Video-based Respiration Monitoring of Newborn Babies: a Feasibility Study

G. Morse[1] and D. Chetverikov[2,3], and D. Egyed[1]

[1] student, Eötvös Loránd University, Budapest, Hungary
[2] Institute for Computer Science and Control, Budapest, Hungary
[3] Eötvös Loránd University, Budapest, Hungary

Abstract
Newborn babies, especially premature ones, need real-time monitoring of vital parameters such as heart and respiratory rates and oxygen saturation. Currently, premature infants kept in clinical incubators are continuously monitored by contact sensors that provide signals for measuring the parameters. The contact sensors have certain disadvantages including the inconvenience caused to babies and the stress induced in parents. In this paper, we address the problem of video-based respiratory rate (RR) monitoring and present the initial related results of an ongoing project aimed at non-contact monitoring of the vital parameters. Different methods for RR estimation are proposed and discussed.

1. Introduction

Infants born before the 37th week of pregnancy are called *premature*. In Hungary, they form about 8–9% of newborn babies [6]; this rate is typical for European countries, e.g., Switzerland.

Potential brain lesions of premature babies frequently (25%) lead to long-term disabilities. To avoid this, the infants are kept in clinical incubators where they receive the necessary treatment under continuous monitoring of their vital parameters including heart rate, respiratory rate (RR), and oxygen saturation (blood oxygen level).

The traditional, wide-spread monitoring systems are based on electroencephalogram (EEG) and polysomnographic devices [5]. The vital parameters are measured based on the signals provided by wired sensors attached to infant body.

The current monitoring systems are expensive and time-consuming [5]. They can suffer from motion artifacts and relatively high false alarm rates. The contact sensors, although moderately invasive, are still inconvenient for babies and induce stress in parents.

Recently, a national research and development project has been started whose aim is to substitute the contact sensing by non-contact video-based one. (See Acknowledgments for the project data.) The project involves work in three major areas to monitor the three vital parameters mentioned above.

This paper is devoted to methods for video-based measurement of *respiratory rate* defined as the number of breath cycles per minute [19]. For information on breath monitoring with other smart sensors, the reader is referred to [3, 8].

For completeness, let us mention some relevant literature on the other two areas of the project. A survey on non-contact heart rate monitoring is available in [11]. Other studies of non-contact heart rate monitoring can be found in [2, 13–15]. Finally, the non-contact monitoring of oxygen saturation is discussed in [2, 7, 10, 18].

The rest of this paper is organized as follow. Section 2 provides a discussion of previous work on video-based respiration monitoring. In section 3, we discuss the clinical setup and the challenges of breath monitoring. Section 4 presents the methods for RR estimation we considered in the initial phase of the project. Initial test results are shown in section 5 to demonstrate the feasibility of the methods. Finally, conclusions and future work are presented in section 6.

2. Previous Work on Respiration Monitoring

Respiratory patterns and rates of adults differ from those of newborn babies. Respiratory signals measured in (healthy) adults are regular, close to periodic, with typical rates of 12–18 breaths per minute [19]. The signals measured in newborn babies are much less regular; they contain frequent pauses and variations, with typical respiratory rates of 30–40 breaths per minute. Another important difference is that the respiratory motion of infant's body is subtle and less visible, which makes its detection and tracking more difficult than in case of adults.

Altogether, this means that video-based respiration monitoring methods developed for adults are not necessarily applicable to newborn babies. In particular, signal frequency analysis approaches such as Fourier Transform may fail in case of infants because of the lack of a definite dominant frequency corresponding to the respiratory rate. This low (quasi-) periodicity of infant's respiratory signals necessitates counting each single breath event separately.

The early survey [1] of respiration monitoring methods reviews both contact and non-contact approaches, with more attention being paid to the former ones. Cases of monitoring children and adult patients are discussed. For non-contact methods, the imaging devices considered are video cameras and thermal cameras. Simple chest motion detection and measurement algorithms such as interframe difference are mentioned. These algorithms are not sufficiently robust to operate in real clinical environment.

