Non-contact monitoring of respiration in the neonatal intensive care unit

João Jorge¹, Mauricio Villarroel¹, Sitithichok Chaichulee¹, Alessandro Guazzi¹, Sara Davis², Gabrielle Green², Kenny McCormick² and Lionel Tarassenko²

¹ Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK
² Neonatal Unit, John Radcliffe Hospital, Oxford University Hospitals Trust, UK

Abstract—An abnormal respiratory rhythm is an early indicator of physiological deterioration. It is of critical importance in the clinical management of critically-ill or premature infants, for whom apnoea of prematurity is a major concern. Nevertheless, respiratory signals are still largely disregarded in neonatal intensive care units due to the high prevalence of noise and high false alarm rates in conventional monitoring. To address this, we present a novel method for the extraction of respiration from camera-based measurements taken from the top-view of an incubator. A total of 107 events from 30 neonatal admissions were annotated by three clinical reviewers as either true cessations of breathing (physiologically relevant) or false (artifact-related). The events were divided into two independent groups for training and validation and our algorithm was trained to classify true cessations. We achieved a good classification performance with 9 out of 10 cessations and 7 out of 10 artefactual events correctly identified in the out-of-sample test set. A reduction in false alarm rate of 77.3% was achieved.

I. INTRODUCTION

A. Remote physiological measurements

Respiratory rate is an important indicator of physiological distress. Conditions such as cardiopulmonary arrest [1], sudden infant death syndrome [2] and other conditions which lead to changes in the arterial partial pressure of blood gases [3] are often preceded by changes in this vital sign. In clinical practice, the respiratory function of spontaneously breathing patients is usually monitored either intermittently, by manual counting, or continuously, using invasive monitoring techniques available in intensive care units. Flow sensors applied to the nose and/or mouth or devices that detect chest wall excursions applied to the chest or abdominal area have also been used [4] although not in clinical practice.

In an effort to reduce patient discomfort and make respiratory measurements more ubiquitous, several methods have been proposed which do not require contact sensors to be attached to the patient, such as thermal imaging [5], microwave-based [6] and radar-based [7] methods. Although technically these are non-contact approaches, they require specialist instrumentation and complex set up procedures.

Under the paradigm of remote sensing, the latest decade has seen the emergence of numerous approaches to the measurement of vital signs using visible-light video cameras. With considerable overlap, most approaches to extract respiration from visible light imaging can be classified into those based on either a photoplethysmographic [8]–[15] or computer vision [16]–[23] approach to signal extraction and analysis.

B. Remote monitoring of neonates

The immaturity of the brain mechanisms governing the respiratory rhythm in neonates can result in irregular breathing patterns with prolonged pauses (apnoeas). These periods pose a clinical problem particularly in infants of gestational age below 36 weeks, for whom apnoea of prematurity is a cause of recurrent episodes of severe oxygen desaturation often associated with adverse clinical outcomes [24], [25].

The automatic monitoring of respiration during neonatal sleep is, therefore, of crucial importance both in (a) intensive care units with high patient-to-staff ratios and in (b) home environments.

Conventional contact devices present clear shortcomings in these scenarios. Not only do they often require cumbersome equipment but they can also cause infant stress and even pain due to damage to their fragile skin. Respiration monitoring in neonatal units is generally done by impedance pneumography (IP). IP is a convenient method if patients are already monitored by electrocardiography (ECG), but is prone to inaccurate readings due to a number of factors including poor ECG electrode placement, motion artefact, and physiologic events which cause thoracic movements unrelated to breathing (such as coughing, or crying) [26], [27].

Although several studies in visible-light monitoring of neonates have so far been advanced, most have been conducted for the extraction of heart rate using camera monitors [14], [15], [28]–[31], with few approaches aimed at measuring respiratory parameters [31]–[33] in this population.

C. Contributions

In this paper, we describe a novel non-contact means of respiratory monitoring and assess its performance for the purpose of detecting cessation of breathing events (COBEs) in a dataset of spontaneously breathing premature infants.
Our method includes the following innovations: firstly, skin segmentation is used in order to localise the subject and select a region-of-interest (ROI). Secondly, cumulative frame differences are used to boost the signal-to-noise ratio (SNR) of subtle breathing motions over the ROI. Lastly, we perform activity analysis for each time window to detect the presence of high-frequency motion, and so exclude these windows from further analysis.

