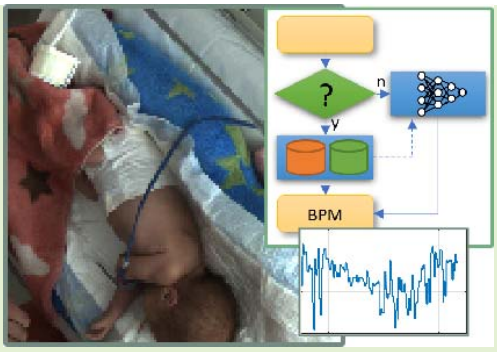
Reference Free Incremental Deep Learning Model Applied for Camera-Based Respiration Monitoring

Péter Földesy, Ákos Zarándy, *Member, IEEE*, and Miklós Szabó

***Abstract*—The article describes a reference and training set free incrementally trained deep learning algorithm for camera-based respiration monitoring systems. The algorithm uses a model based discriminator to find salient areas having respiration like periodic motion. It stores the first principle component of the found waveforms into two slowly growing set along with negative, uncorrelated motion patterns. Using these samples, it trains a deep neural network classifier incrementally to recognize respiration from sudden and motion intensive situations. The classifier had no forgetting mechanism and it is able to adapt quickly the changing respiration patterns and conditions. The algorithm has been validated in a total of 24 hours diverse recording captured in the neonatal intensive care unit (NICU) of the Ist Dept. of Pediatrics and, II. Dept. of Obstetrics and Gynecology, Semmelweis University, Budapest, Hungary and in the COHFACE publicly available dataset of adult subjects. The clinical data set evaluation resulted in mean absolute error (MAE) 6.9 and root mean squared error (RMSE) of 9.8 breaths per minute, respectively, the MAE was below 5 breaths per minute for over 50% of the time. The algorithm was assessed in the COHFACE dataset of adult subjects as well with respiration estimation MAE and RMSE values of 0.95 and 1.7 breaths per minute.**



***Index Terms*— Respiration monitoring, NICU, incubator, real-time, deep learning, incremental learning.**

I. Introduction

IN CAMERA based remote vital sign and specifically respiration monitoring algorithms several solutions have been published in the recent decade [1], also in the special field of newborn infant monitoring. Former methods used classical techniques for segmentation and pulse/breath rates extraction (e.g. blind source separation [2], [3], optical flow [4] for respiration-signal and Fourier analysis for rate value estimation [5], [6]). Recent studies relies on either mixed [7], [8] or neural network solutions [9]–[11] for these tasks to increase motion robustness and overall performance. Another interesting approach is to exploit local motion magnification algorithm [12], which in principle relies on regular, periodic motion enhancement and thus the very irregular respiration of the newborn infants is hard to handle with. The estimation of respiration by following the inspiratory and expiratory movement of the chest and abdominal region is also a common approach. In [13] a median optical flow signal is used to gather respiratory signal after face and chest detection. An example for fusing different techniques is presented in [14] for respiration monitoring. The computationally intensive algorithm performs skin segmentation, which is achieved using a multi-task convolutional neural network, than seeks the similarity between the amplitude modulation by both classical PCA and ICA methods. The involved network was trained by manually annotated skin regions and reached mean absolute error (MAE) 6.9-7.5 breaths per seconds in the collected data set including motion active regions as well. The [15] and [16] demonstrate a complex monitoring system based on a multiple output convolutional neural network. The solution gives a higher level detection capability such as intervention periods. The presented results show low error rate (3.5-4.5 BPM MAE over selected periods) however the data preparation, annotation, augmentation efforts is significant and specific to the observed environment, and the employed neural network requires heavy GPU acceleration.

Manuscript received July 3, 2020; revised August 28, 2020; accepted August 29, 2020. Date of publication September 2, 2020; date of current version December 16, 2020. This work was supported by the National Research, Development and Innovation Office, Hungary, under Project VEKOP-2.2.1-16-2017-00002. The associate editor coordinating the review of this article and approving it for publication was Prof. Guiyun Tian. *(Corresponding author: Péter Földesy.)*

Péter Földesy and Ákos Zarándy are with the Institute for Computer Science and Control, H-1111 Budapest, Hungary (e-mail: foldesy.peter@sztaki.hu).

Miklós Szabó is with the First Department of Paediatrics, Semmelweis University, H-1085 Budapest, Hungary.

