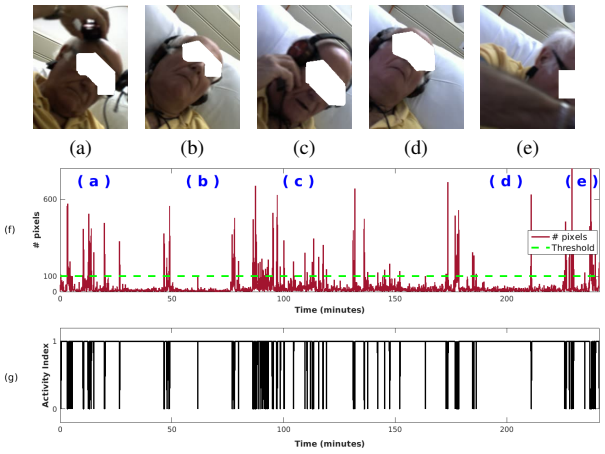

Fig. 3. Activity Index for a typical 4-hour dialysis session: (a) minute 10: wearing headphones to start listening to music, (b) minute 60: sleeping (c) minute 106: awake and engaging in phone conversation, (d) minute 200: sleeping, (e) minute 240: movement and interaction with the clinical staff, (f) Pixels in motion from the activity analysis: areas are labelled (in blue) to correspond to the figures on top. (g) Activity index; a value of 0 is considered a period of patient motion and a value of 1 is considered a stable period suitable for physiological estimation.

an activity index of 0, otherwise it is assigned a value of 1. Figure 3f shows the output of the activity analysis for a typical 4-hour dialysis session.

As most of the video recordings from the dialysis clinical data set were recorded in similar conditions, a fixed threshold is applied to compute the activity index. As shown in figure 3g, the activity index is a binary value, where 0 is considered a period of patient motion and a value of 1 is considered a stable period suitable for physiological estimation.

C. Heart rate estimation

Heart rate estimation is an extension of previous work presented in [20]. The process starts by identifying a reference Region Of Interest, called ROI_R, from areas outside the subject's face such as the background wall. Using the location of the face detected by the algorithms described in the previous section, a grid of multiple regions of interest, labelled $ROI_{S,i}$, is laid out evenly across the total area of the patient's face (in a similar manner to figure 5b).

A multi-channel photoplethysmographic imaging (PPGi) signal is extracted by spatially averaging each $ROI_{S,i}$ (typically 100×100 pixels) for every colour channel. The average colour intensity is also computed from ROI_R.

A beat-by-beat quality assessment is carried out on every PPGi signal to identify data windows suitable for heart rate estimation. Even during periods for which the patient is quiet (sleeping or reading a magazine), video cameras still automatically modify the gain for each colour channel to compensate for changes in the scene, such as sudden changes in the overall lighting (fluorescent lights turned on or off) or shadows, as shown in figure 4a, corresponding to the time when a fluorescent light was turned on. Therefore, the PPGi quality assessment starts by applying a Bayesian change point detection algorithm to find these step changes and discard heart rate estimates during these periods.

Given a data sequence x of N samples from two piece-wise constant inputs μ_1 and μ_2 with Gaussian noise added [21], [22], the probability of a single step change m given the data window provided is defined by:

$$P(m|x) \propto \frac{1}{\sqrt{m(N-m)}} \left[\sum_{i=1}^{N} x_i^2 - \frac{1}{m}(\sum_{i=1}^{m} x_i)^2 - \frac{1}{N-m}(\sum_{i=m+1}^{N} x_i)^2\right]^{-\frac{N-2}{2}} \quad (1)$$

The PPGi waveform is later 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 Hz and 4 Hz (corresponding to 42 and 240 beats/min). These cut-off limits represent the range of expected human heart rates.

Similar to Li et al [23], once the beat onsets are located, the input signal is then divided into 30-second running windows with an overlap of 5 seconds. For each window, a template is constructed from all the valid beats. If a template cannot be computed or is invalid, the template from the previous window is used. The template is used to analyse the morphology of each individual beat within the window. Firstly, beats corresponding to periods of motion (as identified by the activity index algorithm described in previous section or occurring during a step change are flagged as invalid and are assigned a quality value of 0. Secondly, beats that are clipped or are outside a valid physiological or amplitude range are also flagged as invalid.