The rest of this paper is organised as follows: Section II-A introduces the dataset and II-B the proposed methodology. The experimental results are reported in Section III and discussed in Section IV. The main conclusions and directions for future work are presented in Section V.

II. MATERIALS AND METHODS

A. Data collection

Data for this study was acquired during an observational study in the Neonatal Intensive Care Unit (NICU) at the John Radcliffe Hospital (MONITOR Study - REC: 13/SC/0597). This study involved the monitoring of 30 preterm infants of less than 37 weeks of corrected postmenstrual age. Each preterm infant was recorded under ambient light for up to four consecutive days.

We obtained a dataset of (a) RGB video footage acquired using a digital video camera positioned over the incubator inside which study infants were nursed (Giraffe Omnibed® incubator, General Electric, Fairfield, USA), and (b) conventional vital sign data collected concurrently by the patient monitor as part of routine care. This dataset is supplemented by (c) clinical annotations as described in Section II-C. The monitoring equipment is shown in Figure 1.

Video acquisition

We used a 3 CCD (Sony ICX274AL®, Sony, Tokyo, Japan) digital camera (JAI AT-200C®, JAI, Glostrup, Denmark) interfaced via Camera Link to a framegrabber card (Microenable IV AD4-CL®, Silicon Software, Mannheim, Germany). This device was set to acquire 24-bit true colour images (8-bit per colour) at a pre-set rate of 20.3 frames per second and at a resolution of 1628 \times 1236 pixels (with an individual pixel size of 4.4 microns square). The frame buffers were processed using a Spartan® FPGA (Field Programmable Gate Array) board (Xilinx, San Jose, USA) installed in a 1.90 GHz Intel® processor workstation with 4 \times 8 GB RDIMM (registered memory module) and 8 \times 4 TB SATA disks running in-house software under a Fedora 20 Linux operating system.

Patient monitor and sensors

All conventionally-monitored signals were saved on an IntelliVue® MX800 patient monitor (Philips, Amsterdam, Netherlands) and relayed to a separate work station, a standard PC with a 4 TB disk and 8 GB RAM under Windows 7 (iXELLENCE GMBH, iXTrend®) via R-232 serial transmission.

Using proprietary software, the infant’s vital signs were derived from the collected signals and reported at a rate of 1 estimate per second; heart rate (HR) was derived from 3-lead ECG, RR from bipolar impedance pneumogram (IP) signals (both collected using the set of neonatal chest electrodes provided with the monitor at 250 Hz and 62.5 Hz, respectively) or, in the case of spO2 from the photoplethysmographic (PPG) signal acquired by a SET LNCS Neo® pulse oximeter probe (Masimo Corporation, Irvine, USA) sampled at 125 Hz.

B. Methodology

We propose a new method for the measurement of respiration from neonatal video data, which we incorporate into a COBE detection system.

First, skin pixels are identified in each frame using a skin classifier and the region-of-interest (ROI) for motion analysis is selected. The purpose of motion analysis is two-fold; we seek to (1) detect breathing movements and estimate their frequency and (2) produce a binary signal quality index (SQI) for these estimates based on the level of activity on the video sequence. Finally, the information provided by these two signals is combined into a COBE classifier (as shown in Figure 2).

Selection of ROI

The pixels in each frame were clustered into skin and non-skin classes based on their a* chroma value using a 2-class classifier based on Gaussian Mixture models applied to a downsampled a* image. The smallest bounding box which contains the largest contiguous skin region in each frame is selected as the ROI. Figure 3 provides some examples of the ROIs obtained using this method.

Motion analysis

1) for activity detection: A straightforward way to detect motion noise inside the incubator, such as that caused by spurious patient movement, nursing interventions or maternal
touch, is to monitor the high frequency content in raw pixel intensities.

In this analysis, we averaged the value of the blue colour channel over the ROI area and processed this signal with a high-pass filter (3rd order Butterworth; frequency cut-off at 1.6 Hz) designed to attenuate breathing motion as its cut-off frequency approximates the maximum of the neonatal range of respiratory frequencies (100 breaths/min).