Digital Object Identifier 10.1109/JSEN.2020.3021337

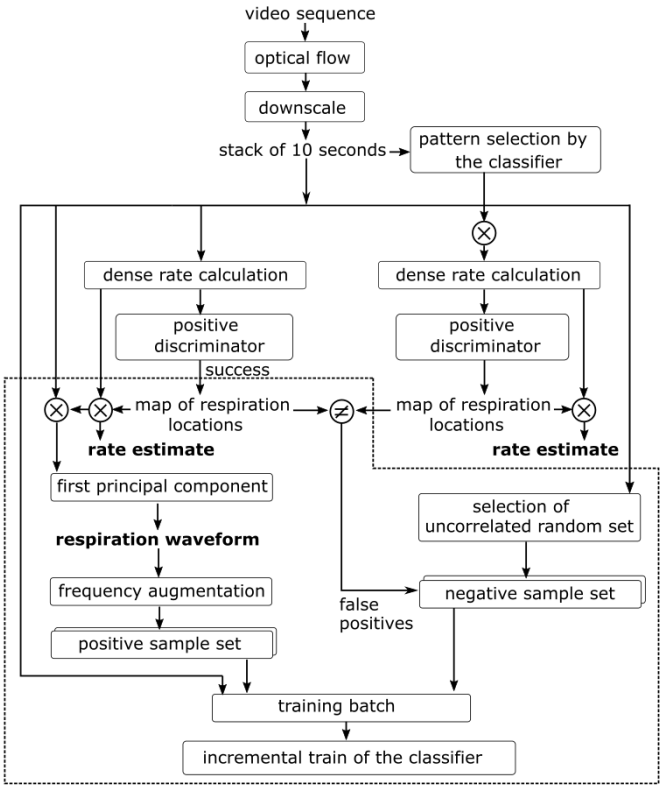


Fig. 1. High level flow chart of the presented algorithm.

Our algorithm is a combination of optical flow based source separation and a low complexity convolutional neural network classifier. The main contribution is the addition of enabling self learning. It takes advantage of the classical approach of optical flow processing to the select region of interest whenever possible, i.e. collects autonomously the training set, which is used to train incrementally the neural network classifier to handle ambiguous situations, when the classical solution fails. The solution has lower computational effort than the full convolutional networks require, and makes any manual intervention or annotation unnecessary.

II. Operation Principles

The architecture is a combination of a dense rate calculation source separation model and a neural network classifier. It can be divided into three major components: a positive discriminator of source separation to select respiration patterns, a negative sample selection mechanism, and a neural network classifier. The discriminator is used to handle motion artifact free situations and provides training set for the classifier, which is use to find these patterns in motion corrupted periods.

The assumption here is that the subject is cooperative, thus non respiratory movement may be present, but not dominant for short time periods. In order to determine whether a salient area can be found or not, the algorithm uses a discriminator.

The flow chart of the algorithm is presented in Fig. 1. It first collects a stack of optical flow amplitude and from this stack, creates a dense respiration rate map. This histogram of the rate map is then analyzed using a simple set of rules. The rules constrains the signal amplitude, frequency band, spatial coherence. The pixel-wise optical flow waveforms resulting a valid respiration detection, is called hereby positive samples, and motion affected, uncorrelated waveform patterns as negative samples. If a significant signal is found, the corresponding waveforms are combined by calculating their first principal component and stored, and the calculated rate is displayed as final result. Besides, portion of the negative samples are also picked and stored. In addition, the neural network classifier is trained incrementally with low iteration count using the stored positive and negative samples.

In time intervals, when there is no dominant periodic motion found, a neural network classifier is used to find areas and patterns that resemble the trained respiration patterns. The candidate waveforms are validated by the same selection rules. In the presence of valid waveforms and rate, they are displayed as valid results.

As the time elapses, the classifier performance becomes better to find weak signal sources and due to the stored sets, long term memorization is preserved and no overfitting occurs due to the increasing and versatile training set.

III. Dense Rate Map Analysis

This section describes the details of how the dense respiration rate map and its histogram and the positive and negative samples are generated.

A. Dense Rate Map

The rate calculation is based on optical flow calculation. First, a dense optical flow is calculated using the Horn-Schunck algorithm [17]. The output of the optical flow is downscaled to a lower resolution, which resolution depends on the viewing angle of the camera, distance of the targets, generally only a low resolution image is necessary (reasoning presented in the end of this section). The downscaled images are collected in a stack, on which the pixel-wise rates are calculated.

The rate of a waveform can be estimated multiple ways. There are a few common solutions: Fourier analysis and peak search, peak and zero crossing counting of the band pass filtered signal [18]. The frequency domain and linear band pass filtering rate estimation is not adequate solution for the very irregularly and broken expiratory patterns of the newborn infants [19]. Hence, we propose the rate estimation as follows: first, linear detrending is done of the raw waveforms, than nonlinear noise filtering, finally, crossing point counting at a threshold level above the noise level. The nonlinear noise filtering removes the high frequency components by first an exponential smoothing of 0.5 filter factor, and a Savitzky-Golay (polynomial) smoothing filter [20]. This filter has 2^{th} order and its window size is double of the shortest expected inhale/exhale peak (i.e. 0.6 seconds for newborn infants). The reason for the choice of nonlinear filtering is to preserve the shape of the waveform without inserting false peaks, which the linear bandpass filter that can cause with overshooting and ringing (see Fig. 2). Threshold level crossing points are finally counted and the rate is extrapolated to 1 minute period.