In recent years, the persistent development of better and cheaper cameras as well as powerful portable computers and graphics cards created new opportunities for more sophisticated, real-time video processing.

A telemedicine-oriented system for respiratory rate measurement of adult patients is proposed in [21]. The patient sits in front of a near-IR camera sensitive to light in the visible and near-infrared domain. The upper body (chest/abdomen) of the patient is selected as the region of interest (ROI). Brightness variations in the ROI are used as temporal indicators of respiration events. Fast Fourier Transform is applied to obtain the respiratory rate of the patient. Although the system [21] is not suitable for newborn monitoring, the relation between the brightness variations and the respiratory rate deserves attention as a potentially useful indicator.

The study [12] addresses the problem of feature point selection and tracking for respiration monitoring, primarily in telemedicine. A smart phone camera is used whose frame size is reduced to 1080×720 pixels for faster operation. The distortion caused by camera motion is compensated by mapping each frame to the initial one. The original single-level implementation of the Kanade-Lucas-Tomasi (KLT) feature point tracker [16] is applied, and large persistent motions are selected as indicators of breath. The tests in [12] only check the breath feature point selection against manually provided ground truth. No respiratory rate is calculated. However, the idea of applying the efficient KLT tracker is useful, and we also use this tracker in our study.

An initial study of video-based monitoring of breath for adult mobile health control is presented in [15]. The setup of this study resembles that of [21]: the patient sits in front of the camera close to it. ROIs for breath monitoring are parts of the shoulders near the neck. The experimental study only demonstrates correlation between the respiration and the motion of the ROIs tracked in a short video, without calculation of the respiratory rate.

The recent study [5] aims at monitoring newborn babies to detect neonatal clonic seizure (sustained rhythmical jerking) from limbs' motion and apnea (suspension of breathing) from chest motion. The well-known approach [17] is used for motion area extraction. The limbs are monitored at the original video resolution to detect periodic motion that indicates clonic seizures. As the chest motion due to breath is small and subtle, the authors use the method [20] to magnify the chest video, then extract motion in the magnified video. The lack of chest motion for a certain period is interpreted as the indicator of apnea. No respiratory rate estimation is provided, and it is not clear if the obtained chest motion data can be used for this purpose.

Note that the method [20] relies on an optical flow algorithm closely related to [16]. In [18], the method [20] is used to enhance the visibility of face details for pulse visualization and counting. In our future work on the project, we consider its application, as well.

3. Environmental Setup, Challenges and Concepts

Ideal *conditions* for video-based breath monitoring are defined by the nurses taking care of the baby and the technicians operating the monitoring setup. The setup should not disturb the work of the nurses. The camera should view the baby from his/her side looking down at an appropriate angle.

The baby should be centered and fill up the majority of the viewfield. The clothing of the baby should be textured to reveal motion. The resolution of the camera should be sufficient for reliable motion detection in presence of noise. The lighting should be adequate so that the baby is clearly visible.

The key *challenges* are the ability to handle infants with different activity level, rhythmicity and intensity of reaction, as well as the ability to cope with babies placed on their back or belly. Further challenges are robustness to environmental changes, including initial start up, temporary occlusions, and changes in lighting conditions, camera angle, position and magnification, whether gentle or abrupt. Finally, the monitoring system should be able to detect and handle emergency conditions.

The first *concept* we use is that of a calibration time window, where a certain time period is considered sufficient to observe the data relative to its neighboring data, for both purposes of positive correlation or negative filtering. Motion detection algorithms which typically use background subtraction to differentiate foreground and background also use a history window, which serves the similar purpose of initial learning.

We assume that changes in intensity and/or location reveal respiratory motion. This necessitates light intensity and distance measurements. The measured signal data must be normalized over controlled temporal windows to provide a sufficiently strong signal. The normalization, however, must not result in noise amplification and false emergency alarms in periods of stillness.