Finally, a multi-scale Dynamic Time Warping algorithm is used to compute the minimum distance from the beat to the window template. Beats for which this distance lies within a given threshold are ruled to be of good quality, otherwise they are flagged as invalid.

Heart rate estimates are computed from each of the $ROI_{S,i}$ (simply labelled ROI_S), following the strategy below:

1) Each ROI_S and ROI_R is divided into 15-second windows. A window length of 15 seconds corresponds

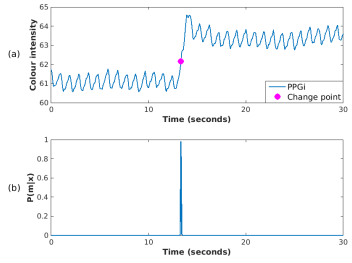

Fig. 4. Bayesian change point detection algorithm applied to a 30-second window during which a fluorescent light was turned on producing a step change : (a) Input PPGi signal with the change point marked around second 13, (b) The probability of the change point.

to approximately 20 cardiac cycles, a sufficient number of cycles for accurate estimation, without introducing too long a processing delay.

2) We fit an auto-regressive (AR) model to the time-series derived from ROI_R. At 12 frames/sec and for a window of 15 seconds, 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 [24].
3) We then fit a separate AR model to the time-series derived from ROI_S in the same way as for ROI_R.
4) 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 k degrees of each other (k = 1 or 2, typically). Pole cancellation in ROI_S gives the new AR model (heart rate information only). More detail on pole cancellation is given in [20].
5) The highest-magnitude pole between 0 Hz and the half-sampling frequency in the AR model for ROI_S is the "heart rate pole". Its angle corresponds to the heart rate in beats/min, the latter being obtained by multiplying θ by $60 f_s/2\pi$, where θ 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 [25].
6) 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.
7) We slide the 15-second window by one second and repeat steps 1 to 8 for the new window. The use of a one-second offset between consecutive windows allows us to derive heart rate estimates (based on the previous 15 seconds of data) every second.

Once the heart rate estimates are computed for each colour channel from each $ROI_{S,i}$, the overall heart rate estimate is computed using a data fusion technique. Each colour channel from every $ROI_{S,i}$ is tracked with an individual Kalman filter, producing one estimate per ROI and per colour channel. The overall heart rate estimate for the current window HR_w is computed by combining the output from each Kalman filter [26], [27] as:

$$HR_w = \sum_{l=1}^{n} \left(\frac{\prod_{k=1,j\neq l}^{n} \sigma_i^2}{\sum_{i=1}^{n} (\prod_{j=1,j\neq i}^{n} \sigma_j^2)} * HR_{ROI_{S,l}} \right) \quad (2)$$

where $l = 1, 2, ..., n$ is the list or ROI_S. The SQI-weighted residual σ is given by:

$$\sigma_i^2 = \left(\frac{r_i}{SQI_i} \right)^2 \quad (3)$$

where r_i is the Kalman filter residual for the $ROI_{S,i}$ and SQI_i is the signal quality index for the data window for the ROI.

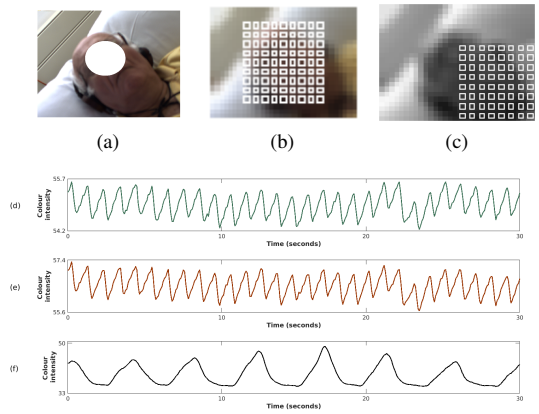

Fig. 5. Regions of interest selection for respiratory rate estimation. (a) The original image frame. (b) the 9x9 grid of all the ROI on the patient's face, used both for computing heart rate and respiratory rate. (c) the 9x9 grid of all the ROI on the patient's chest. 30-second sample waveforms extracted using (d) the green channel from a region of interest on the patient's face, (e) the red channel from a region of interest on the patient's face, (f) the gray-scale image from a region of interest on the patient's chest.