The evaluations had shown that the rate map size affects the performance of the algorithm in two ways. First, the too dense

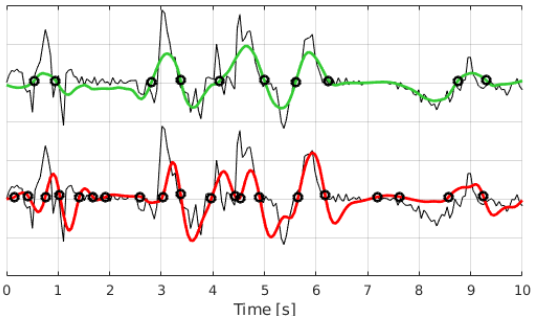


Fig. 2. The rate calculation method. The top, green curve is the detrended, nonlinear filtered optical flow signal with threshold crossing points. The bottom, red curve is the linear bandpass filtered (20-120 BPM, 6th-order Butterworth filter) signal with the crossing points showing several false hits due to ringing and overshoots.

map, such as native resolution of the input frames, results in noisy rate estimation and thus makes the histogram peaks blurred. Secondly, the very low resolution rate map makes the rate histogram too discrete to analyze it meaningfully and provides inadequate samples for the training process. Between the two limiting cases, the rate estimation performance did not show significant variation. In practice, the sufficient resolution is reached, when the number of rate estimated pixels of the subjects manifesting respiration pattern exceeds approximately 50-100. Applying this simple rule of thumb to the presented NICU and COHFACE experiments, considering the field of view (see Fig. 7), at least 48×48 and 32×24 pixels rate map resolutions are required and thus used.

B. Positive Sample Discriminator

The task of the positive discriminator is to find a single dominant source that is reached by analysing the rate distribution across the image. The rate values of all waveforms of the stack are determined providing a dense rate map and a rate histogram constructed from its values using 2 BPM bin size. A sample histogram can be seen in Fig. 5a.

The analysis of the histogram is done by a rule set, which is a composition of simple algebraic relations. The detailed selection rules are the followings:

1) Noise Level Constraint: The waveforms having standard deviance below a limit (noise level) are removed from further processing. The noise level is calculated on empty view recordings and consequent optical flow, downscaling and filtering steps. Its value is fixed before on-site execution.

2) Single Dominant Peak: The largest peak is greater by 10-30% than the second largest bin in the histogram. Adjacent peaks are handled as a single dominant peak.

3) Valid Frequency Range Constraint: Outlier frequency valued waveforms are removed (e.g. smaller than 10 breaths per minute (BPM) and larger than 90 BPM for newborn infants, 5-80 BPM for adults, etc.).

4) Area Constraint: The height of the highest peak must be larger than 5-10% of number of the all pixels.

5) Geometrical Outlier Rejection: Though the pattern classification is done independently for the waveforms in each pixel, for spatial filtering, it is remaped back to image form. In this form, the individual, separated points are removed by

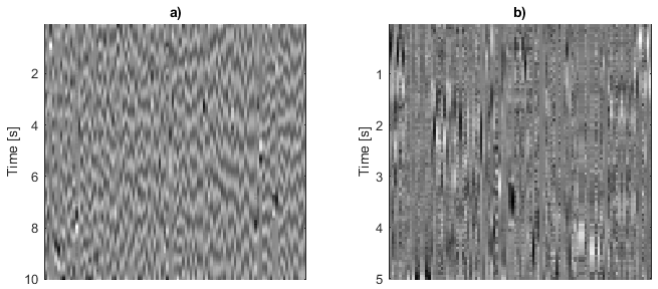


Fig. 3. Examples of a stored data sets: a) positive samples set, b) negative samples set.

binary morphological operation enforcing spatially cohesive distribution (see Fig. 6c). The operation is a short sequence of binary erosion of 2×2 kernel and a reconstruction step.

6) Correlation Constraint: The first principal component is calculated of the remaining waveforms. Those waveforms, which have low component scores (score is defined here as the dot product of the principal component vector and a waveform), are removed from the positive hits.

If the above process resulted in valid outcome, the first principal component, i.e. respiration waveform, is frequency augmented and multiplied by $+1/-1$ and stored in the positive data set. The peak rate is displayed as the estimated rate. The data augmentation is crucial for improving the generalization capability of the classifier. The augmentation contains new frequency augmented waveforms generated by oversampling resulting in $5, 10, 15\%$ rate reduced versions (e.g. 80 BPM rated waveform is oversampled in the time dimension, the data that exceeds the original length is dropped, and the result is 76, 72, and 68 BPM signals). Overall, a positive outcome generates eight new waveforms in the dataset.

C. Negative Sample Selection

The role of negative sample selection is to build up a set of samples, where possibly no periodic, respiration like patterns are present. The classifier is also inferred on the same stack content beside the positive discriminator (Fig. 1), and waveforms from the false positive detection (difference between the discriminator and the classifier output) are selected as new patterns to be included in the negative set. If no false positives found or there was no positive peak, the non-peak region of the image is used for random selection. The picked candidate waveforms are checked not to be correlated by Pearson correlation: if a newly selected waveform is correlated $r > 0.4$ with an existing selection from the same stack, a new one is picked and checked. This way the diversity of the samples is increased and the risk to include unwanted periodic cases is minimized, and the contrastive capability of the classifier is increased as well. The amount of negative samples is maintained about the twice of the positive samples. A sample of such sets after continuous operation can be see in in Fig. 3.

