Respiratory Rate Monitoring by Video Processing Using Local Motion Magnification

Davide Alinovi*, Gianluigi Ferrari*, Francesco Pisani§ and Riccardo Raheli*
*Department of Engineering and Architecture, University of Parma, Italy
§Department of Medicine and Surgery, University of Parma, Italy
E-mail: {davide.alinovi, gianluigi.ferrari, francesco.pisani, riccardo.raheli}@unipr.it

***Abstract*—Breathing monitoring by non-contact video processing has been the subject of recent research. This paper presents an advanced video processing algorithm for reliable Respiratory Rate (RR) monitoring based on the analysis of local video variations and a motion magnification method. This novel algorithm improves over the existing solutions in terms of estimation accuracy and excision of large body movements unrelated with respiration. Applications to adults and infants are presented to demonstrate the performance of the proposed algorithm and compare it with previous work.**

I. Introduction

Respiration is an important physiological task and plays a fundamental role in human body functions. It consists in repetitive movements, performed by specific muscles, to "import" oxygen (inhalation) and "throw out" carbon dioxide (exhalation). The Respiratory Rate (RR) is one of main vital signs and is defined as the number of breathing cycles in a time unit. Changes in the RR need to be carefully monitored, as potential signs of deterioration of health condition.

Recent research has addressed non-contact and non invasive monitoring systems of vital signs [1]; in this scenario, video processing-based methods are very attractive [2]–[4].

Body movements can be exploited for the analysis of the breathing effort. Since tiny movements are involved, motion magnification techniques [3] and analysis of pixel-wise variations [5] were exploited. These systems are among the first to present reliable video processing schemes to monitor effectively the RR. However, the work in [3] is sensitive to large body movements and scenario variations; moreover, in various conditions, it may estimate a RR doubled respect to the real one. The method in [5] requires pixel-wise analysis of video streams and may be computationally inefficient for possible real-time applications.

In this paper, a new method for non-invasive measurement of the RR by video processing is proposed to improve over the previous approaches. The main contribution is an algorithm for the extraction of motion information, relying on a recent and effective subtle motion enhancement technique [6]. Moreover, the system increases efficiency by selecting suitable regions mainly related with breathing movements and integrating a smart method to limit the interference due to environmental changes, other people/medical staff or large movements of the framed person/patient. Results are presented considering two different conditions: first the method is tested on healthy volunteers in a controlled environment; then, preliminary results on hospitalized newborn patients are proposed.

The remainder of the paper is organized as follows. In Section II, the algorithm for breathing motion information extraction is described. In Section III, validation of the method and performance on RR estimation are presented. Finally, conclusions are drawn in Section IV.

II. Video Processing Methodology for Respiration Analysis

Relying on the description of the breathing cycle introduced in Section I, chest wall movements can be detected and analyzed to monitor the respiration and estimate the RR.

Applying motion magnification to patient monitoring, the method used for the detection of tiny breathing movements of the system is inspired by [6]; moreover, it limits problems related to large global movements introducing local motion analysis based on Regions of Interest (ROI), with an approach similar to the one presented in [5]. In Fig. 1, a diagram of the proposed system is shown: the blocks and the involved steps will be described in the following subsections.

Before describing the algorithm, useful notation is introduced. A video stream acquired by a camera sensor is defined as a multidimensional signal consisting of a sequence of digital images with $U_1 \times U_2$ pixels acquired over time with a sampling rate f_s, equivalent to a period $T_s = 1/f_s$. Mathematically, each frame at time instant nT_s can be described as a discrete function $f[\mathbf{u}, n]$ containing pixel intensity values for each spatial coordinate $\mathbf{u} = (u_1, u_2)$ and having U_1 rows and U_2 columns. For multi-chromatic videos, a proper number of color channels has to be considered, depending to the sensor color space [7]. In this work, videos are recorded by Red, Green and Blue (RGB) cameras: to reduce the computational complexity and simplify the notation, the performed processing is based on a single-channel gray scale video, after proper conversion [7].