Temporal smoothing of the data in high frame rate videos is another useful concept that helps remove noise and decrease error rate. Last but not least, we rely on the principle of fusion of different multiple measurements to provide redundancy and sufficient support for stable calculations and robust decision-making. In the very beginning of our study, we only used feature point tracking to select the motion best representing the respiration activity. Later on, we decided to consider other cues, as well, in a multi-cue framework. Below, we discuss both options.

4. Methods for RR Estimation

4.1. Tracking-based estimation

The well-known Kanade-Lucas-Tomasi (KLT) method [16] for feature point selection and tracking can be efficiently used to select and track features that reveal breathing. Two versions of tracking-based measurements were developed at the beginning of our study. In these methods, the KLT implementation [4] was used.

The version developed by Dániel Egyed is presented in his thesis [6]. In the initialization (learning) phase, the method uses the KLT tracker to select and track a given number of prominent feature points. During the learning period, displacements in the horizontal direction are considered and the point with the largest and persistent motion is selected to represent the respiration activity.

Then, the method switches to the respiratory rate estimation mode. The signal formed by the horizontal displacement of the selected point is smoothed and its relevant maxima are continuously counted as the signal develops. The operation of this method is illustrated in figure 1 where the boxes show the selected and tracked features, and the breath counting signal is also displayed. More technical details are available in [6]. Note that this version is not rotation-invariant as it is based on a single coordinate rather than the complete displacement vector.

Another version of tracking-based RR estimator was independently developed by Dmitry Chetverikov. His version uses magnitudes of displacement vectors, and it is rotation-invariant. Similarly to the previous version, a given number

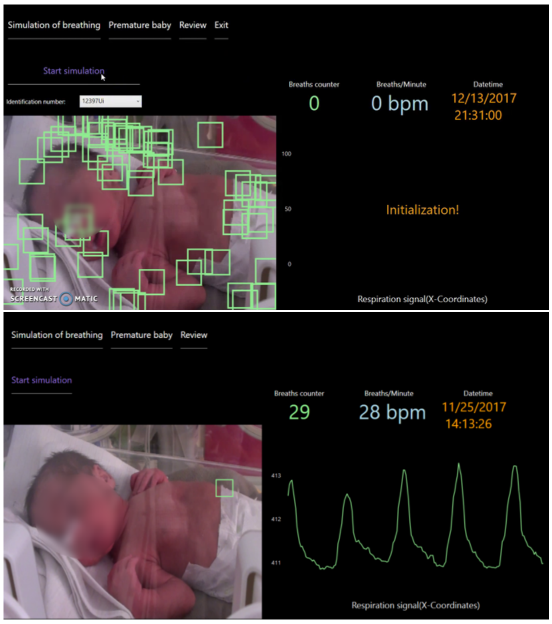

Figure 1: *Initialization (top) and breath counting (bottom) by the method [6].*

Figure 2: *Breath counting signal produced by Chetverikov's method.*

of features is tracked. A large number of features is discarded based on the following criteria: (a) the track is too short (noisy feature); (b) the temporal mean magnitude of displacement vector is too small (noisy motion or static feature); (c) the track contains too large displacements (limb motion, etc.). Then a preset number of features showing distinct, persistent and coherent motion is selected for RR estimation based on the displacement signals. The procedure is illustrated in figure 2 where the features are marked by crosses, and the resulting signal is shown in the bottom.

The preliminary single-cue methods demonstrated that

Figure 3: *The first frame with the ROI.*

feature tracking can be efficiently applied for RR estimation. However, it also became clear that the tracking cue should be supported by other independent cues to enhance the applicability and robustness. Below, we present the fusion-based method developed by Gregory Morse.

4.2. Fusion of multiple measurements

The region of interest is currently set to a centered ellipse mask whose height and width are half that of the video. In the future work, we plan to automatically adapt the ROI to the position of the infant. Figure 3 shows the first frame of the processed video with the ROI indicated by the ellipse. The parameters described below are measured in the ROI. The gray elliptic area on the right-hand side serves for indication of the level of activity in the ROI.