D. Respiratory rate estimation

The previous section showed that heart rate can be estimated using regions of interest on the patient's face. This can be expanded to analyse the image for frequency content in the respiratory rate physiological range, typically between 6 to 42 breaths per minute. As opposed to heart rate, respiratory rate can be estimated not only from the subject's face, but also from the chest area, mostly due to tge patient's thoracic motion due to breathing [20].

For a robust estimation of respiratory rate, regions of interest corresponding to areas on the patient's skin and upper torso have to be considered. The nature of these signals differ in principle, as shown in figure 5. The signals extracted from the patient's skin are pulsatile time series, mainly caused by colour changes due to blood flow (figure 5d and 5e). The signals extracted from the upper torso (such as the thorax) are mainly affected by subject motion due to breathing, as shown in figure 5f.

The breathing-related amplitude variations are extracted from each region of interest in the patient's face 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/minute). 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.

As the patient's upper torso is covered by clothing, the signal extracted does not depend on colour changes due to blood flow, but rather depend on the subject's motion patterns. Each colour image frame from the video data is therefore converted to a gray-scale version and the mean over each region of interest is computed, as shown in figure 5f.

As the number of respiration signals extracted from both the patient's face and torso is large, Principal Component Analysis (PCA) is applied and the first three components are selected. Similarly to heart rate, a data fusion algorithm based on Kalman filters is used to combine the respiratory rate estimates form all the selected PCA components.

E. Identifying changes in SpO2

The methods presented in the previous sections mainly find a frequency component to estimate heart rate or respiratory rate. Identifying changes in SpO_2 depends on the colour reproduction of the video camera sensor from skin regions.

Unlike the photodiode used in a pulse oximeter, colour cameras cannot directly measure the spectra of colour signals because spectral accuracy is sacrificed for spatial resolution [28]. Since the spectral data for a single point in the scene is described with three values (called tristimulus values), the (R,G,B) values are only an approximation of the true incoming colour signal spectra. There exists another major problem caused by this spectral data compression: colour samples with different reflectances can become metameric, so that different combinations of light across the wavelength range can produce an equivalent sensor response. This is more evident when the same colour object is recorded as two different colours when affected by different illumination sources, or when two different colours cannot be discriminated under the effects of other illuminants [29]. Colour changes can be due to several factors such as varying light levels, the temporary presence of shadows during the intervention of the clinical staff, or the varying light colour from changes in the spectral power distribution of light sources (daylight, fluorescent or incandescent).

It is important, therefore, to properly balance the colour reproduction of the video recordings and to choose the right colour channels to be used for identifying changes in SPO_2. Once stable periods are identified from the video recordings, the images are colour-balanced following the generalised diagonal transformation white-balancing model

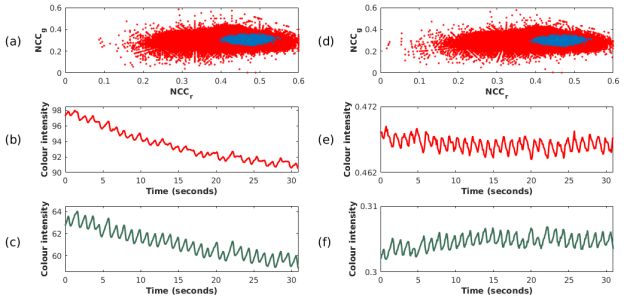

Fig. 6. Comparison of the extraction of a 30-second pulsatile signal from a single region of interest (on the patient's forehead) between the original video (left column) and after colour processing (right column). (a) Chromaticity diagram: the red colour represents the chromaticity for the whole image, whereas the blue is computed from the skin areas only; the pulsatile time series for the original red and green channels are shown in (b) and (c) respectively. (d) Chromaticity diagram after colour balancing and colour space conversion; the pulsatile time series for the NCC_r and NCC_g channels are shown in (e) and (f) respectively.

[30] expressed in the following equation:

$$RGB_{out} = F_{out} \times D \times F_{in} \times RGB_{in} \tag{4}$$

where RGB_{in} is a set of 3×1 camera input RGB values for a given pixel, F_{in} is 3×3 matrix transforming RGB to an intermediate colour space, D is 3×3 diagonal matrix whose values vary with illuminant and effectively performs the colour balancing, F_{out} is 3×3 matrix transforming the intermediate colour space value back into RGB colour space, and RGB_{out} is a set of 3×1 colour balanced RGB values. The matrix D is computed from colour values of the original non-colour balanced image, in the intermediate colour space, for a point manually selected as white.