A. Selection of Areas of Interest

The first step consists of the automatic selection of the ROI in which the frame variations are related with breathing movements. To this purpose, extraction of periodic features is needed. A block of N frames of the original video $f[\mathbf{u}, n]$ is analyzed; after estimating the common fundamental frequency, the amplitude of the periodic component for each

ISBN 978-90-827970-1-5

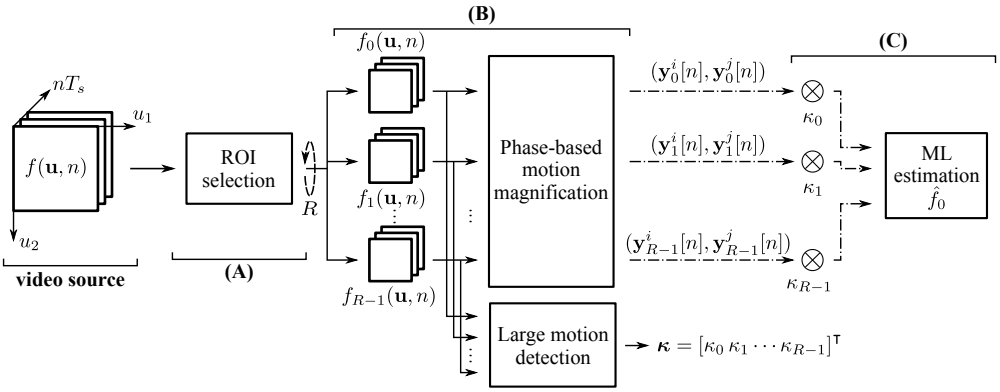

Fig. 1. Overview of the proposed video processing-based framework. The system includes: ROI identification and selection (A), breathing motion magnification and detection (B) and, finally, RR estimation (C) of the framed patient.

pixel may be obtained. The matrix of estimated amplitudes can be interpreted as a map describing the position of periodic variations inside the video, where higher values correspond to larger periodic components. For the details on this approach to periodic features extraction, the reader is referred to [8]. Restricting to the context of respiratory monitoring, such areas correspond to the ones where breathing movements cause main variations in pixel intensities.

From this map, R areas with size $W \times W$ can be selected, choosing the regions where the amplitude of the estimated periodic variations are higher: in Fig. 1 this step is indicated by part (A) and the algorithm for the extraction of the ROI is described in [5]. The selection of ROI leads to the extraction of smaller video streams, denoted as $f_r[\mathbf{u},n]$, where $r \in \{0,1,\ldots,R-1\}$.

B. Breathing Information: from Video to Motion Signals

The core of the proposed algorithm exploits the principles of the advanced motion magnification system presented in [6] for the extraction of breathing information. This corresponds to the block (B) shown in Fig. 1; it is applied on each selected ROI and composed of three steps: multi-scale decomposition, phase-based motion magnification and extraction of signals representing local movements.

The extraction of breathing information is performed in the same way for each region $f_r[\mathbf{u},n]$. However, to simplify the notation, in this subsection the method is presented on a generic video $f[\mathbf{u},n]$.

1) Complex Pyramidal Decomposition: Given a video $f[\mathbf{u},n]$, it is first processed frame-by-frame to obtain a multi-scale spatial decomposition based on a Laplacian pyramid [9]. Specifically, a single frame is decomposed into L images $\{p_\ell[\mathbf{u},n]\}_{\ell=0}^{L-1}$ with scaled resolution, called levels, representing different spatial frequency sub-bands. Then, an analytic representation of the images is invoked.