The blue channel was selected due to the low penetration depth of light in skin over this wavelength range [34]. As blue light is more affected by specular reflection than other colour channels, the intensity signal collected at this wavelength is much less modulated by the blood-volume pulse, thus rendering it a good proxy for subject motion.

A simple frame-wise measure of activity was obtained by adding the filtered pixel intensities for each frame \( k \). Lastly, this series was filtered using a moving 5-point average filter to achieve a smoother description of activity at the time of each frame, \( a(t_k) \).

2) for breathing detection: During normal sleep, breathing causes subtle movements of the thorax and abdomen (and, to a lesser extent, the entire body). These are observable as a periodical variations in the temporal differences between consecutive frames in continuous image data:

\[
\bar{\Delta}I_k(i,j) = |I_k(i,j) - I_{k-1}(i,j)| \quad (1)
\]

where \( k \) is the frame index, \( I_k \) is the array of intensity values over that frame (or, in our case, the blue colour component) and \( i \) and \( j \) are spatial coordinates.

This approach has been applied successfully in [17], [18] to extract respiratory frequencies of subjects under controlled environments for illumination. By thresholding \( \bar{\Delta}I_k \), one can obtain a difference mask. This mask might include some non-subject pixels (e.g. in image regions affected by shadows or other changes in illumination) as well as exclude moving regions with low contrast. To address this, we used the temporal information over a time interval longer than the sampling period by accumulating the frame differences over the past \( N \) frames, as given in Eq. 2.

\[
A_N^k(i,j) = \sum_{l=k}^{k-N+1} \bar{\Delta}I_l(i,j) \quad (2)
\]

A reasonable value of \( N \) is that which corresponds to the average duration of the upslope of the respiratory waveform, a time interval typically associated with inspiration. To exploit the intrinsic parameters of the system, we relate this quantity to the respiratory time constant \( \tau_e \), a property of the respiratory system analogous to the time constant of RC electrical circuits [35]; one time constant is defined as the time it takes the lung alveoli (capacitor) to discharge 63% of tidal volume (electrical charge) through the respiratory airways (resistance).

For a premature infant with a lung compliance of 0.005 L/cm H\(_2\)O and an airway resistance of 30 cm H\(_2\)O/Ls [36], \( \tau_e \) is then 0.15 s [36]. Hence, it takes about 0.45 s (3\( \tau_e \)) for 95% of tidal volume to be exhaled.

In spontaneously breathing infants, there is a ratio of approximately 1:2 between the inspiration and expiratory constants [37]:
Thus, at an effective frame rate of 20.3 Hz, the value of $N$ corresponding to 3 inspiratory constants is roughly $N_{3\tau_{i}} = 5$ frames. A camera-based respiratory signal $\phi_k$ was then computed as

$$\phi_k = \sum_{i,j \in \text{ROI}} A^{N_{3\tau_{i}}} (i, j)$$

Finally, respiratory rate $RR_{\phi}$ was estimated from 10 s windows of $\phi$ (with 1 s overlap) using a 5th order AR model in the manner described in [111] adjusted to the neonatal respiratory range.

C. Construction of a data set for validation

For the creation of training and validation subsets, we have gathered a set of 107 time epochs (of 300 s) comprising episodes of cessations of breathing (positive events) and instances of normal breathing (negative events). These were extracted from a dataset of 30 NICU stays (comprising a total of 455 monitoring hours) as described below.

In severe cases, episodes of neonatal apnoea can last for 20 seconds or longer [38]. The following was then considered to be a necessary condition for detecting COBEs:

$$RR_{IP} < 20 \text{ breaths/min} \text{ for a period of at least 20 seconds,}$$

where $RR_{IP}$ is the respiratory rate estimated (at 1 Hz) by the patient monitor using bipolar IP.

The set $\Omega_\phi$ of events detected by applying the criterion above to the dataset was reviewed by a team of three reviewers (including a consultant neonatologist, a clinical research fellow and a neonatal nurse) to ascertain whether each occurrence was artefactual or an actual COBE, and (if so) apnoea-related. For each event, reviewers decided among three options: (a) the event was caused by noise or artefacts in the raw signals (b) the event was caused by a cessation of breathing (positive event), or (c) the event was not caused by cessation of breathing nor by artefacts in the raw signals (i.e. $RR_{IP}$ accurately reported a low respiratory rate value).

