www.elsevier.com/locate/cbm

Monitoring Infants by Automatic Video Processing: A Unified Approach to Motion Analysis

Luca Cattani, Davide Alinovi, Gianluigi Ferrari, Riccardo Raheli, Elena Pavlidis, Carlotta Spagnoli, Francesco Pisani

PII: S0010-4825(16)30303-1
DOI: http://dx.doi.org/10.1016/j.compbiomed.2016.11.010
Reference: CBM2545

To appear in: *Computers in Biology and Medicine*

Received date: 6 August 2016
Revised date: 20 November 2016
Accepted date: 23 November 2016

Cite this article as: Luca Cattani, Davide Alinovi, Gianluigi Ferrari, Riccardo Raheli, Elena Pavlidis, Carlotta Spagnoli and Francesco Pisani, Monitoring Infants by Automatic Video Processing: A Unified Approach to Motion Analysis, *Computers in Biology and Medicine* http://dx.doi.org/10.1016/j.compbiomed.2016.11.010

Monitoring Infants by Automatic Video Processing: A Unified Approach to Motion Analysis☆

Luca Cattani[a], Davide Alinovi[a,*], Gianluigi Ferrari[a], Riccardo Raheli[a], Elena Pavlidis[b], Carlotta Spagnoli[b], Francesco Pisani[b]

[a] *Department of Information Engineering, University of Parma, Parco Area delle Scienze 181/A, IT-43124, Parma, Italy*
[b] *Department of Neuroscience, University of Parma, Via Volturno 39, IT-43126, Parma, Italy*

Abstract

A unified approach to contact-less and low-cost video processing for automatic detection of neonatal diseases characterized by specific movement patterns is presented. This disease category includes neonatal clonic seizures and apneas. Both disorders are characterized by the presence or absence, respectively, of periodic movements of parts of the body—e.g., the limbs in case of clonic seizures and the chest/abdomen in case of apneas. Therefore, one can analyze the data obtained from multiple video sensors placed around a patient, extracting relevant motion signals and estimating, using the Maximum Likelihood (ML) criterion, their possible periodicity. This approach is very versatile and allows to investigate various scenarios, including: a single Red, Green and Blue (RGB) camera, an RGB-depth sensor or a network of a few RGB cameras. Data fusion principles are considered to aggregate the signals from multiple sensors. In the case of apneas, since breathing movements are subtle, the video can be preprocessed by a recently proposed algorithm able to emphasize small movements. The performance of the proposed contact-less detection algorithms is assessed, considering real video recordings of newborns, in terms of sensitivity, specificity, and Receiver Operating Characteristic (ROC) curves, with respect to medical gold standard devices. The obtained results show that a video processing-based system can effectively detect the considered specific diseases, with increasing performance for increasing number of sensors.

☆This paper has supporting multimedia material available.

*Corresponding author.

Email addresses: luca.cattani@unipr.it (Luca Cattani), alinovi@tlc.unipr.it (Davide Alinovi), gianluigi.ferrari@unipr.it (Gianluigi Ferrari), riccardo.raheli@unipr.it (Riccardo Raheli), elena.pavlidis@studenti.unipr.it (Elena Pavlidis), carlotta.spagnoli@studenti.unipr.it (Carlotta Spagnoli), francesco.pisani@unipr.it (Francesco Pisani)

Keywords: Neonatal clonic seizure, apnea, breath monitoring, periodicity analysis, maximum-likelihood detection.

1. Introduction

Monitoring neonatal movements may be exploited to detect symptoms of specific disorders. In fact, some severe diseases are characterized by the presence or absence of rhythmic movements of one or multiple body parts. Clonic seizures and apneas can be identified by sudden periodic movements of specific body parts or by the absence of periodic breathing movements, respectively.

Seizures are clinically defined as paroxysmal alterations of the neurological functions (i.e., behavioral, motor or autonomic function) and represent a distinctive symptom of acute brain dysfunction in the newborn [1]. Leading causes of neonatal seizures are intracranial haemorrhage, hypoxic-ischaemic encephalopathy and sepsis. As major symptoms of acute central nervous system impairment, almost 80% of these paroxysmal events occur in the first two days of life [2]. The estimated incidence of seizures is between 1‰ and 3.5‰ in full-term newborns and even higher in preterm infants [3]. Essentially, four main clinical seizure types can be recognized in neonates, namely: subtle, clonic, tonic and myoclonic [1]. Prompt and accurate detection of neonatal seizures is crucial to administer timely treatments that may prevent further seizure-induced brain damage. It is important to stress that the various categories of neonatal seizures are characterized by very different kinds of movements and require different analysis approaches. This article analyzes the monitoring of *clonic* seizures. Analysis systems for other types of neonatal seizures are discussed in the following articles: [4, 5] for subtle seizures, [6] for myoclonic seizures.

Apneas consist of the temporary absence of spontaneous respiration [7], which manifests itself by the lack of breathing movements for a certain time period. Seizures, cerebrovascular events [8] and congenital diseases, such as Congenital Central Hypoventilation Syndrome (CCHS) [9], are among the main causes of these events in the neonatal period. CCHS, also referred to as Ondine's curse, is a life-threatening disorder. Discovered as a disease caused by mutations in the paired-like homeobox 2B (PHOX2B) gene, CCHS manifests itself, in the neonatal period, especially during quiet sleep, with cyanosis, apnea events or even cardiorespiratory arrests [9]. CCHS is a relatively rare disorder: in fact, just approximately 1000 individuals with this condition have been identified. Researchers believe that some cases of Sudden Infant Death Syndrome (SIDS) may be caused by undiagnosed CCHS [10].

Monitoring of vital signs for the diagnosis of such type of diseases is possible almost exclusively in hospital environments. Currently, the standard monitoring systems are based on ElectroEncephaloGram (EEG) polygraphy and polysomnographic devices [11], composed of several wired sensors directly connected to the body of the patient: electrodes positioned on the scalp, chest and other body parts, elastic belt around the chest, nasal flow meter and pulse oximeter. The polysomnographic device allows to monitor the brain electrical

activity, the cardiac and muscular activities, the respiratory movements, the breathing pattern and the blood oxygen saturation. These systems, typically used for short-term monitoring, are expensive, time-consuming, moderately invasive (especially for newborns), and require experienced medical staff, not always available full-time in a Neonatal Intensive Care Unit (NICU). For home monitoring, many systems are available, but they require sensors attached to the body of the newborn (e.g., smart bed [12], wearable sensor system [13]). At the opposite, the objective of our research is to study a contact-less monitoring system. Thus, automatic, real-time and non-invasive equipment able to reliably recognize these diseases would be strategic in hospital environments, helping to monitoring 24/7 all the newborns present in a NICU, or even at home, in order to implement remote monitoring systems, reducing time and cost of hospitalization.