As in one dimension (1D), the analytic signal and related local amplitude and phase can be obtained through the Hilbert transform [10]. A multi-dimensional generalization can be defined, named monogenic signal, strictly related with the Riesz transform [11]. The latter can be defined in the Fourier domain: in the two-dimension (2D) case, given the vector of normalized angular frequencies $\boldsymbol{\omega} = (\omega_1,\omega_2)$, the transform can be identified with the two components: $H_1(\boldsymbol{\omega}) = -j\omega_1/\left\|\boldsymbol{\omega}\right\|$ and $H_2(\boldsymbol{\omega}) = -j\omega_2/\left\|\boldsymbol{\omega}\right\|$, where $\|\cdot\|$ is the norm of a vector. Given the input $p_\ell[\mathbf{u},n]$, the Riesz transform of the ℓ-th level can be expressed as

$$\mathcal{R}\left\{p_\ell[\mathbf{u},n]\right\} = \begin{pmatrix} r_{1,\ell}[\mathbf{u},n] \\ r_{2,\ell}[\mathbf{u},n] \end{pmatrix} = \begin{pmatrix} h_1[\mathbf{u}] * p_\ell[\mathbf{u},n] \\ h_2[\mathbf{u}] * p_\ell[\mathbf{u},n] \end{pmatrix} \tag{1}$$

where $h_i[\mathbf{u}] = \mathcal{F}^{-1}\left(H_i(\boldsymbol{\omega})\right)$, $i \in \{1,2\}$, $*$ denotes 2D convolution and the operator $\mathcal{F}^{-1}(\cdot)$ is the inverse 2D Fourier transform.

Applying the Riesz transform on the levels $\{0,1,\ldots,L-2\}$ of the pyramid, the set of three elements

$$\mathbf{p}_{\mathrm{m},\ell}[\mathbf{u},n] = (p_\ell[\mathbf{u},n], r_{1,\ell}[\mathbf{u},n], r_{2,\ell}[\mathbf{u},n]) \tag{2}$$

is obtained and defined as the monogenic signal of the ℓ-th level. The set $\{\mathbf{p}_{\mathrm{m},\ell}[\mathbf{u},n]\}_{\ell=0}^{L-2}$ jointly with $p_{L-1}[\mathbf{u},n]$ is named as the Riesz pyramid.

Given the 2D monogenic signal of the ℓ-th levels, local amplitude, phase and orientation can be respectively defined [12]

$$A_\ell = \sqrt{p_\ell^2 + r_{1,\ell}^2 + r_{2,\ell}^2} \tag{3a}$$

$$\varphi_\ell = \arctan\left(\left(\sqrt{r_{1,\ell}^2 + r_{2,\ell}^2}\right)\Big/ p_\ell\right) \tag{3b}$$

$$\vartheta_\ell = \arctan\left(r_{2,\ell}/r_{1,\ell}\right) \tag{3c}$$

where dependency on $[\mathbf{u},n]$ is omitted for brevity. To solve problems of sign ambiguity on phase and orientation, a quaternionic representation of the Riesz pyramid can be exploited [13]. So, the monogenic triple in (2) can be represented as the quaternion

$$\mathbf{q}_\ell[\mathbf{u},n] = p_\ell[\mathbf{u},n] + i\, r_{1,\ell}[\mathbf{u},n] + j\, r_{2,\ell}[\mathbf{u},n] + k \cdot 0 \tag{4}$$

with imaginary units i, j and k and where, following quaternionic algebra, the amplitude in (3a) and the complex quaternionic phase [13] can be expressed as

$$A_\ell = \left\|\mathbf{q}_\ell\right\| \tag{5a}$$

$$i\,\varphi_\ell\cos\left(\vartheta_\ell\right) + j\,\varphi_\ell\sin\left(\vartheta_\ell\right) = \log\left(\mathbf{q}_\ell/\left\|\mathbf{q}_\ell\right\|\right) \tag{5b}$$

ISBN 978-90-827970-1-5

where $\|\cdot\|$ now denotes the quaternionic norm, $\log(\cdot)$ is the logarithm of a quaternion and, as in (3), dependency on $[\mathbf{u}, n]$ is omitted.

2) Phase-Based Motion Magnification: After the pyramidal decomposition, amplification of small movements is needed. It relies on a pixel-wise processing with the ability to process local phases of each spatial sub-bands by temporal filters in order to select, and then amplify, only motion of interest [14].