The three prior signals fused by the method are measured based on the following cues:

1. Feature point selection and tracking.
2. Motion detection using background subtraction to find subtle motion.
3. Brightness monitoring in the motion area.

The prior signals are normalized and smoothed, then fused into the final combined signal.

The already mentioned Kanade-Lucas-Tomasi algorithm is used for *feature tracking*. In this case, the software provided by OpenCV [9] is used, where the algorithm is called differently: Lucas-Kanade.

The feature point tracking component of the combined signal is based on the point displacements with respect to the reference positions in the ROI. Initially, the reference positions are the point positions in the first frame. They are periodically updated to account for lost points.

A reference frame containing these reference points is initialized to the starting frame. In each frame, the algorithm selects the points whose displacement vectors w.r.t. the corresponding reference points in the reference frame are larger than is the previous frame. The total number of such ‘positive’ points is denoted by N_p, while the sum of the ‘positive’ displacements in the current frame is denoted by D_p. In a similar way, N_n and D_n are calculated for the points whose displacements decreased. Finally, the tracking-based signal in the current frame is obtained as

$$S_{tr} = \frac{D_p - D_n}{\max\{N_p, N_n\}}$$

The reference frame is changed to the current frame if S_{tr} is negative, or if the reference frame has exceeded the calibration time window. The well-known *background subtraction* technique [17] is used to determine a background model and subtract it to obtain the foreground. The motion detection signal S_{mt} is calculated as the total number of foreground pixels in the region of interest.

The *brightness signal* is calculated in the motion area obtained by background subtraction. Similarly to the tracking signal, the brightness monitoring algorithm sums the ’positive’ points where the brightness increased and the ’negative’ ones where the brightness decreased relative to a reference frame which is initialized and updated in the same way as for the point tracking. The difference between the two sums forms the brightness signal S_{br}.

Each of the three signals is normalized in each frame by finding its extrema values in the previous 25 frames (1 second) and the next 25 frames. After the normalization, the range of all signals is $[0, 1]$. The signals are then smoothed by averaging in a moving temporal window whose width is 7 frames. The final combined signal is the sum of the three normalized and smoothed components.

The peaks and troughs of the fused signal are used to *count breaths* and calculate the respiratory rate. If a trough is detected, the next one can be detected after at least 1 second. The larges peak between each two neighboring troughs is then selected for the counting.

5. Initial Test Results

The main goal of the initial tests was to investigate the feasibility of the proposed multisignal method. To obtain initial ground truth for the test, an approximately 15-minute video had been recorded at 25 frames per second (fps) rate, then processed by the method and simultaneously monitored by a human observer. The observer had to click the mouse each time he saw a breath. These counts and the resulting respiration rates, i.e., counts per minute, were compared to the corresponding numbers obtained by the method.

Figure 4 shows sample frames of the processed video from its first minute. The three frames were selected so as to demonstrate the robustness of the method to camera zoom variation. The first (top) image shows the frame No.502 (20 seconds from the start). This frame was recorded with the initial zoom. One can see the output signal and two counters,