An attractive contact-less monitoring tool to automatically detect movement- or breathing-dependent diseases, such as seizures or breathing disorders, may rely on properly processing video signals, acquired through single or multiple video cameras. The movements of the newborn's body (e.g., limbs and chest), framed by the cameras, can be analyzed to detect specific behaviors which may be symptoms of neurological dysfunction.

In [14], the acquisition, through sophisticated video processing, of the motion strength was proposed as expedient to detect the presence of neonatal seizures. In [15], the authors used an optical flow-based technique to track and characterize the movements of a newborn during prolonged monitoring. Neural networks were then used to obtain a diagnosis based on a previous training phase. Taking into account the long monitoring time and the fact that the implementation of long reliable tracking of jerky movements of the newborn limbs may be very complex, this approach is not suitable for real-time detection and requires expensive hardware for accurate optical flow processing (especially for dense optical flow techniques). In [16], a real-time video processing-based approach to the detection of neonatal clonic seizures by recognizing characteristic periodic movements, was proposed. This approach relies on a periodicity detector, based on hybrid pitch estimation, to analyze a motion signal extracted from the video. This method has also received attention in the medical literature [17].

In this paper, an improved method to estimate the periodicity of pathological movements, based on the use of the Maximum Likelihood (ML) criterion [18], is presented. In particular, motion signals from multiple digital cameras or depth-sensor devices (e.g., Kinect [19]) are extracted and properly processed in order to detect potential abnormal motor patterns. We propose a monitoring system based on the detection of pathological movements, characterized by the presence or absence of a significant periodic component (i.e. rhythmic movements). Thus, the field of application includes any disease presenting this type of symptoms: clonic seizures and CCHS are relevant examples. The novelty lies in combining known techniques (more or less recent) to create an innovative monitoring system that is, unlike existing systems, contact-less and low-cost. In fact, as an example, we were able to develope an Android application, denoted as Smartphone-based Contactless Epilepsy Detector (SmartCED), able to detect

in real time, capturing images through the phone camera, the occurrences of clonic seizures: the computing resources of modern smartphones have proven to be sufficient for our purpose [20]. This paper unifies and expands preliminary contributions [21] and [22]. The approach behind these two studies, namely, detection and estimation of periodic movements, is common, so that a unified vision is here proposed.

The remainder of the paper is organized as follows. In Section 2, the method for the analysis of periodic movements is described. In Section 3, performance results are presented. Finally, in Section 4 conclusions are drawn.

2. Video Processing for Periodic Motion Analysis

2.1. Extraction of Temporal Motion Signal

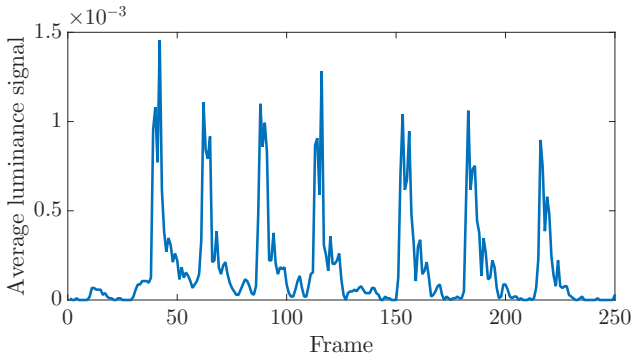
Extraction of a relevant motion signal from every sensor is the first key step of the proposed approach. The application to standard Red, Green and Blue (RGB) [23] cameras is initially presented, and the extension to depth sensors is later discussed.

We start considering a video (i.e., a sequence of frames) with sampling period T. Frames are numbered as $i = 1, 2, ...$, specifying the indexes related to time instants as integer multiples of T. Each frame is described by a matrix of $W \times H$ pixels. The images acquired by every camera are processed, through a sequence of standard image processing operations, in order to highlight the movements of the body parts [23]. Following the approach in [16], every frame is first converted to gray scale; then, a simple Finite Impulse Response (FIR) filtering operation based on the difference between consecutive frames is performed. With the objective of a low computational complexity, the resulting frames can be converted to a binary scale and the *erosion* morphological operation [23] can be applied. In this way it is also possible to reduce some residual noise, as discussed in [16]. The resulting binary-element frame matrix is denoted as $\boldsymbol{I}_s[i]$, $i \in 1, 2, ...$, where the subscript s identifies the s-th sensor. Since a binary image is composed of white pixels (having a luminance value equal to 1) and by black pixels (having a luminance value equal to 0), a spatial average luminance signal $\overline{L}_s[i]$ can be defined as:

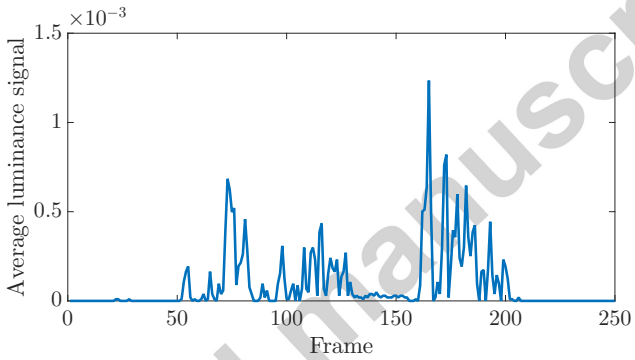
$$\overline{L}_s[i] \triangleq \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} I_s[x, y, i] \tag{1}$$

where $I_s[x, y, i]$ is the $[x, y]$ entry of $\boldsymbol{I}_s[i]$. Thus, the signal $\overline{L}_s[i]$ is the average number of white pixels in the i-th binary frame $\boldsymbol{I}_s[i]$ coming from the s-th sensor. This signal may represent the movement "pattern" of the involved body parts and is also referred to as motion signal.