The first step consists of temporal filtering the local phase. As the breathing movements are periodic with a repetition time related with the age and health condition of the monitored patient, a band-pass filter may be used to extract phase variations on frequencies of interest. To keep low the computational complexity and the delay introduced by the temporal filter, a first order Infinte Impulse Response (IIR) Butterworth digital band-pass filter is employed. This is designed selecting proper cut-off frequencies $f_{\mathrm{co},l}$ and $f_{\mathrm{co},h}$ [10], named lower and higher cut-off frequency, respectively, and with $f_{\mathrm{co},l} < f_{\mathrm{co},h}$. The filter is applied on the phase computed at each level of the corresponding Riesz pyramid.

To avoid issues on wrapped phase [13], the pixel-wise temporal processing does not filter directly the local phases as defined in (5a), but a cumulative sum of unwrapped quaternionic phases: in this way, at the n-th discrete time the following coefficients are computed

$$\begin{cases} \log\left(\bar{\mathbf{q}}_\ell[\mathbf{u}, n]\right) & \text{for } n = 0, \\ \log\left(\bar{\mathbf{q}}_\ell[\mathbf{u}, n]\bar{\mathbf{q}}_\ell^{-1}[\mathbf{u}, n-1]\right) & \text{for } n = 1, 2, \ldots \end{cases} \tag{6}$$

where the term $\bar{\mathbf{q}}_\ell[\mathbf{u}, n] = \frac{\mathbf{q}_\ell[\mathbf{u},n]}{\|\mathbf{q}_\ell[\mathbf{u},n]\|}$ is the monogenic normalized quaternion. As (6) for $n > 0$ represents the quantity

$$\begin{aligned} &i\left(\varphi_\ell[\mathbf{u}, n] - \varphi_\ell[\mathbf{u}, n-1]\right)\cos(\vartheta_\ell[\mathbf{u}])+ \\ &j\underbrace{\left(\varphi_\ell[\mathbf{u}, n] - \varphi_\ell[\mathbf{u}, n-1]\right)}_{\text{phase difference: } \varphi'_\ell[\mathbf{u},n]}\sin(\vartheta_\ell[\mathbf{u}]) \end{aligned} \tag{7}$$

where the local orientation $\vartheta_\ell[\mathbf{u}]$ is supposed to be time invariant, the cumulative sum of its terms is

$$i\varphi''_\ell[\mathbf{u}, n]\cos(\vartheta_\ell[\mathbf{u}]) + j\varphi''_\ell[\mathbf{u}, n]\sin(\vartheta_\ell[\mathbf{u}]) \tag{8}$$

where the unwrapped phase is[1]

$$\varphi''_\ell[\mathbf{u}, n] = \varphi_\ell[\mathbf{u}, 0] + \sum_{k=1}^{n}\varphi'_\ell[\mathbf{u}, k] \quad \text{for } n = 1, 2, \ldots. \tag{9}$$

The IIR temporal filter is now applied at every pixel of the quantity in (8) to isolate periodic movements, obtaining the two imaginary quaternionic components $i\delta_\ell[\mathbf{u}, n]\cos(\vartheta[\mathbf{u}])$ and $j\delta_\ell[\mathbf{u}, n]\sin(\vartheta[\mathbf{u}])$.

The filtered quaternionic phase now defines the translation in space, coherently with the selected temporal frequencies, due to motion included in the video signal. The motion information of the framed object is contained inside this phase. Following the phase-based magnification approach in [14], such a phase can be amplified to increase the translational movement. To this purpose, the two imaginary components of the quaternion containing the phase information $\delta_\ell[\mathbf{u}, n]$ are multiplied by a coefficient $\alpha > 1$.

As discussed in Section II-B, the described method is applied on each of the R selected ROI: therefore R filtered and amplified quaternionic phase differences $i\,\alpha\delta_{r,\ell}[\mathbf{u}, n]\cos(\vartheta_r[\mathbf{u}])$ and $j\,\alpha\delta_{r,\ell}[\mathbf{u}, n]\sin(\vartheta_r[\mathbf{u}])$ are obtained, $r = 1, 2, \ldots, R$.