Even after colour-balancing, the colours of objects perceived by a colour camera, specifically skin tones, are also dependent on the changing lightness (how dark or light the scene is) and conditions that have a strong effect on the intensities of the recorded colours. The choice of a colour space representation for modelling the skin colour changes under different illumination conditions is critical when developing robust techniques against illumination changes. Therefore, the colour-balanced RGB images from the video recordings are converted to the Normal Colour Coordinates (NCC) space as it has been shown to be among the most usable colour spaces for skin chromaticity modelling [31], [17].

From the NCC colour space, the "r,g" chromaticity values (noted as r, g lowercase) are computed using equation 5. Figure 6 shows the results of colour procesing.

$$\begin{aligned} r &= \frac{R}{R+G+B} \\ g &= \frac{G}{R+G+B} \end{aligned} \tag{5}$$

Following the colour processing stage, the video is the subdivided in 15-second windows. For each window, all

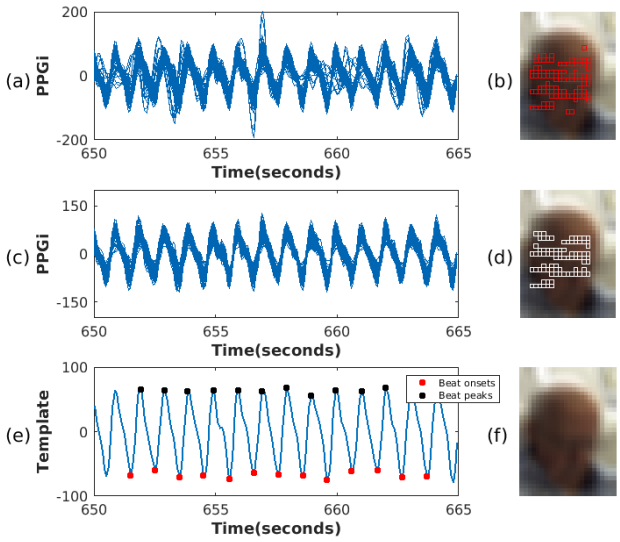

Fig. 7. Building a template for a 15-second window from the red channel. (a) 99 PGGi signals for which the beat-by-beat SQI method used in heart rate analysis labelled as valid, (b) location of the good quality ROI in the image frames. (c) PPGi signals for which the correlation coefficient between the ROI and the window template is greater than 0.8, only 84 ROI are valid; (d) location of the new selected ROI on the image frame; (e) Representative template for the window from which changes in SpO_2 can be identified. (f) The input image shown as a reference.

the regions of interest on the patient's face (as detected by the face-tracking algorithms) are processed using the beat-by-beat signal quality methods used for heart rate analysis. A template is computed from all regions of interest of good quality. Subsequently, a representative PPGi signal is extracted from the ROI for which the correlation coefficient between the ROI and the template is greater than 0.8. This process is shown in figure 7a. The median of the AC and DC values are computed from the representative PPGi signal (figure 7e) from which SpO_2 can be estimated using the ratio of ratios:

$$SpO_2 = A - B\frac{(I_{AC}/I_{DC})_{\lambda_1}}{(I_{AC}/I_{DC})_{\lambda_2}} \quad (6)$$

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 reflected light at wavelengths λ_1 and λ_2. The coefficients A and B are determined using the reported values from the two pulse oximeters. The chosen wavelengths are from the "r,g" NCC chromaticity conversion.

IV. Results

A. Heart rate

According to the table I in the previous section, the number of video sessions recorded was 104, comprising a total video length of 369.1 hours. The mean video length per dialysis session was 3.5 hours. For a total of 20 sessions, the recordings were interrupted due to several reasons, including:

TABLE II

HEART RATE ESTIMATION RESULTS

Device	MAE	MAD	Time %
Finger vs Ear pulse oximeter	1.87 bpm	1.73 bpm	100%
Camera vs Reference HR	2.8 bpm	2.6 bpm	65.3%

TABLE III

RESPIRATORY RATE ESTIMATION RESULTS

ROI locations	MAE	MED	Time
Only face	2.2 bpm	2.1 bpm	60.3 %
Only upper torso	1.8 bpm	1.6 bpm	65.1 %
All combined	2.1 bpm	1.8 bpm	69.2 %

patient discomfort, medical intervention, family visits, video equipment malfunction or other external factors.