The (a) set was subdivided into $(a_2)$ motion artefacts and $(a_1)$ IP artefacts (i.e. related to suboptimal probe placement or other acquisition artefacts which cause the IP signal to deviate from true respiratory effort). The (b) set was subdivided into cessations of breathing with $(b_1)$ and without $(b_2)$ an associated apnoea (i.e. with desaturation and bradycardia). Examples of signal segments for scenarios $a_2$ and $b_1$ are given in Figures 4 and 5, respectively.

D. Classification of COBEs

To detect COBEs, $RR_{\phi}$ values for times with $a(t) > 0$ were discarded due to the presence of non-breathing motion in the video stream. $RR_{\phi}$ was obtained by linearly interpolating $RR_{\phi}$ over the time intervals between the remaining
segments where these were less than $\Delta t^*$ apart. $\Delta t^*$ is the detection threshold, a parameter of the classifier. For each event, a COBE was deemed to have occurred if

$RR_\phi < 20$ breaths/min for a period of at least 20 seconds,

We used ROC analysis to determine $\Delta t^*$ on the training set.

III. RESULTS

The clinical reviewers annotated a total of $|\Omega_c| = 107$ potential COBEs. Figure 6 summarises the findings of this annotation process. The analysis of potential COBEs by the clinical reviewers revealed that 82 out of 107 (76.6%) such occurrences were due to the presence of artefacts in the IP signal (11 due to motion and 71 due to poor signal acquisition). Of the remaining 25 artefact-free events, only 20 (28.6% of potential COBEs) corresponded to true cessations of breathing.

The manually classified events were then distributed at random between test and training sets in a manner which preserved the balance between the number of elements in each subclass (Table I). In practice, only a subset of 10 negative events was used in each experiment on the training set to ensure an equal number of positive and negative events in this set.

By following the methodology described, it was possible to classify the true cessations of breathing. The performance

\[ \begin{array}{c|c|c|c|}
\text{Events, classes} & \text{Train set} & \text{Test set} & \text{Total} \\
\hline
\text{IP artifacts, } a_1 & 36 & 35 & 71 \\
\text{Motion artifacts, } a_2 & 5 & 6 & 11 \\
\text{Apneic COBE, } b_1 & 8 & 9 & 17 \\
\text{Non-apneic COBE, } b_2 & 2 & 1 & 3 \\
\text{Low } RR, c & 2 & 3 & 5 \\
\text{Total, } \Omega_c & 53 & 54 & 107 \\
\hline
\text{Positives, } P = a_1 \cup a_2 \cup c & 10 & 10 & 20 \\
\text{Negatives, } N = b_1 \cup b_2 & 43 & 44 & 87 \\
\text{Total, } \Omega_c & 53 & 54 & 107 \\
\end{array} \]

\[ \begin{array}{c|c|c|}
\text{Train set} & \text{Test set} & \\
\hline
P & (10) & (44) \\
N & (10) & (10) \\
\hline
9 & 7 & 34 \\
\hline
b_1 & (9) & (35) \\
b_2 & (1) & (6) \\
a_1 & (3) & \\
a_2 & (6) & 2 \\
c & (3) & 6 \\
\end{array} \]
of the proposed method for COBE detection on the training and test sets is reported in Table II.

The majority of IP artefacts ($a_1$) in the test set were correctly identified as negative events (26 out of 35 events) and all the movement-related artefacts ($a_2$) in the same set were correctly identified (6 events). Regarding positive events ($b_1$ and $b_2$), 7 out of 10 COBEs in the test set were detected. In conjunction, these results yield a reduction in the false alarm rate of 77.3% on the test set.

Figure 7 shows activity $a(t)$ and respiratory signals during a clinically validated episode of neonatal apnoea. We observe that motion events manifest as perturbations in the activity signal $a(t)$. After motion events, marked by $a(t) > 0$ (shown as grey-shaded areas in the top panel) are excluded, the lower intensity changes in the camera-derived signal due to the respiratory cycle become apparent (as illustrated by the $\tilde{\phi}$ plot).