3) Regional Motion Information Extraction: Local filtered and amplified cumulative quaternionic phases computed for each $r \in \{0, 1, \ldots, R-1\}$ can vary with positive and negative values depending on the direction of the local movement. This can be assimilated to the Fourier shifting theorem [10], in which a delay or an advance in time is closely related to the sign of the phase shift in the frequency domain. As ROI are employed, it is expected that the variations inside them is almost coherent, namely there are not areas which move in opposite directions. In this way, a reliable solution to obtain signals describing the motion inside the ROI is based on computing the spatial average of the local phase variations as functions of time, which can be obtained as

$$\begin{aligned} y^i_{r,\ell}[n] = &\quad \frac{1}{W^2}\sum_{u_1=0}^{W-1}\sum_{u_2=0}^{W-1}\alpha\delta_{r,\ell}[\mathbf{u}, n]\cos(\vartheta_r[\mathbf{u}]) \\ y^j_{r,\ell}[n] = &\quad \frac{1}{W^2}\sum_{u_1=0}^{W-1}\sum_{u_2=0}^{W-1}\alpha\delta_{r,\ell}[\mathbf{u}, n]\sin(\vartheta_r[\mathbf{u}]) \end{aligned} \tag{10}$$

where the dependency on r is now specified.

C. Data Fusion for Respiratory Rate Estimation

Once motion signals are extracted, an approach for RR estimation is needed. The aim of the last part of the RR analysis algorithm, denoted as block (C) in Fig. 1, is to use data information from various regions and estimate the principal periodic component: this method is reliable because the fundamental frequency obtained from motion signals is assumed to be strictly related to respiration. Moreover, the algorithm is able to exclude ROI affected by large movement not related to respiration, which can invalidate the estimation of the RR.

Signals $y^i_{r,\ell}[n]$ and $y^j_{r,\ell}[n]$ describe motion for each ROI and on different levels of the Riesz pyramid for the two imaginary quaternionic components. Such signals are expected to be quasi-periodic as breathing is intrinsically a repetitive movement whose frequency can vary over time. Assuming that they are representative of the respiration cycle, it is possible to group them in a matrix $\mathbf{Y}[n]$ with L rows and $2R$ column defined as

$$\mathbf{Y}[n] = \left[\ \mathbf{Y}^i[n]\,\mathrm{diag}\,(\boldsymbol{\kappa})\ \middle|\ \mathbf{Y}^j[n]\,\mathrm{diag}\,(\boldsymbol{\kappa})\ \right] \tag{11}$$

where

$$\mathbf{Y}^\iota[n] = \left[\ \mathbf{y}^\iota_0[n] \quad \mathbf{y}^\iota_1[n] \quad \cdots \quad \mathbf{y}^\iota_{R-1}[n]\ \right] \tag{12}$$

are the matrices for the two imaginary quaternionic components with $\iota \in \{i, j\}$, in which

$$\mathbf{y}^\iota_r[n] = \left[\ y^\iota_{r,0}[n] \quad y^\iota_{r,1}[n] \quad \cdots \quad y^\iota_{r,L-1}[n]\ \right]^\mathsf{T} \tag{13}$$

[1] $\varphi_\ell[\mathbf{u}, n] = \varphi''_\ell[\mathbf{u}, n] \mod 2\pi$.

ISBN 978-90-827970-1-5 © EURASIP 2018

and $\boldsymbol{\kappa} = [\kappa_0 \kappa_1 \cdots \kappa_{R-1}]^\intercal$ is the vector of binary parameters which includes/excludes ROI depending on the absence/presence of large motion.

With the introduced notation, a possible model of periodicity for RR estimation may be defined as

$$\mathbf{Y}[n] = \mathbf{C} + \mathbf{A}\cos\left(2\pi f_0 T_s n + \mathbf{\Phi}\right) + \mathbf{W}[n] \qquad (14)$$

where $\mathbf{C}$ are continuous components, $\{\mathbf{W}[n]\}$ are sequences of independent and identically distributed (i.i.d.) zero-mean Gaussian noise samples, $\mathbf{A}$ are amplitudes, f_0 is the common frequency of the periodic movements, $\mathbf{\Phi}$ are temporal phases, T_s is frame sampling period and the addition of a scalar to a matrix and trigonometric functions are applied element-wise. A large motion avoidance system is applied extracting a motion signal from the original ROI videos following the method in [5]: if the r-th area includes large movements over a heuristically chosen threshold, this area is excluded from RR estimation. Following the generalized Maximum Likelihood (ML) approach described in [8], data fusion can be exploited and the estimation $\hat{f}_0$ of the fundamental frequency on a window with duration NT_s can be obtained. The estimation $\hat{f}_0$ is defined as the RR of the monitored patient.