IV. Classifier

The goal of this classifier is to distinguish the positive, respiration alike, waveforms from the negative samples, so it

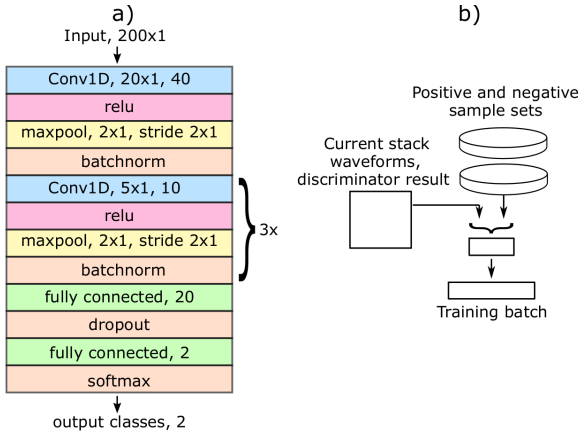


Fig. 4. a) network architecture of the waveform classifier. b) composition of the training batch for one incremental learning step.

is a two-class classification problem on wave patterns. The selected solution is a binary output convolutional neural network. Its input size is 200×1, it is composed of z-score input normalization (scaling with the average and standard deviation of one), 1D convolutions, batchnorm, relu activation function, maxpooling, and a two fully connected layers before its output layer. The architecture is illustrated in Fig. 4a. The architecture was selected and tuned on multiple runs on the recorded video datasets.

A. Incremental Training

The incremental training is a key feature of the concept. No pretraining nor recorded annotated full training set is required, instead it is collected during operation. The constant learning rate training with a moving window [21] or with forgetting mechanisms [22] are well known techniques. The disadvantage of these methods could be that the size of the window and forgetting factor is critical and an non optimal choice could result in instability or undertraining. Another broadly accepted method is to increase the network complexity with structural adaptation [23].

The hereby chosen method is batch based training with limited number of epochs per steps. Though the batch learning paradigm has its limit compared to classic incremental or online learning, namely it needs to handle the whole dataset for training, the proposed method do not results in unmanageable data amount. Therefore, the incremental training (weight update) is performed with batches and small number of epochs. The training batches are composed of the collected positive and negative sets that are concatenated with the waveforms of the current time stack (see Fig. 4b). As the time elapses, the positive and negative sets are increasing with a moderate and decaying rate. E.g. only a few hundreds of new uncorrelated waveforms per hour was collected in our clinical test series. The consequence of using the whole training set in each weight update is that there is no need for forgetting mechanism and the training window is the gathered complete dataset. The observation is that the generalization and stability of the performance is better if the training is performed only when respiration signal was found by the discriminator. This way, the batches contained clear input samples in addition to the gathered sample sets. Practical details are disclosed in the following subsections.

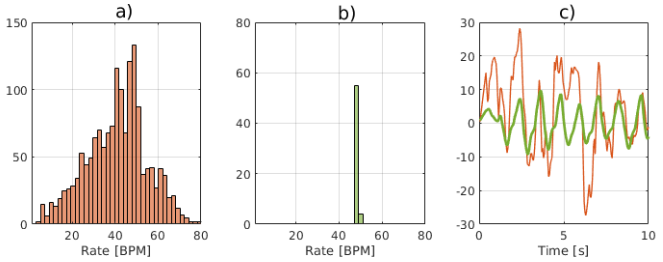


Fig. 5. a) A sample rate histogram of a 10 seconds period during running the algorithm, b) result of the classifier combined with the discriminator, c) red curve shows the associated waveforms: unprocessed average of the optical flow amplitude of the whole period and the green curve shows the resulting extracted waveform.

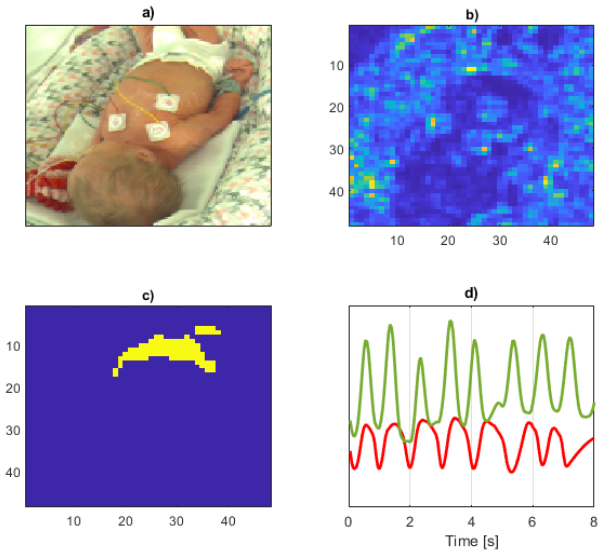


Fig. 6. (a) sample snapshot, at the moment, when baby moves, (b) the optical flow amplitude, (c) the classifier output can be seen showing the areas of the detected respiration like locations at the abdominal, rejecting the upper body movement. The (d) figure compares the ECG contact impedance pneumographic signal (red) and this algorithm generated results (green).