In Figure 1, illustrative examples of average luminance signals, extracted from a single RGB camera according to the above procedure, are shown, considering: (a) a clonic seizure and (b) random movements. For a comparison of the signal that can be obtained by this technique with the corresponding EEG signal associated with the same clonic seizure, please refer to [16].



(a) Clonic seizure



(b) Random movements

Figure 1: Examples of extracted average luminance signals: (a) clonic seizure and (b) random movements.

2.2. Maximum-Likelihood Approach for Periodicity Detection

Once the motion signals from every sensors have been extracted, a method to decide on the possible presence of a periodic movement is needed. Since the diseases under study have in common symptoms described by the presence or absence of quasi-periodic movements, the goal is to determine whether the extracted signals have a common periodic component and, if so, to estimate it. To this purpose, the motion signal in (1) related to the s-th sensor is assumed to be modeled as:

$$\overline{L}_s[i] = c_s + A_s \cos[2\pi f T i + \phi_s] + n_s[i] \tag{2}$$

where: c_s is a continuous component, $\{n_s[i]\}$ is a sequence of independent identically distributed (i.i.d.) zero-mean Gaussian noise samples and A_s, ϕ_s, f are amplitude, phase, and frequency, respectively, of the periodic component. The sampling period T is assumed identical on all signals. The periodic component

of interest is assumed to have the same frequency on all sensors, whereas continuous components, phases and amplitudes are expected to take on different values for each sensor. The main goal is to detect the presence of a common sinusoidal component and estimate its frequency f. The signals coming from multiple sensors have to be aggregated using data fusion principles, in order to reinforce the estimation of the common parameter f.

The ML estimation of A_s, ϕ_s and f can be achieved minimizing the following likelihood function [18]:

$$J\left(A_s, f, \phi_s\right) \triangleq \sum_{s=1}^{S} \sum_{i=1}^{N} \left(\overline{L}_s[i] - A_s \cos[2\pi f T i + \phi_s]\right)^2 \tag{3}$$

where: N is the number of consecutive video frames (i.e., an N-frame window), S is the number of sensors and $\{\overline{L}_s[i]\}_{i=1}^{N}$ are the observed samples for the s-th sensor.

Under the assumption of i.i.d. zero-mean Gaussian noise, after standard algebraic manipulations, the approximate ML frequency estimator is [18, 24]:

$$\hat{f} = \underset{f}{\operatorname{argmax}} \sum_{s=1}^{S} \left| \sum_{i=1}^{N} \overline{L}_s[i] e^{-j2\pi f T i} \right|^2 . \tag{4}$$

The s-th term of the outer sum in (4) represents the periodogram of $\overline{L}_s[i]$ and, thus, its peak indicates the most significant harmonic component and equals the estimated frequency. Note that it is possible to estimate the frequency from one sensor alone using (4) with $S = 1$.

After frequency estimation, every amplitude components $\hat{A}_s$ can be estimated as [24]:

$$\hat{A}_s = \frac{2}{N} \left| \sum_{i=1}^{N} \overline{L}_s[i] e^{-j2\pi \hat{f} T i} \right| \qquad s = 1, 2, ..., S. \tag{5}$$

A significant periodic component is declared if a certain threshold η is exceeded according to:

$$\frac{N}{S} \sum_{s=1}^{S} \hat{A}_s^2 > \eta. \tag{6}$$

The threshold η must be experimentally set in order to maximize the system performance.

The signal in Figure 2(a) is relative to the movements of a sleeping cat's chest. The signal was extracted from a video realized with a single camera with a sampling rate of 15 frames/s: rhythmic movements are clearly visible. In Figure 2(b), the corresponding likelihood function is shown as a function of f and the estimated frequency $\hat{f}$ (in correspondence to the peak of the likelihood function) is highlighted.

According to (3)–(6), the described ML method operates on sequences of N frames; hence, the periodicity can be analyzed on disjoint temporal windows of

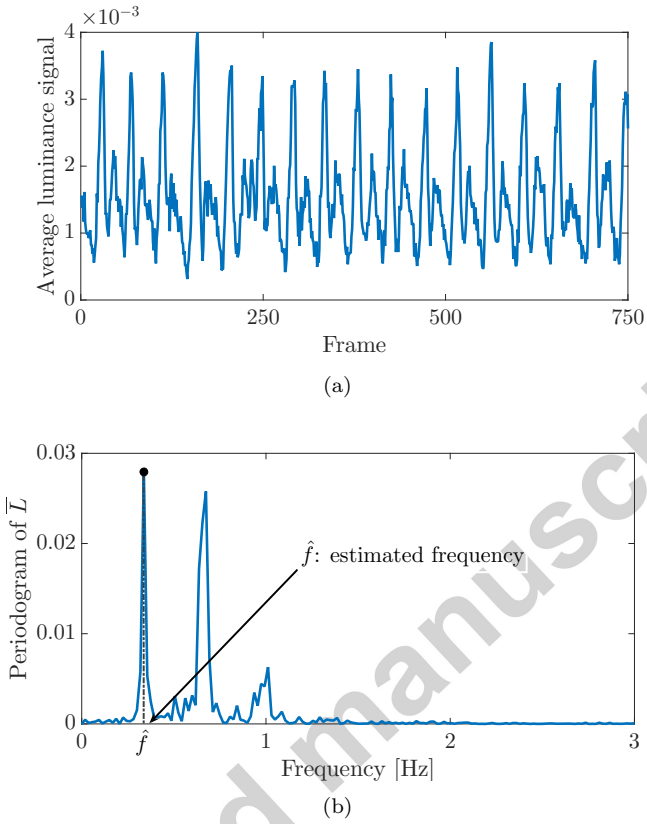
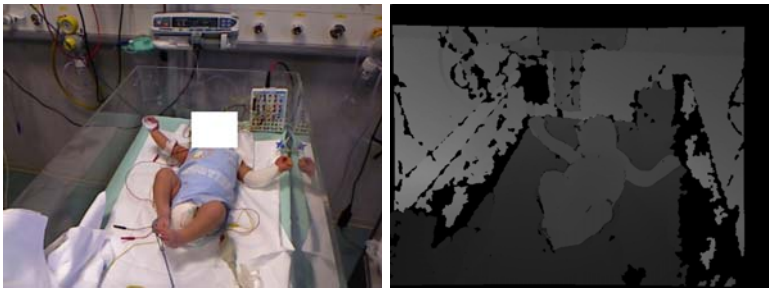


Figure 2: (a) Example of a rhythmic motion signal and (b) corresponding periodogram from the likelihood function.

duration NT. However, a motion-related symptom of a disease could manifest itself across two consecutive windows and the algorithm could miss the detection in both windows [16]. To avoid these cases and improve detection performance, the analysis of interlaced windows is considered, using an interlacing factor (i.e., the percentage of overlap of consecutive windows) equal to 50%.