III. Validation and Results

The introduced method is now validated and results on RR estimation for adults and newborns are presented. A video camera is placed in front of or laterally with respect to a steady subject, framing the upper part for adults or the whole body for newborns. At first, motion signals are compared with references: accelerometric signal for adults and pneumogram for infants. The employed hardware consists of Shimmer3 sensors by Shimmer Sensing™ placed on the chest of the adult, whereas the pneumogram is recorded by an elastic belt around the newborn's chest. Motion signals are then compared with a previously proposed method [3], here referred to as Spatio-Temporal video-processing for RR Estimation (STRE), showing significant improvement. Finally, performance in RR estimation, also compared with STRE, is carried out.

In order to verify the ability of the proposed algorithm to extract signals strictly related with respiration, a comparison with the reference signals is proposed. In Fig. 2(a) a 20 s window of signal obtained by processing a video recording of an infant is compared with the corresponding portion of pneumogram. In Fig. 2(b) a signal of the same time duration extracted by a video stream framing an adult at rest is compared with the corresponding accelerometric signal. In both cases the video-processing system shows an excellent agreement with the signal employed as gold-standard. Differences on the reference signals in Fig. 2 can be ascribed to the technologies employed as gold standard and the age of monitored patient.

In previous works, signals describing framed movements were obtained by analyzing frame variations with a Difference of Frames (DoF) algorithm or proper IIR frame filtering, frame thresholding and extraction of an average luminance signal [3], [15]. These signals describe the "amount" of motion detected inside the frames calculating the amount of the pixel intensity

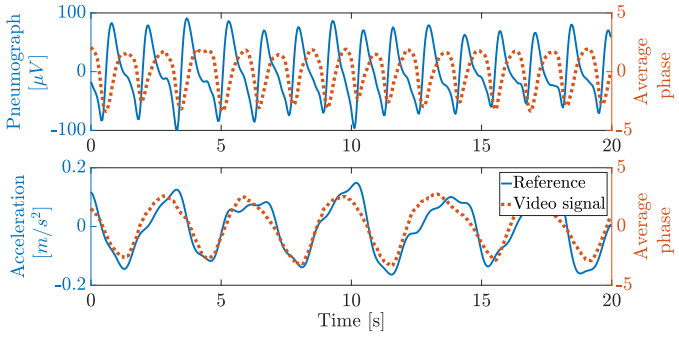

Fig. 2. Average local phase variations compared with the reference signals for RR estimation: (a) a pneumogram of a newborn and (b) an accelerometric sensor on an adult.

variations over time: this quantity is always positive and it is unable to distinguish motions in opposite directions.

In Fig. 3(a) a comparison between the signal extracted by the STRE algorithm and the signals $y^i_{r,0}[n]$ and $y^j_{r,0}[n]$ on a selected region near the chest of an infant breathing around 0.71 Hz is shown. The signal obtained with the STRE algorithm exhibits periodicity: albeit inhalation/exhalation movements are here partially distinguishable, for adult or different camera locations they may be not, creating signals which describe faster variations with a halved repetition time. Using the proposed algorithm, the inhalation/exhalation movements can be clearly distinguished as the two "motion" signals $y^i_{r,0}[nT_s]$ and $y^j_{r,0}[nT_s]$ exhibit movements with opposite direction, as highlighted by positive/negative values of the signals. This key difference can be noticed also in the magnitude spectra of the motion information signals, shown in Fig. 3(b). Although a peak around 0.75 Hz is clearly visible in all spectra, in the magnitude spectrum of the signal extracted by [3] (in which the mean is removed for better visualization), peaks corresponding to higher order harmonics near 1.4 Hz and 2.2 Hz are also included. These peaks, caused by the shape of the signal, may be higher than the fundamental one and, if exhibited also in other levels of the pyramidal decomposition, may cause an error in the estimation of the RR. The extraction of motion signals by the proposed algorithm solves this issue, as the obtained signals are quasi-sinusoidal and reduce the higher order harmonic peaks in the related spectrum, as shown in Fig. 3(b).