During the apnoeic episode, $RR\tilde{\phi}$ (shown in the third panel of the same Figure) drops from a resting rate of 50 breaths/min to assume values below the threshold set in Section II-D for the detection of true cessations (20 breaths/min) for a period of approximately 40 seconds, therefore generating a positive detection.

IV. DISCUSSION

Methods for measuring motion based on temporal differencing can adapt rapidly to changes in background, environment lighting or spurious motion. However, the small amplitude of abdominal and chest movements associated with neonatal breathing (typically in the order of a few millimetres) means that these can be imperceptible between consecutive frames. We successfully address this issue through the use of accumulated temporal differences. The relatively low computational cost of computing temporal differences when compared with other techniques for estimating motion based on optical flow makes our approach ideally suited for real-time implementation.

A resting respiratory rate below 20 breaths/min is abnormal in premature infants (Figure 8). Such low rates could indicate a progression into an apnoeic event, and therefore, they should prompt immediate clinical attention. Our study in the Oxford NICU revealed that a high number of these oc-
currences (76.6%) are due to the presence of artefacts in the IP signal, the standard modality for monitoring respiration in neonatal units. This poses a clear problem as high false alarm rates may desensitize NICU staff and reduce reaction speed by alarm fatigue [39].

Through the analysis of the video signals acquired from the inside of the cot, we classified the potential cessation of breathing events in the dataset. By identifying where high-frequency movement occurs, it was possible to distinguish between subtle body movements associated with the respiratory effort from spurious motion inside the cot. Further analysis of the segments of low activity allowed us to classify breathing versus non-breathing behaviour.

A meaningful reduction in the number artefact-related events of was achieved. All the movement-related artefacts ($a_2$) and the majority of IP artefacts ($a_1$) in an out-of-sample test set were correctly identified as artefactual detections. Given that the detection of true cessations is paramount in patient care, it is of critical importance that any increase specificity is not achieved at the expense of the capacity to detect positive events ($b_1$ and $b_2$). In the work presented here, 7 out of 10 COBEs in the test set were detected. On closer inspection, 2 of these cases were found to correspond to periods when the monitored infant was undergoing phototherapy. This procedure involves prolonged exposure to LED light predominantly in the blue region of the spectrum (460 - 490 nm) to an irradiation level of 30 to 40 $\mu$W/cm$^2$/nm. This causes bright regions in the acquired images to become saturated, thus decreasing SNR over these regions and depressing the $a(t)$ signal.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a novel technique to obtain activity and respiratory signals from neonatal infants nursed inside an incubator through the temporal processing of the acquired video. Our approach combines these two signals to obtain a more reliable respiratory signal and identify periods of its cessation.

Our results show that the method presented can successfully distinguish between artefactual and true decreases in RR in a range of critically low values for this variable (RR < 20 breaths/min). Although sample size was small, the feasibility of our approach based on motion analysis was clearly demonstrated.

Conventional neonatal monitors could, therefore, benefit greatly from the collection of video signals. Specifically, the implementation of video-based COBE detection systems could potentially help reduce the high rate of false apnoea alarms in these units. Hence, we believe our findings should motivate extensive validation of such techniques.

In this work, we have used a simple colour-based skin detector to segment the ROI. To build in the capability for continuous monitoring over longer periods of time, there is a need for more robust subject segmentation. Methods which include additional features such as texture, and optical flow with high level features such as pose and knowledge of anatomical structures could be explored. Recent developments in deep learning research have provided a framework for embedding visual features within convolutional neural networks to yield highly accurate classifiers [40].

Finally, the possibility to extract additional vital signs using the same camera-based device would also offer significant advantages. In particular, the ability to obtain camera-derived estimates of HR and spO$_2$, in addition to RR, during episodes of bradycardia and desaturation, would leverage the system presented here from a COBE detector to an apnoea detector, thereby paving the way for truly non-contact surveillance of apnoea-related cessations of breathing.

VI. ACKNOWLEDGMENTS

The authors gratefully acknowledge Sheula Barlow and Sharon Baron for their contribution to the clinical annotation of COBEs. We would like thank the FG’17 reviewers for their comments.

REFERENCES