2.3. Use of a Depth Sensor

Microsoft Kinect [19] is a device that associates a standard RGB video stream with a depth map stream. The provided depth information can be used to improve the ability of a standard video-based system to distinguish pathological movements from background noise or random movements not concerning the framed patient. Nevertheless, these depth estimating systems are afflicted by a significant and systematic issue: the shadowing noise [25]. This problem, common to all structured-light approaches which use an offset camera to determine a depth map, consists of the presence of regions where the projected pattern



(a) RGB image (b) Depth image

(c) Foreground mask image (d) Filtered depth image

Figure 3: Background-foreground segmentation starting from data acquired by RGB and depth sensors.

is shadowed by foreground objects, making it impossible to estimate the corresponding depth. Because of this issue, it becomes difficult to apply a simple difference between consecutive depth frames: in fact, in the processed image, besides movement parts, background areas could not be correctly detected.

For the sake of limiting the shadowing noise, a background-foreground segmentation approach can be used. To this purpose, the Mixture-of-Gaussian method described in [26], which is based on [27], is employed. The aim is to build a foreground mask, i.e., a binary image of the same size $W \times H$ of the video frame, where white and black pixels refer to ones in the original frame belonging to foreground or background, respectively. Applying the background-foreground segmentation to the RGB frame and the obtained mask to the depth frame, it is possible to filter out all unnecessary pixels.

Figure 3 shows: (a) an RGB image, (b) the associated depth map, (c) the foreground mask extracted from the RGB image, and (d) the depth map filtered by the foreground mask.

After the depth sequence has being denoised, according to the above procedure, the motion signal can be extracted by the same technique described in Subsection 2.1.

2.4. Application of Eulerian Video Magnification

The proposed algorithm in Subsections 2.1–2.3 is a reliable approach to monitor and analyze large and clearly visible body movements. If motions of interest are subtle, like respiration, they can be easily confused with noise or small environmental changes. The methods described above may not be sensitive enough to handle these movements, especially for breathing disorders, because the small excursion of a newborn's body parts may be difficult to detect. To improve detection of such small movements, a motion magnification algorithm can be applied; the proposed solution relies on a video pre-processing algorithm, in order to amplify motions of interest, followed by the approach described above. The employed motion magnification method, known as Eulerian Video Magnification (EVM) algorithm, was proposed in [28], to overcome the spatio-temporal restrictions of the human visual system for revealing subtle changes in video signals. The EVM algorithm consists of a cascade of spatial multi-scale frame decomposition and temporal processing, with a final reconstruction into a new video where small changes in the framed area are enhanced [28].

3. Performance Analysis

The acquisition system used at the NICU of the University Hospital of Parma and some illustrative examples of images acquired from a few cameras are shown in Figure 4. The system is composed of three cameras: the first two cameras are arranged orthogonally in order to frame the baby by the front and the side, the third one is attached to the cradle to frame the face of the newborn [4]. In the following, the use of the first camera only will be considered first. Then, performance results using all three cameras will be discussed. A multi-sensor system is less affected by the patient's position, because all main viewing angles are monitored by different sensors. Using a single sensor system, instead, the performance may degrade[1].

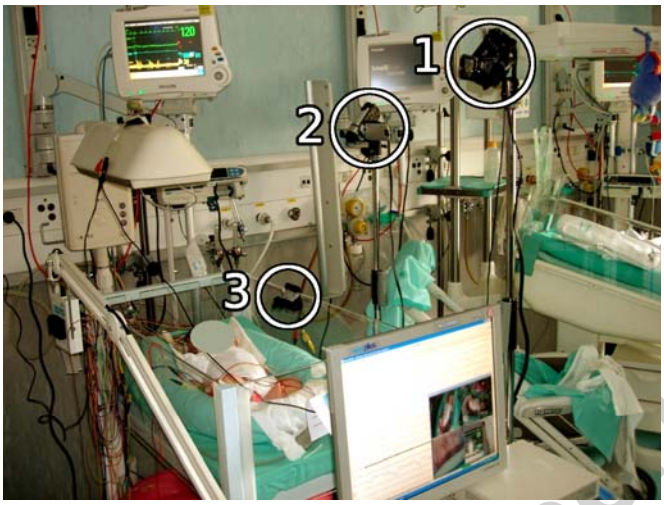
Examples of motion analysis by video processing according to the methods described in Section 2 are provided as supplemental material.

3.1. Detection of Neonatal Clonic Seizures

A neonatal clonic seizures is characterized by a duration of at least 10 s [1] with a period ranging between 0.5 s and 3.5 s [29]. Accordingly, the monitoring system described in Section 2 has been calibrated to temporal windows of 10 s, thus the value of N in (4) becomes $10f_s$, where f_s is the sampling rate. As an example, for a camera with a sampling rate of 25 frames/s, N will be 250 frames.

The performance of the proposed detection system is investigated considering a binary test, which classifies results as "presence of clonic seizures" in the video of the newborn (positive) and "absence of movements or presence of random

[1] This system is very versatile and is suitable for many kinds of sensors. It is convenient to use the most suitable sensor for the environment of interest. In case of low light conditions it is recommended the use of an infrared sensors or high performance cameras.



(a)

(b)

Figure 4: (a) Acquisition system; 1,2,3: position of the cameras and (b) corresponding frames.

movements" (negative). The performance of a clinical binary test is typically presented in terms sensitivity and specificity, which are defined, respectively, as follows [30]:

$$\alpha \triangleq \frac{n_{\mathrm{TP}}}{n_{\mathrm{TP}} + n_{\mathrm{FN}}} \tag{7a}$$

$$\beta \triangleq \frac{n_{\mathrm{TN}}}{n_{\mathrm{TN}} + n_{\mathrm{FP}}} \tag{7b}$$

where n_{TP}, n_{TN}, n_{FP}, and n_{FN} denote, respectively, the numbers of True Positives, True Negatives, False Positives, and False Negatives in the sequence of all considered windows. A perfect predictor would have a sensitivity $\alpha = 1$ (e.g., all sick individuals would be correctly identified as sick) and a specificity $\beta = 1$ (e.g., no healthy individuals would be incorrectly identified as sick). We shall refer to sensitivity and specificity values in terms of corresponding percentages, i.e. 100α and 100β, respectively.