For performance evaluation, two video sets of recordings are considered: the first set consists of 4 videos of different adults on a chair and regularly breathing with an overall time duration of 20 min and 11 s; the second set consists of 2 videos of a newborn assisted in the Neonatal Intensive Care Unit (NICU) of the University Hospital of Parma,[2] Italy, with normal respiration and an overall time duration of 8 min and 16 s. The summary of parameters employed for the analysis of the two sets is reported in Table I. As a remark, the estimation of the RR is performed on temporal windows of duration

[2]Analysis and use of patients' biomedical signals and video recordings was approved by the Ethical Local Committee.

ISBN 978-90-827970-1-5 © EURASIP 2018

TABLE I
SUMMARY OF VIDEO TESTS AND PARAMETERS USED FOR THE ANALYSIS

Video set	No. samples	Camera resolution	f_s [Hz]	L	W [pixel]	R	$[f_{co,l}, f_{co,h}]$ [Hz]	α	NT_s [s]	Interlacing factor	Reference device
Adults	4	800×600	30	4	41	3	$[0.19, 0.9]$	20	20	50%	Accelerometer
Newborns	2	360×288	25	3	21	4	$[0.3, 1.1]$	25	20	50%	Pneumograph

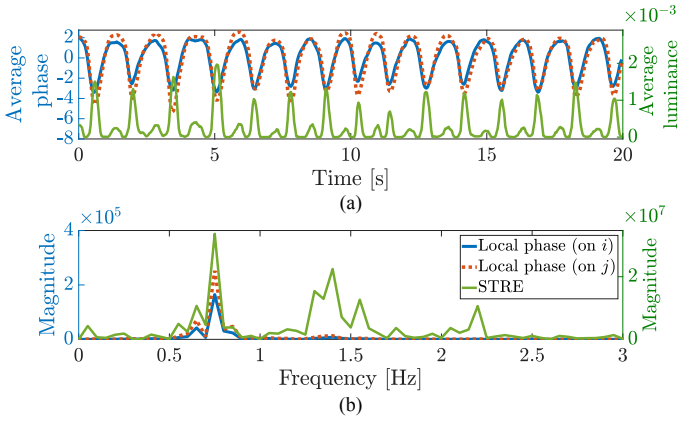

Fig. 3. Comparison (a) of signals $y^i_{r,0}[nT_s]$ and $y^j_{r,0}[nT_s]$ with one extracted by STRE and (b) corresponding magnitude frequency spectra.

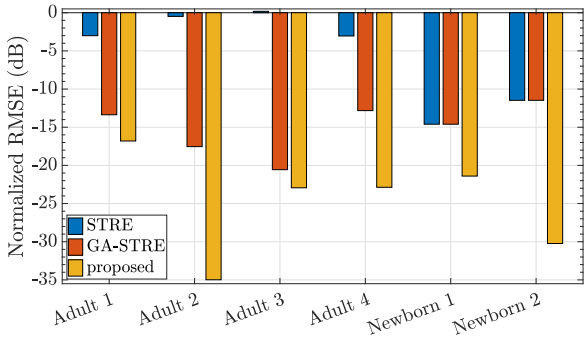

Fig. 4. Normalized RMSE in RR estimation for videos under test. GA version of STRE is employed in the case of adults to reduce the RR doubling effect.

NT_s s and a proper interlacing factor.