B. Inference

When the discriminator fails to find a significant peak, the output of the classifier is used for select respiration patterns. The candidate waveform set, using the corresponding rate histogram, are validated and further filtered by the positive discriminator. The rates of the approved waveforms are averaged to provide the final rate estimation (see Fig. 5). An example of an input snapshot and the results are shown in Fig. 6.

C. Execution Time and Parameter Optimization

The execution time of algorithm, including both training and processing, enabled real-time operation on an i3 powered computer with no GPU acceleration and using a single execution thread and no optimized code. In order to control and limit the execution time of the training steps, only a fix length random permutation of the sample sets are included after their size exceeded a system dependent limit. Also, at regular intervals,

the two data sets are purged by removing highly correlated samples. In the experiments, training set size of 2000-5000, batch size of 100-500 samples, 1-10 epochs, constant and exponentially decreased learning rate, and Adam/Stochastic Gradient Descent with Momentum (SGDM) weight update methods was evaluated and compared to find optimal settings. The question of the investigation was how the performance changes with this technique compared to full length training. It turned out, that the precision of the classifier slightly depends on the size of the sample sets and on the batch size, learning rate of the training steps, but in general, the performances are very close and usually slightly better. The presented experimental results has been reached with batch size of 200 and two epoch per training steps, using SGDM update, constant learning 0.001 rate, and 4096 training set size limit.

D. Performance Stabilization Time

The question arose, that how long is the adaptation time of the algorithm. This time strongly depends on the behavior of the subjects. In case of newborn infants, the different sleeping stages and awake periods follow each other with a few hours periodicity. Adults change their respiration patterns as well on daily bases and depending on their activity. The algorithm training process contains the latest respiration waveforms overrepresented, consequently it immediately follows the respiration pattern changes and capable of recognizing the new pattern robustly afterwards. The long term pattern changes learned as they first occur. In the experiments, infant monitoring stabilized after 4-6 hours, while in case of the COHFACE database, it took less than an hour of recordings. This period was estimated by disabling the training process and letting the inference alone to determine the respiration rate.

V. Experimental Results

Though the algorithm and its training do not rely on external reference signals, comparisons are presented of its results using a clinical trial and a publicly available dataset.

A. NICU Experiments

Recordings with reference data were collected in the NICU of the Ist Dept. of Pediatrics and, II. Dept. of Obstetrics and Gynecology, Semmelweis University, Budapest, Hungary. The population demographics of the participants can be seen in Table I.

The camera was a Basler acA2040-55uc model. The videos had 500 × 500 pixels resolution and 20 frames per second speed. Synchronized reference data for respiration had been gathered from vital-sign monitoring systems, namely Philips IntelliVue MP20/MP50 models. The videos were captured from different camera-angles and with different optics under ambient illumination (see Fig. 7).

The experiment was performed in two steps. First, the algorithm had run on a 12 hours balanced set of the different scenarios including different infants, camera-angles, illumination changes, periods of intensive motion, caring, and phototerapy

TABLE I
Population Demographics

Subject	1	2	3	4	5	6	7
Recording time (hours)	96.7	5.5	39.4	27.4	51	105.5	50.1
Gender	F	M	M	F	F	M	F
Gestatioanal age (weeks)	32	32+3	31+4	35+4	39	32	33
Birth weight (g)	2020	1840	1850	1870	3150	2120	2080
Postnatal age (days)	4	4	10	8	4	7	2
Actual weight (g)	1900	1850	1680	1820	2905	2040	1960
Length (cm)	46	44	-	45	57	45	44
Head circumference (cm)	32	29.5	-	32	34	30	32
Respiratory support	no	no	no	no	no	no	no
Pharmacological cardiovascular support	no	no	no	no	no	no	no
Any drugs a)	no	no	no	no	yes	no	no
Fitzpatrick scale	2	3	2	2	2	2	2

a) Drugs may alter muscle tone / physical activity (sedatives, anticonvulsants etc.).

Fig. 7. Snapshots of the clinical video set.

device lights collected from the 375 hours database. In the second step, another independent 12 hours test is composed to compare the rate estimate to a reference monitor. The second run was a continuation of the first one by using and expanding the collected sets and preserving the classifier weights. The reason for the two test phases is that the motion activity and respiration pattern varies in a broad range of the infants, and the necessary samples collected by the algorithm helps in analyzing new patterns. This way the long term, continuous operation performance is assessed.

In the test collection, the infants slept for about 20% of the time, otherwise had shown usual daylight motion activity. The span of the respiration rate was 10-90 BPM with 46 BPM mean value.