Initially, results based on the retrospective off-line analysis of 10 single camera-based video recordings of newborns are presented. A single sensor, i.e., $S = 1$ in (4)–(6), and a standard linear combination of the RGB sig-

	Real Positive	Real Negative
Positive Test	$n_{\mathrm{TP}} = 51$	$n_{\mathrm{FP}} = 16$
Negative Test	$n_{\mathrm{FN}} = 7$	$n_{\mathrm{TN}} = 210$
Performance	Sensitivity: 88%	
	Specificity: 93%	

Table 1: Detection of clonic seizures (one B&W camera)

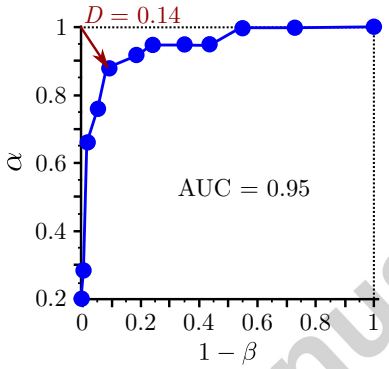


Figure 5: ROC curve, using one RGB sensor.

nals for conversion to gray scale are considered, with corresponding weights $w_R = 0.299$, $w_G = 0.587$ and $w_B = 0.114$ [31]. We refer to this sensor as Black & White (B&W). The set of videos has been chosen to represent all possible behaviors of a newborn (no movements, natural movements, pathological movements). Each video has these settings:

- sampling rate: 25 frames/s ($T = 40$ ms);
- resolution: 360 × 288 pixels;
- length: 5 min.

In order to evaluate the performance of the proposed system, we have considered only 284 of the 580 available interlaced windows, discarding windows without movements. In fact, without discarding these windows, an unrealistic increase in specificity (7b) can occur, because all the windows without movements would be too easily classified by the algorithm as true negatives. The corresponding results are shown in Table 1, where real patient conditions are compared with results obtained by the video-processing system in terms of n_{TP}, n_{TN}, n_{FP} and n_{FN}. The corresponding sensitivity and specificity are then evaluated according to (7a) and (7b) .

In Figure 5, the Receiver Operating Characteristic (ROC) curve [32], obtained by varying the decision threshold η in (6) and used to measure the performance of the system, is shown. The obtained minimum euclidean distance D

	Real Positive	Real Negative
Positive Test	$n_{\mathrm{TP}} = 44$	$n_{\mathrm{FP}} = 8$
Negative Test	$n_{\mathrm{FN}} = 14$	$n_{\mathrm{TN}} = 218$
Performance	Sensitivity: 76%	
	Specificity: 96%	

Table 2: Detection of clonic seizures (Red channel)

	Real Positive	Real Negative
Positive Test	$n_{\mathrm{TP}} = 49$	$n_{\mathrm{FP}} = 10$
Negative Test	$n_{\mathrm{FN}} = 8$	$n_{\mathrm{TN}} = 217$
Performance	Sensitivity: 86%	
	Specificity: 96%	

Table 3: Detection of clonic seizures (Green channel)

	Real Positive	Real Negative
Positive Test	$n_{\mathrm{TP}} = 44$	$n_{\mathrm{FP}} = 13$
Negative Test	$n_{\mathrm{FN}} = 13$	$n_{\mathrm{TN}} = 214$
Performance	Sensitivity: 77%	
	Specificity: 94%	

Table 4: Detection of clonic seizures (Blue channel)

from the "ideal" operational point $(0, 1)$, i.e., the point associated with sensitivity and specificity both equal to 1, is 0.14 and the Area Under Curve (AUC) is 0.95. Note that, in the medical literature, a test with AUC between 0.9 and 1 is considered as highly reliable [32].

Using the same set of video recordings, the achievable performance using just one color channel of a single RGB sensor is investigated. This option is equivalent to a particular linear combination in which only one of the three weights w_R, w_G and w_B is set to 1, whereas the other two are set to 0. The corresponding results are shown in Tables 2, 3 and 4 for the Red, Green and Blue channels, respectively. These tables can also be compared with Table 1 relative to the standard RGB linear conversion to gray scale (i.e., B&W). This analysis shows that the sole green channel yields a performance similar to that obtained by the use of the standard linear combination of the three channels. Since the standard conversion to gray scale is not necessarily optimal, a slightly larger value of β, using a single color channel, is indeed possible.

Multi-sensor video recordings can increase the reliability of the detection system. With this in mind, the performance of the system, using the three channels of the RGB sensor as if they were single monochrome sensors, is analyzed, combining the extracted signals according to the ML criterion in (4)–(6) with $S = 3$. In Table 5, the obtained results are shown. It may be worth under-

	Real Positive	Real Negative
Positive Test	$n_{\mathrm{TP}} = 50$	$n_{\mathrm{FP}} = 9$
Negative Test	$n_{\mathrm{FN}} = 7$	$n_{\mathrm{TN}} = 218$
Performance	Sensitivity: 88%	
	Specificity: 96%	

Table 5: Detection of clonic seizures – 3 RGB channels

	Real Positive	Real Negative
Positive Test	$n_{\mathrm{TP}} = 138$	$n_{\mathrm{FP}} = 10$
Negative Test	$n_{\mathrm{FN}} = 12$	$n_{\mathrm{TN}} = 78$
Performance	Sensitivity: 92%	
	Specificity: 88%	

Table 6: Detection clonic seizures (B&W and depth sensors)

lining how the use of the ML criterion allows to achieve a performance better than or equal to that of any of the considered linear combinations of the 3 RGB channels (i.e., gray scale or single color channels).