The heart rate algorithms described in this paper require video sessions during which the patient is stable and the location of the face is known. Following the criteria described in section III-B, a further 23 sessions were rejected. Therefore, the resulting number of valid dialysis sessions is 61, comprising a total video length of 219.8 hours (approximately 60% of the total number of hours recorded), with a mean video length of 3.6 hours per dialysis session.

Table II presents the overall heart rate estimation results. The mean absolute error (MAE), mean absolute deviation (MAD) and the proportion of estimated time are reported for all the valid dialysis sessions, comprising a total recording time of 219.8 hours. The errors are calculated by comparing the heart rate estimates from the camera against the ground truth heart rate, computed from the mean of the two estimates from the pulse oximeters, as described in section III-A.

B. Respiratory rate

The total time for which the reference respiratory rates (as computed from the ECG and PPG) agree with each other within 2 breaths per minute is 108.82 hours. Table III presents the overall respiratory rate estimation results comparing when taking regions of interest on the subject's face only, torso or both.

C. SpO2

From all the dialysis sessions recordings, 10 segments, each with a duration of 5 minutes, were manually selected. For these, the SpO_2 estimates provided by the two pulse oximeters differ by less than 2% and the computed reference SpO_2 decreased by more than 5% in each 5-minute recording. Table IV compares the reference SpO_2 from the pulse oximeters with the camera calibration.

Figure 8 shows a comparison between the camera estimates and the reference signals for a 6-minute video segment.

TABLE IV

SpO_2 CALIBRATION RESULTS FROM CAMERA

Device	MAE	MAD
Finger vs Ear pulse oximeter	0.97 %	0.78 %
Camera vs Reference SpO_2	2.5 %	1.7 %

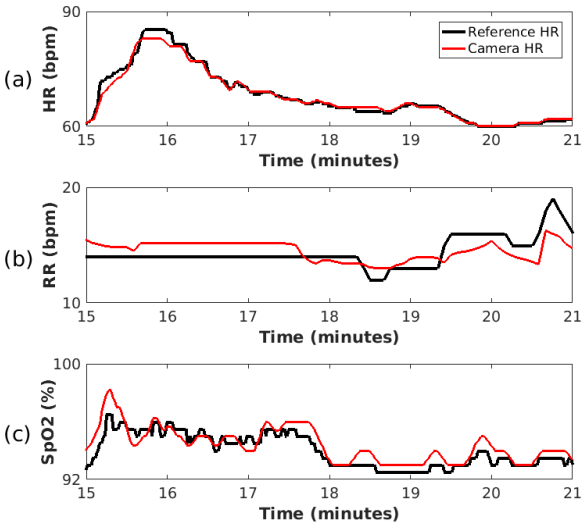

Fig. 8. Summary of vital sign estimation using the proposed methods for a 6-minute video segment. The camera estimates are plotted in red, whereas the reference signals are plotted in black: (a) Heart rate, (b) Respiratory Rate, (c) SpO2

V. Conclusion

The accuracy of the heart rate and respiratory rate estimates from the video camera, in periods during which the subject is stable, is comparable to that of the reference signals computed from devices used in regular clinical care. The camera estimates show a strong positive correlation with the reference signals, with minimal bias for approximately 65% of heart rate estimates and 69% of respiratory rate estimates when the reference signals are valid.

As most of the video recordings from the clinical data set were recorded in similar conditions (patients' upper torso at a similar distance from the camera), some common thresholds and parameters were chosen. The use of algorithms than can learn from data, such as Convolutional Neural Networks (CNN), will be needed to cope with changes in video recording conditions.

References

[1] J. Prior and J. Silberstein, *Physical diagnosis: the history and examination of the patient*. CV Mosby, 1977.

[2] D. Shier, J. Butler, and R. Lewis, *Hole's Human Anatomy*. McGraw-Hill, 2001.

[3] F. Shann, K. Hart, and D. Thomas, "Acute lower respiratory tract infections in children: possible criteria for selection of patients for antibiotic therapy and hospital admission," *Bulletin of the World Health Organization*, vol. 62, no. 5, p. 749, 1984.

[4] W. Lim, M. Van der Eerden, R. Laing, *et al.*, "Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study," *Thorax*, vol. 58, no. 5, pp. 377–382, 2003.

[5] E. Marcantonio, S. McKean, M. Goldfinger, *et al.*, "Factors associated with unplanned hospital readmission among patients 65 years of age and older in a medicare managed care plan," *The American journal of medicine*, vol. 107, no. 1, pp. 13–17, 1999.