The time window of calculation was 10 seconds, which is moved in 5 seconds steps. The size of the dense rate map was 48 × 48 pixels downscaled from the optical flow including the whole camera view. In 18% of the recordings, no respiration pattern was found due to intensive motion, caring, body twisting or rotation. The prediction over positive peak found ratio was above 28%. The RMSE of rate estimation for the test set was 9.8 BPM, and the MAE 6.9 BPM with −1.55 BPM bias. The MAE was below 5 BPM for over 50% of observation time. The Bland-Altman analysis showed −21.2 and 17.1 BPM at 95% limits of agreement. Samples of rates comparison can be seen in Fig. 8. The Bland-Altman plot is presented in Fig. 9.

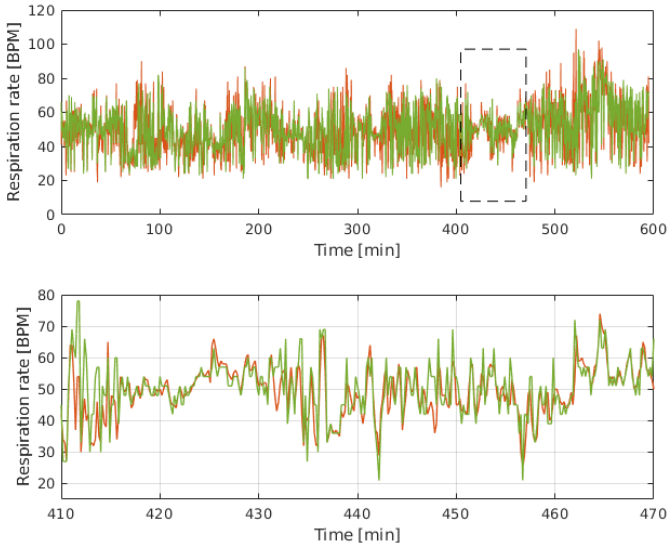


Fig. 8. Rate estimation (green) compared to the respiration rate signal of the reference Philips IntelliVue MP50 monitor (red). The top curve is a 10 hours cut, the bottom curve is a part of it shown as a dashed rectangle in the top figure.

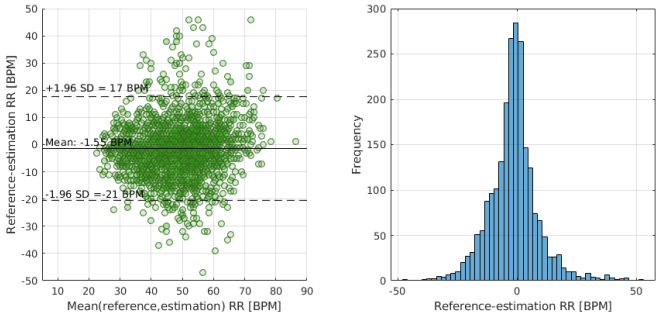


Fig. 9. Bland-Altman plot and the histogram of the rate estimation and the reference Philips IntelliVue MP20/MP50 monitor respiration rates.

The results are comparable to the findings of recent and relevant publications, such as [9] presenting the limits of agreement were −22 to 23.6 BPM on a smaller database. Two methods (PCA and ICA blind separation techniques) have been disclosed in [14] based on a similar database having MAE 6.9 BMP and 7.5 BPM values. The [16] describes an implementation of video-based non-contact technologiy to monitor the vital signs of preterm infants. The used dataset is composed of several patients and long recording time. The corresponding respiration rate estimate had −15.1 to 10.9 limits of agreement and 3.5-4.5 BPM MAE after discarding noisy periods (33.3% of the complete recording time was considered valid).

B. COHFACE Dataset Evaluation

The publicly available COHFACE dataset is evaluated with the algorithm as well [24] with adjusted model parameters. Beside video recordings, breathing waveform, recorded by respiration belt, have also been recorded in the dataset. The dataset is meant for remote heart-rate monitoring, a part of the chest and shoulder is visible of most of the subjects allowing the video based respiration rate measurement. The set contains 160 one-minute long RGB video sequences of 40 adult subjects recorded in a real-world scenario (12 females and

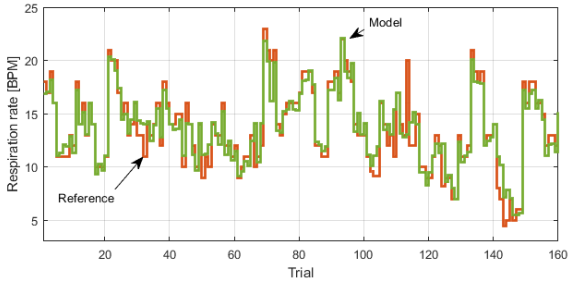


Fig. 10. The COHFACE dataset test results (green) compared to respiration belt based respiration rate (red).