As a different application of multi-sensor processing, a gray-scale camera is augmented with a depth sensor [19], namely $S = 2$ is considered in (4)–(6). The necessary video and depth signals are obtained using an RGB-depth sensor in the position of camera 1 in Figure 4. Two video recordings of a newborn are considered: in the first one, he performs pathological movements related to a clonic seizures; in the second one, instead, he performs only physiological movements. Each video has these settings:

- sampling rate: 30 frames/s ($T = 33.333$ ms);
- resolution: 640×480 pixels;
- length: 10 min.

Considering 10-second half-interlaced windows, we could analyze 238 observation windows. The results shown in Table 6 confirm the effectiveness of the proposed approach.

Finally, the results based on the analysis of four recordings of a newborn with three cameras as shown in Figure 4, i.e., $S = 3$ in (4)–(6), are presented. Standard conversion to gray scale is used for each camera. In the first two videos, the baby performs pathological movements (related to clonic seizures); in the second two videos, instead, he performs only physiological movements. Each video has these settings:

- sampling rate: 25 frames/s ($T = 40$ ms);
- resolution: 360×288 pixels;
- length: 1 min.

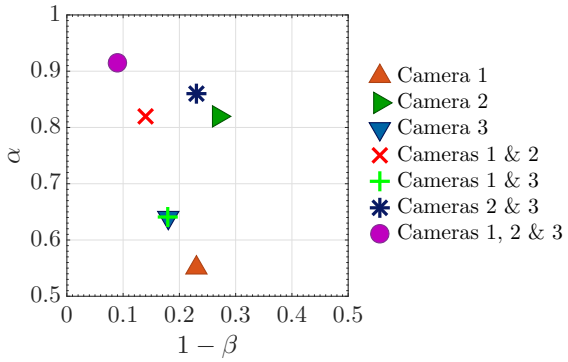


Figure 6: Performance of clonic seizure detection with different configurations of 3 RGB cameras.

Considering 10-second half-interlaced windows, we could analyze 44 observation windows. In Figure 6, the results obtained using individual cameras ($S = 1$) are compared with the results obtained using various combinations of the three available cameras ($S = 2$ or $S = 3$). Each considered case is represented by the point of coordinates $(\alpha, 1 - \beta)$. It clearly emerges how the use of multiple cameras may significantly improve the system performance, especially in the presence of clonic movements of limited amplitude, as in the case of the considered videos.

3.2. Detection of Apneas

Clinically, an apnea event is defined as an episode of absence of breathing lasting at least 20 s. The medical literature defines as apneas also the episodes of absence of breathing lasting between 10 s and 20 s, provided that they are associated with other clinical signs/symptoms [33]. In order to detect apnea events, we first emphasize the respiratory movements with the EVM algorithm presented in [28], then the proposed methods can be used to extract the chest motion signal from a set of sensors. In Figure 7, the motion signal extracted from a video sequence of an abnormal event is shown. A single RGB camera with conversion to gray scale is considered. As it can be observed, the time period without breathing movements is clearly highlighted.

We now presents the results of a retrospective analysis performed on the video of a newborn, who previously received a timely diagnosis of central apneas by means of a polysomnogram. As ground truth, the results obtained by the video-EEG polysomnography with thermistor and a pletismopgraph, lasting a total of 1 hour and 35 s, are considered. These indicate:[2] 23 apnea episodes

[2]Due to the documented dysregulation of respiratory control, during the recording the patient was also monitored by an elastic belt sensor, a nasal flow-meter and a pulse oximeter. Furthermore, oxygen therapy was applied using nasal cannulas whenever clinically indicated.

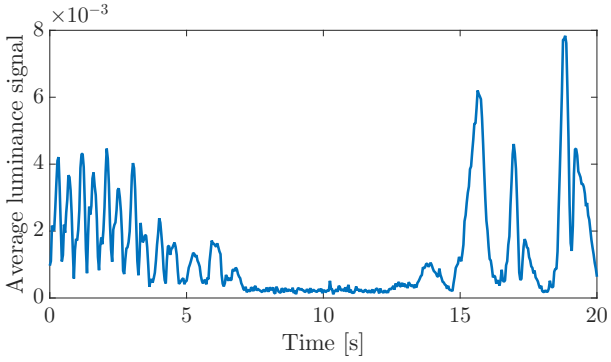


Figure 7: An illustrative example of extracted average luminance signal of a newborn experiencing an abnormal breathing event.

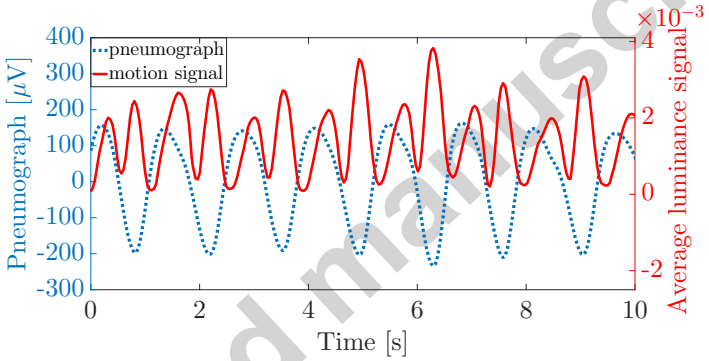


Figure 8: Comparison between the motion signal extracted by the proposed technique and the signal extracted from the pneumograph of the video-EEG device.

with a total duration of 28 min and 31 s, and an average single apnea duration of 74 s. In Figure 8, a comparison between the motion signal extracted by the proposed system and pneumographic signal, considered as the ground truth, is shown. For performance analysis, we discarded 6 of these available 23 apneas, because: in 4 of them the cameras were covered by medical personnel; 2 of them were too short (less than 20 s), since our system is set up to detect apneas of at least 20 s. The performance of the proposed monitoring system is investigated, once again, considering a binary test, which classifies results as "presence or absence of breathing movements." Sensitivity and specificity are redefined, in analogy with (7a) and (7b), in terms of T_{TP}, T_{TN}, T_{FP}, and T_{FN}, which denote, respectively, the total duration of the time intervals with apnea correctly detected (True Positives), no apnea correctly detected (True Negatives), normal breathing incorrectly reported as apnea (False Positives) and apnea incorrectly reported as normal breathing (False Negatives). In the evaluation of these time intervals, a Tolerance Delay (TD), defined with respect to the end of the first processing window where the apnea episode can be detected, is considered. An

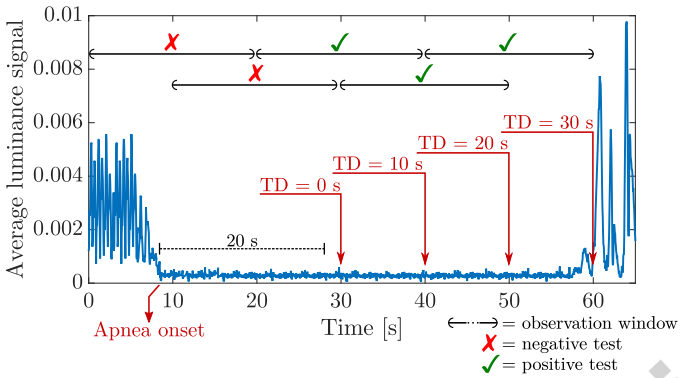


Figure 9: Half-interlaced observation windows of 20 s length, with TD of 10, 20 and 30 s, apnea starting at 8 s and lasting approximately 47 s.