[6] M. Pollack, U. Ruttimann, and P. Getson, "Pediatric risk of mortality (prism) score.," *Critical care medicine*, vol. 16, no. 11, pp. 1110–1116, 1988.

[7] A. Lumb, *Nunn's applied respiratory physiology*. Churchill Livingstone, Elsevier, 2010.

[8] D. Meredith, *Continuous Monitoring During Haemodialysis*. PhD thesis, University of Oxford, 2014.

[9] T. R. Association, "The seventeenth annual report." `https://www.renalreg.org/reports/2014`, December 2014.

[10] W. Tharion, M. Buller, C. Clements, *et al.*, "Human factors evaluation of the hidalgo equivital eq-02 physiological status monitoring system," tech. rep., DTIC Document, 2013.

[11] M. Kurella Tamura, K. Covinsky, G. Chertow, *et al.*, "Functional status of elderly adults before and after initiation of dialysis," *New England Journal of Medicine*, vol. 361, no. 16, pp. 1539–1547, 2009.

[12] D. Lamping, N. Constantinovici, P. Roderick, *et al.*, "Clinical outcomes, quality of life, and costs in the north thames dialysis study of elderly people on dialysis: a prospective cohort study," *The Lancet*, vol. 356, no. 9241, pp. 1543–1550, 2000.

[13] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, vol. 28, no. 3, p. R1, 2007.

[14] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *International Journal of Nursing Studies*, vol. 47, no. 8, pp. 931–936, 2010.

[15] I. Smith, J. Mackay, N. Fahrid, and D. Krucheck, "Respiratory rate measurement: a comparison of methods," *British Journal of Healthcare Assistants*, vol. 5, no. 1, p. 18, 2011.

[16] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "Probabilistic estimation of respiratory rate using gaussian processes," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pp. 2902–2905, IEEE, 2013.

[17] S. Z. Li and A. K. Jain, *Handbook of face recognition*. springer, 2011.

[18] J. Sivic, M. Everingham, and A. Zisserman, "Who are you? – learning person specific classifiers from video," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[19] M. Everingham, J. Sivic, and A. Zisserman, ""Hello! My name is... Buffy" – automatic naming of characters in TV video," in *Proceedings of the British Machine Vision Conference*, 2006.

[20] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. Clifton, and C. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *Physiological measurement*, vol. 35, no. 5, p. 807, 2014.

[21] J. Ruanaidh and W. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing (Statistics and Computing)*. Springer, ISBN 978-0-387-94629-0, 1996.

[22] R. P. Adams and D. J. MacKay, "Bayesian online changepoint detection," *arXiv preprint arXiv:0710.3742*, 2007.

[23] Q. Li and G. Clifford, "Dynamic time warping and machine learning for signal quality assessment of pulsatile signals," *Physiological Measurement*, vol. 33, no. 9, p. 1491, 2012.

[24] J. Pardey, S. Roberts, and L. Tarassenko, "A review of parametric modelling techniques for EEG analysis.," *Medical engineering & physics*, vol. 18, pp. 2–11, 1996.

[25] R. Takalo, H. Hytti, and H. Ihalainen, "Tutorial on univariate autoregressive spectral analysis.," *Journal of clinical monitoring and computing*, vol. 19, pp. 401–410, 2005.

[26] L. Tarassenko, L. Mason, and N. Townsend, "Multi-sensor fusion for robust computation of breathing rate," *Electronics Letters*, vol. 38, no. 22, pp. 1314–1316, 2002.

[27] Q. Li, R. G. Mark, and G. D. Clifford, "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a kalman filter," *Physiological measurement*, vol. 29, no. 1, p. 15, 2008.

[28] B. Fortner, "a meyer, t. e.(1997) number by colors: A guide to using color to understand technical data."

[29] G. Wyszecki and W. S. Stiles, *Color science: concepts and methods, quantitative data and formulae*. John Wiley & Sons, 1982.

[30] F. Xiao, J. E. Farrell, J. M. DiCarlo, and B. A. Wandell, "Preferred color spaces for white balancing," in *Electronic Imaging 2003*, pp. 342–350, International Society for Optics and Photonics, 2003.

[31] B. Martinkauppi, *Face colour under varying illumination-analysis and application*. PhD thesis, University of Oulu, 2002.