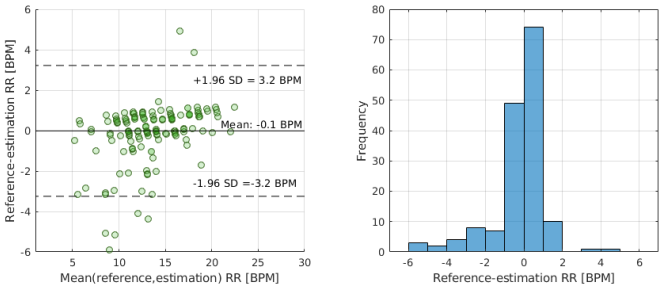


Fig. 11. Bland-Altman plot and the histogram of the rate estimation error for the COHFACE dataset.

28 males, in age 20-65 years). The video sequences have a resolution of 640 × 480 pixels and the frame rate of 20 Hz. The respiration rate varied between 5-25 BPM across the set.

The following model parameters have been modified to fit the adult subjects: the frame rate of the original video has been dropped to fit one minute window to the input width of the classifier (3.3 frames per seconds); the rate map resolution was 32 × 24 pixels; the noise level and the crossing point thresholds were 0.005 and 0.002, respectively; the Savitzky-Golay filter order had 2^{nd} order and 2.5 seconds window size; the histogram bin size was 1 BPM; the constraints of the positive discriminator were: single dominant peak parameter was 5%; valid range was set to 4-40 BPM; area constraint was 5%; and the PCA score limit was 0.2. The negative sample selection process remained unchanged.

The algorithm has been executed twice on the dataset. In the first run, the first half of the one-minute videos has been involved in order to let the incremental classifier to build its representation of the observed respiration behavior. In the second run, the second half of the videos are processed. 10 trials was classified not to have respiration like pattern (not visible or out of range). The rate estimate error values were RMSE = 1.7 BPM, MAE = 0.95 BPM with −0.1 BPM bias, and the limits of agreement were −3.2 to 3.2 BPM (as shown in Fig. 10-11). This error rate is only slightly worse compared to a sophisticated respiration model based solution of [4] providing limits of agreement of −2.67 to 2.81 BPM in a similar dataset.

VI. Conclusion

A reference and training set free deep learning model has been described. A simple rule set applied by a discriminator on dense rate histograms helps to incrementally train a classifier that is capable to identify respiratory movements in motion

corrupted videos. The transformation from the waveforms to the scalars, i.e. rate calculation, is defined in general way and can be easily reformulated to highlight different properties. By changing it, other multichannel measurement sources can be trained online and analyzed with the algorithm, whenever there is no training set or difficult to gather such.

ACKNOWLEDGMENT

The authors would like to thank to Ádám Nagy, Dániel Terbe, Máté Siket, Imre Jánoki for collecting and processing NICU measurement data. Portions of the research in this article used the COHFACE Dataset made available by the Idiap Research Institute, Martigny, Switzerland.

REFERENCES

[1] C. H. Antink, S. Lyra, M. Paul, X. Yu, and S. Leonhardt, "A broader look: Camera-based vital sign estimation across the spectrum," *Yearbook Med. Informat.*, vol. 28, no. 1, pp. 102–114, Aug. 2019.

[2] M. Lewandowska, J. Rumiński, T. Kocejko, and J. Nowak, "Measuring pulse rate with a webcam—A non-contact method for evaluating cardiac activity," in *Proc. Federated Conf. Comput. Sci. Inf. Syst. (FedCSIS)*, Sep. 2011, pp. 405–410.

[3] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in non-contact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.

[4] R. Janssen, W. Wang, A. Moço, and G. de Haan, "Video-based respiration monitoring with automatic region of interest detection," *Physiolog. Meas.*, vol. 37, no. 1, p. 100, 2015.

[5] P. De Chazal *et al.*, "Sleep/wake measurement using a non-contact biomotion sensor," *J. Sleep Res.*, vol. 20, no. 2, pp. 356–366, Jun. 2011.

[6] W. Wang, A. C. den Brinker, and G. De Haan, "Full video pulse extraction," *Biomed. Opt. Express*, vol. 9, no. 8, pp. 3898–3914, 2018.

[7] X. Niu, H. Han, S. Shan, and X. Chen, "SynRhythm: Learning a deep heart rate estimator from general to specific," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 3580–3585.

[8] R. Song, S. Zhang, C. Li, Y. Zhang, J. Cheng, and X. Chen, "Heart rate estimation from facial videos using a spatiotemporal representation with convolutional neural networks," *IEEE Trans. Instrum. Meas.*, early access, Mar. 20, 2020, doi: 10.1109/TIM.2020.2984168.

[9] K. Gibson *et al.*, "Non-contact heart and respiratory rate monitoring of preterm infants based on a computer vision system: A method comparison study," *Pediatric Res.*, vol. 86, no. 6, pp. 738–741, Dec. 2019.

[10] Z. Yu, X. Li, and G. Zhao, "Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks," in *Proc. BMVC*, 2019, pp. 1–12.

[11] R. Špetlík, V. Franc, and J. Matas, "Visual heart rate estimation with convolutional neural network," in *Proc. Brit. Mach. Vis. Conf.*, Newcastle, U.K., 2018, pp. 3–6.