Performance on Apnea Detection							
TD	DA	T_{TP}	T_{TN}	T_{FP}	T_{FN}	α [%]	β [%]
0	13	1200	1800	500	140	90%	78%
10	14	1230	1800	500	110	91%	78%
20	16	1280	1880	420	60	95%	81%
30	17	1340	1920	380	0	100%	83%

Legend: TD = Tolerance Delay (seconds); DA = number of Detected Apneas; T_{TP}, T_{TN}, T_{FP} and T_{FN} (seconds).

Table 7: Real Apneas Detection (one RGB camera)

apnea episode is assumed correctly detected if the algorithm reports it within a time interval not longer than the assigned TD. Figure 9 helps to better understand this concept, as here a long apnea is considered and consecutive interlaced windows are shown: different TDs are shown jointly with consecutive observation windows. As TD increases, a larger number of interlaced windows fall inside the TD, thus increasing the probability of correctly detecting the presence of an apnea. In the upper part of Figure 9, examples of detection results are shown; for a value of TD equal to 0 s, the apnea episode is missed, whereas for TD equal to 10 s, 20 s or 30 s, it is correctly detected. In Table 7, the sensitivity and specificity values obtained by the proposed method are shown, considering various values of the assumed TD. In particular, sensitivity and specificity are computed by considering the portions of minutes of apnea (out of 22 min) and regular breathing (out of 38 min) correctly/incorrectly diagnosed. The proposed system was able to correctly detect from 13 to 17 episodes, out of a total of 17, depending on the acceptable TD. Interestingly, all apnea episodes are correctly detected using the maximum TD of 30 s.

We remark that the system was tested also in the presence of blankets cover-

ing the newborn. The obtained results, with an example visible in the supporting material, show that the small movement enhancement method may indeed enable a correct detection of the diseases.

4. Conclusion

This paper describes a unified wire-free, low-cost, non-invasive technique for automatic detection of abnormal paroxysmal events in newborns, namely, clonic seizures and life-threatening apnea events. According to the proposed approach, motion signals are extracted from RGB cameras or depth sensors in order to estimate whether a periodic component is present, which is the characteristic feature of clonic seizures or breathing movements. In the case of apnea detection, the video may be pre-processed using the EVM technique, in order to emphasize small breathing movements (e.g., chest/abdomen movements). The diagnostic performance of the detection systems has been analyzed by comparison with the gold standard represented by the video-EEG polysomnography. The results are very promising, as they show that it is indeed possible to identify both clonic seizures or apnea events with contact-less devices. This innovative detection system could represent a timely, user-friendly and 24/7 tool to be used in the NICU or at the patient's home in order to monitor newborns with severe diseases.

Standard Protocol Approvals, Registrations, and Patient Consents

In accordance with current practice at our Institution, an informed consent form was signed by parents of each newborn patient, and the informed consent was placed in the patients' hospital chart. The Ethic Local Committee approval to perform retrospective studies was obtained.

Conflict of Interest Statement

None.

References

[1] J. J. Volpe, Neurology of the Newborn, 5th Edition, Saunders Elsevier, Philadelphia, PA, USA, 2008.

[2] C. P. Panayiotopoulos, The Epilepsies: Seizures, Syndromes and Management, 1st Edition, Bladon Medical Publishing, Chipping Norton, UK, 2005.

[3] H. C. Glass, E. Wirrell, Controversies in neonatal seizure management, J. Child Neurol. 24 (5) (2009) 591–599. doi:10.1177/0883073808327832.

[4] F. Pisani, E. Pavlidis, L. Cattani, G. Ferrari, R. Raheli, C. Spagnoli, Optimizing detection rate and characterization of subtle paroxysmal neonatal abnormal facial movements with multi-camera video-electroencephalogram recordings, Neuropediatrics 47 (3) (2016) 169–174. doi:10.1055/s-0036-1582245.

[5] G. M. Kouamou Ntonfo, F. Lofino, G. Ferrari, R. Raheli, F. Pisani, Video processing-based detection of neonatal seizures by trajectory features clustering, in: IEEE Int. Conf. Commun. (ICC 2012), Ottawa, Canada, 2012, pp. 3456–3460. doi:10.1109/ICC.2012.6364396.

[6] G. M. Kouamou Ntonfo, G. Ferrari, F. Lofino, R. Raheli, F. Pisani, Extraction of video features for real-time detection of neonatal seizures, in: IEEE Int. Symp. World of Wireless, Mobile and Multimedia Netw. (WoWMoM 2011), Lucca, Italy, 2011, pp. 1–6. doi:10.1109/WoWMoM.2011.5986193.

[7] Mosby, Mosby's Medical Dictionary, 10th Edition, Elsevier, Saint Louis, MO, USA, 2016.

[8] J. J. Tramonte, H. P. Goodkin, Temporal lobe hemorrhage in the full-term neonate presenting as apneic seizures, J. Perinatol. 24 (11) (2004) 726–729. doi:10.1038/sj.jp.7211181.

[9] C. M. Cielo, C. L. Marcus, Central hypoventilation syndromes, Sleep Med. Clin. 9 (1) (2014) 105–118. doi:10.1016/j.jsmc.2013.10.005.

[10] C. M. Rand, P. P. Patwari, M. S. Carroll, D. E. Weese-Mayer, Congenital central hypoventilation syndrome and sudden infant death syndrome: Disorders of autonomic regulation, Semin. Pediatr. Neurol. 20 (1) (2013) 44–55. doi:10.1016/j.spen.2013.01.005.