Performance in RR estimation is reported in Fig. 4, where for each video sample the Root Mean Squared Error (RMSE) between the estimated RR and the reference one, normalized to the Root Mean Square (RMS) value of the latter, is shown. It can be noticed that the method here proposed is more accurate in RR estimation in every test, in particular on adults: this is due to the fact that STRE yields wrong estimations on adults caused by the frequency doubling effect previously described. In order to compare the accuracy of both algorithms without accounting for this specific drawback, an idealized Genie-Aided (GA) version of STRE is also considered, in which RR estimates affected by frequency doubling are ideally corrected.

IV. Conclusion

In this paper, a video processing algorithm for non-contact monitoring of the RR, has been proposed. The method, once ROI have been selected, is able to process video streams in order to obtain signals describing breathing movements; this step is achieved relying on a recent phase-based motion magnification algorithm. Then, RR is estimated applying a signal fusion-based ML criterion, by automatically excluding unreliable ROI affected by large motions or variations which can hinder the RR estimation accuracy. The performance in breathing signal extraction and RR estimation has been analyzed by comparing with gold-standard devices and demonstrating a remarkable improvement over a previously proposed algorithm.

References

[1] F. Q. Al-Khalidi, R. Saatchi, D. Burke, H. Elphick, and S. Tan, "Respiration rate monitoring methods: A review," *Pediatr. Pulm.* (Wiley), vol. 46, no. 6, pp. 523–529, June 2011.

[2] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. W. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *IOP Physiol. Meas.*, vol. 35, no. 5, pp. 807–831, May 2014.

[3] D. Alinovi, L. Cattani, G. Ferrari, F. Pisani, and R. Raheli, "Spatio-temporal video processing for respiratory rate estimation," in *Proc. IEEE Int. Symp. Med. Meas. and Applicat.* (MeMeA), Turin, Italy, June 2015, pp. 12–17.

[4] R. Janssen, W. Wang, A. Moço, and G. de Haan, "Video-based respiration monitoring with automatic region of interest detection," *IOP Physiol. Meas.*, vol. 37, no. 1, pp. 100–114, Jan. 2016.

[5] D. Alinovi, G. Ferrari, F. Pisani, and R. Raheli, "Respiratory rate monitoring by maximum likelihood video processing," in *Proc. IEEE Int. Symp. Signal Process. and Inf. Technol.* (ISSPIT), Limassol, Cyprus, Dec. 2016, pp. 172–177.

[6] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, "Riesz pyramids for fast phase-based video magnification," in *Proc. IEEE Int. Conf. Comput. Photography* (ICCP), Santa Clara, CA, USA, May 2014, pp. 1–10.

[7] J. W. Woods, *Multidimensional Signal, Image, and Video Processing and Coding*, 2nd ed. Waltham, MA, USA: Academic Press, 2012.

[8] D. Alinovi and R. Raheli, "Extraction of periodic features from video signals," in *Proc. Ninth IARIA Int. Conf. Advances Multimedia* (MMEDIA), Venice, Italy, Apr. 2017, pp. 95–100.

[9] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.

[10] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd ed., ser. Prentice Hall Signal Processing. Upper Saddle River, NJ, USA: Pearson - Prentice Hall, 2010.

[11] M. Felsberg and G. Sommer, "The monogenic signal," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3136–3144, Dec. 2001.

[12] C. P. Bridge, "An introduction to the monogenic signal," *arXiv Computer Research Repository* (CoRR), pp. 1–21, Mar. 2017, available on arXiv.org [arXiv:1703.09199 (cs.CV)].

[13] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, "Quaternionic representation of the Riesz pyramid for video magnification," Computer Science and Artificial Intelligence Laboratory (MIT), Cambridge, MA, USA, Technical Report, 2014.

[14] ——, "Phase-based video motion processing," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 80:1–80:10, July 2013.

[15] L. Cattani, D. Alinovi, G. Ferrari, R. Raheli, E. Pavlidis, C. Spagnoli, and F. Pisani, "Monitoring infants by automatic video processing: a unified approach to motion analysis," *Comput. Biol. Med.* (Elsevier), vol. 80, pp. 158–165, Jan. 2016.

ISBN 978-90-827970-1-5