[12] D. Alinovi, G. Ferrari, F. Pisani, and R. Raheli, "Respiratory rate monitoring by video processing using local motion magnification," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 1780–1784.

[13] K.-Y. Lin, D.-Y. Chen, and W.-J. Tsai, "Image-based motion-tolerant remote respiratory rate evaluation," *IEEE Sensors J.*, vol. 16, no. 9, pp. 3263–3271, May 2016.

[14] J. Jorge, M. Villarroel, S. Chaichulee, K. McCormick, and L. Tarassenko, "Data fusion for improved camera-based detection of respiration in neonates," *Proc. SPIE*, vol. 10501, Feb. 2018, Art. no. 1050112.

[15] S. Chaichulee *et al.*, "Cardio-respiratory signal extraction from video camera data for continuous non-contact vital sign monitoring using deep learning," *Physiolog. Meas.*, vol. 40, no. 11, Dec. 2019, Art. no. 115001.

[16] M. Villarroel *et al.*, "Non-contact physiological monitoring of preterm infants in the neonatal intensive care unit," *NPJ Digit. Med.*, vol. 2, no. 1, pp. 1–18, Dec. 2019.

[17] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, nos. 1–3, pp. 185–203, Aug. 1981.

[18] P. H. Charlton, M. Villarroel, and F. Salguiero, "Waveform analysis to estimate respiratory rate," in *Secondary Analysis of Electronic Health Records*. Cham, Switzerland: Springer, 2016, pp. 377–390, doi: 10.1007/978-3-319-43742-2_26.

[19] A. B. te Pas, C. Wong, C. O. F. Kamlin, J. A. Dawson, C. J. Morley, and P. G. Davis, "Breathing patterns in preterm and term infants immediately after birth," *Pediatric Res.*, vol. 65, no. 3, pp. 352–356, Mar. 2009.

[20] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures.," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964.

[21] G. Widmer and M. Kubat, "Learning in the presence of concept drift and hidden contexts," *Mach. Learn.*, vol. 23, no. 1, pp. 69–101, Apr. 1996.

[22] D. Martínez-Rego, B. Pérez-Sánchez, O. Fontenla-Romero, and A. Alonso-Betanzos, "A robust incremental learning method for non-stationary environments," *Neurocomputing*, vol. 74, no. 11, pp. 1800–1808, May 2011.

[23] B. Pérez-Sánchez, O. Fontenla-Romero, and B. Guijarro-Berdiñas, "A review of adaptive online learning for artificial neural networks," *Artif. Intell. Rev.*, vol. 49, no. 2, pp. 281–299, Feb. 2018.

[24] G. Heusch, A. Anjos, and S. Marcel, "A reproducible study on remote heart rate measurement," 2017, *arXiv:1709.00962*. [Online]. Available: http://arxiv.org/abs/1709.00962

Péter Földesy received the Ph.D. degree in integrated circuit design for early vision chips from the Budapest University of Technology and Economics, Hungary, in 2002, and the D.Sc. degree in sub-THz and mmwave sensory technology contribution from the Hungarian Academy of Sciences, Hungary, in 2019. Since 1996, he has been with the Institute for Computer Science and Control (SZTAKI), Budapest, Hungary, where he is currently a Research Fellow. Since 2007, he has also been with the Péter Pázmány Catholic University, Faculty of Information Technology and Bionics. His current research interest includes the application of life sign monitoring with non-contact technologies of newborn and preterm infants.

Ákos Zarándy (Member, IEEE) received the Ph.D. and D.Sc. degrees in electrical engineering and computer science from the Hungarian Academy of Sciences in 1997 and 2010, respectively. He is currently a Research Advisor with the Institute for Computer Science and Control (SZTAKI), Budapest, Hungary, and also a Professor with Péter Pázmány Catholic University. His research interests include computer vision and image sensing with special sensors and optics. He led several successful research and development projects, including medical vision system development, locally adaptive sensor development, and solved ultrahigh-speed vision problems. He is also active in remote and video-based photoplethysmography.

Miklós Szabó received the Med Habil degree in neonatology from Semmelweis University, Budapest, Hungary, in 1988, the M.D. degree, and the Ph.D. degree in extracellular antioxidant defence mechanisms of neonatal adaptation in 2005. He was a Neonatologist with Semmelweis University, where he has been with the First Department of Paediatrics since 1993. He is specialized paediatrics and later neonatology. Since 2007, he has been the Head of Neonatal Services and the Chief of Neonatal Intensive Care. Since 2017, he has also been a Lecturer of Semmelweis University and has also been the Head of the Division of Neonatology since 2019.

He initiated the first Hungarian web-based health database on neonatal health in 2002 and was pioneering the hypothermia treatment as neuroprotective therapy in asphyxiated infants in 2004. He has authored several peer reviewed publication on clinical research regarding neonatal hypoxic ischemic encephalopathy. He is an acknowledged physician and a clinical researcher. His scientific activity earlier focused to neonatal antioxidant defense mechanisms and later to neonatal neuroprotection, quality of the care and to humanize the early neonatal period.