[11] R. J. Martin, A. J. Block, M. A. Cohn, W. A. Conway, D. W. Hudgel, A. P. Powles, M. H. Sanders, P. L. Smith, Indications and standards for cardiopulmonary sleep studies, Sleep 8 (4) (1985) 371–379.

[12] W. B. Spillman Jr., M. Mayer, J. Bennett, J. Gong, K. E. Meissner, B. Davis, R. O. Claus, A. A. Muelenaer Jr., X. Xu, A 'smart' bed for non-intrusive monitoring of patient physiological factors, Meas. Sci. Technol. 15 (8) (2004) 1614–1620. doi:10.1088/0957-0233/15/8/032.

[13] A. Pantelopoulos, N. G. Bourbakis, A survey on wearable sensor-based systems for health monitoring and prognosis, IEEE T. Syst. Man. Cyb. 40 (1) (2010) 1–12. doi:10.1109/TSMCC.2009.2032660.

[14] N. B. Karayiannis, S. Srinivasan, R. Bhattacharya, M. S. Wise, J. D. Frost Jr., E. M. Mizrahi, Extraction of motion strength and motor activity signals from video recordings of neonatal seizures, IEEE Trans. Med. Imag. 20 (9) (2001) 965–980. doi:10.1109/42.952733.

[15] N. B. Karayiannis, E. M. Mizrahi, B. Varughese, G. Tao, J. D. Frost Jr., M. S. Wise, Quantifying motion in video recordings of neonatal seizures by regularized optical flow methods, IEEE Trans. Image Process. 14 (7) (2005) 890–903. doi:10.1109/TIP.2005.849320.

[16] G. M. Kouamou Ntonfo, G. Ferrari, R. Raheli, F. Pisani, Low-complexity image processing for real-time detection of neonatal clonic seizures, IEEE Trans. Inf. Technol. Biomed. 16 (3) (2012) 375–382. doi:10.1109/TITB.2012.2186586.

[17] J. Jacobs, Detecting neonatal seizures: A challenge accepted!, Clin. Neurophysiol. 125 (8) (2014) 1501–1503. doi:10.1016/j.clinph.2014.02.003.

[18] S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Vol. 1, Prentice Hall, Upper Saddle River, NJ, USA, 1993.

[19] J. Webb, J. Ashley, Beginning Kinect programming with the Microsoft Kinect SDK, Apress, New York, NY, USA, 2012.

[20] L. Cattani, H. P. Saini, G. Ferrari, F. Pisani, R. Raheli, SmartCED: An Android application for neonatal seizures detection, in: IEEE Int. Symp. Med. Meas. and Applicat. (MeMeA), Benevento, Italy, 2016, pp. 1–6. doi:10.1109/MeMeA.2016.7533708.

[21] L. Cattani, G. M. Kouamou Ntonfo, F. Lofino, G. Ferrari, R. Raheli, F. Pisani, Maximum-likelihood detection of neonatal clonic seizures by video image processing, in: 8th Int. Symp. Med. Inf. and Commun. Technol. (ISMICT), Florence, Italy, 2014, pp. 1–5. doi:10.1109/ISMICT.2014.6825219.

[22] L. Cattani, D. Alinovi, G. Ferrari, R. Raheli, E. Pavlidis, C. Spagnoli, F. Pisani, A wire-free, non-invasive, low-cost video processing-based approach to neonatal apnoea detection, in: IEEE Workshop Biometric Meas. and Syst. Security and Med. Applicat. (BIOMS), Rome, Italy, 2014, pp. 67–73. doi:10.1109/BIOMS.2014.6951538.

[23] R. C. Gonzalez, R. E. Woods, Digital Image Processing, 3rd Edition, Pearson - Prentice Hall, Upper Saddle River, NJ, USA, 2010.

[24] N. Patwari, J. Wilson, S. Ananthanarayanan, S. Kasera, D. Westenskow, Monitoring breathing via signal strength in wireless networks, IEEE Trans. Mobile Comput. 13 (8) (2014) 1774–1786. doi:10.1109/TMC.2013.117.

[25] T. Mallick, P. P. Das, A. K. Majumdar, Characterizations of noise in Kinect depth images: A review, IEEE Sensors J. 14 (6) (2014) 1731–1740. doi:10.1109/JSEN.2014.2309987.

[26] P. Kaew Tra Kul Pong, R. Bowden, Video-Based Surveillance Systems: Computer Vision and Distributed Processing, Springer Science + Business

Media, New York, NY, USA, 2002, Ch. 11 An improved adaptive background mixture model for real-time tracking with shadow detection, pp. 135–144. doi:10.1007/978-1-4615-0913-4_11.

[27] C. Stauffer, W. E. L. Grimson, Adaptive background mixture models for real-time tracking, in: IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognition, Vol. 2, Fort Collins, CO, USA, 1999, pp. 246–252. doi:10.1109/CVPR.1999.784637.

[28] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, W. Freeman, Eulerian video magnification for revealing subtle changes in the world, ACM Trans. Graph. 31 (4) (2012) 65:1–65:8. doi:10.1145/2185520.2185561.

[29] F. Pisani, C. Spagnoli, E. Pavlidis, C. Facini, G. M. Kouamou Ntonfo, G. Ferrari, R. Raheli, Real-time automated detection of clonic seizures in newborns, Clin. Neurophysiol. 125 (8) (2014) 1533–1540. doi:10.1016/j.clinph.2013.12.119.

[30] A. Lalkhen, A. McCluskey, Clinical tests: sensitivity and specificity, Contin. Educ. Anaesth. Crit. Care Pain 8 (6) (2008) 221–223. doi:10.1093/bjaceaccp/mkn041.

[31] A. Bovik, The Essential Guide to Image Processing, 1st Edition, Academic Press, San Diego, CA, USA, 2009.

[32] J. Swets, Measuring the accuracy of diagnostic systems, Science 240 (4857) (1998) 1285–1293. doi:10.1126/science.3287615.

[33] N. N. Finer, R. Higgins, J. Kattwinkel, R. J. Martin, Summary proceedings from the apnea-of-prematurity group, Pediatrics 117 (Supplement 1) (2006) S47–S51. doi:10.1542/peds.2005-0620H.