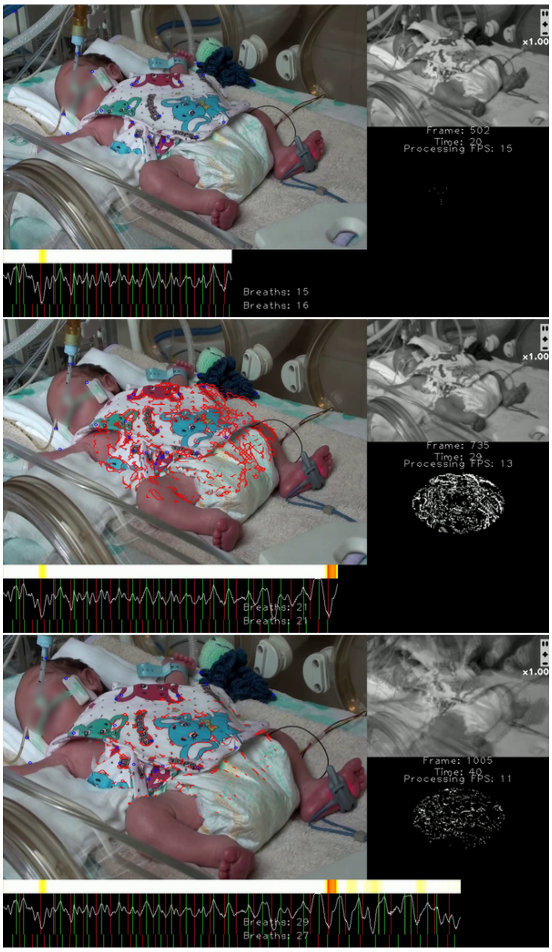

Figure 4: *Sample frames from the first minute of monitoring. Growing breath counts are indicated. Top: before zooming. Middle: during zooming. Bottom: after zooming.*

by the proposed method (top) and by the observer (bottom). The two numbers are quite close. The RR is not obtained yet since the method starts calculating the RR after one minute of monitoring.

The second (middle) image of figure 4 is a frame recorded during the short transitional period of zooming in. The activity level indicator shows large values as changes due to zooming are detected in many parts of the ROI. The amplitude of the output signal increases, but method is not disturbed by the event: the breath counts of the two counters coincide.

After zooming in, the baby occupies a larger part of the viewfield. The third (bottom) image of figure 4 shows a frame recorded after the transitional period. The activity level decreases, the calculated breath count is again close to the manual ground truth.

Figure 5: *Sample frames after the first minute of monitoring. Respiratory rates are indicated as breath counts per minute.*

After the first minute of monitoring, the method starts measuring the respiratory rate. Figure 5 shows two examples of respiratory rate monitoring. The first one (top) was obtained just after the start of RR monitoring, the second one almost 50 seconds later. One can see that the calculated RR is close to, but somewhat higher than, the manual ground truth.

The processing rate indicated on the screen was 11–15 fps in the initialization period and 6–8 fps afterwards. Through all the video, the respiration rate obtained by the proposed method was close to the manual ground truth. However, we experienced that the method tends to count more breaths that the observer. The algorithm can be sensitive to minor motions (peaks in the signal) which are not perceived by the observer. As discussed earlier, breath of premature babies is subtle, non-regular and not easy to sense and perceive. It is possible that a specialist watching the video would count the breaths more precisely that the inexperienced observer used in our test.

A remarkable observation we made in the test is a strong correlation between feature point and light intensity signals. This phenomenon will be analyzed and taken into account in the further development of the multisignal method.

6. Conclusion and Outlook

The results of our initial feasibility study of video-based RR monitoring of newborn babies support the claim that such monitoring is possible. However, many critical issues are still to be addressed to make the system applicable and acceptable in clinical environment:

- Developing methods for automatic ROI extraction.
- Analysis of correlation between the signals in order to modify and enhance them.
- Improving the algorithms for signal normalization and breath counting based on signal features.
- Analysis of traditional contact sensor data:
  - breath-in and breath-out times
  - relations between video-based signal and sensor signal
  - comparison of RR measurements based on the signals
- Ensuring robustness to varying conditions as discussed in section 3:
  - illumination, visibility and camera setup
  - baby activity level, rhythmicity, intensity of reaction
  - differently dressed babies placed on their back or belly
- Setting the optimal values of the parameters.
- Large-scale tests with different babies in varying conditions.
- Achieving real-time operation.
- ...

Due to the critical nature of the task, emergency exception handling and monitoring and calibration to very accurate and precise handling of these cases are absolutely needed before a viable product could result from this project.

Acknowledgments

This study is a part of the project entitled ‘Development of novel intelligent digital neonatal reanimation board (open incubator) to decrease the mortality rate and increase the chances of healthy life of newborn and premature babies’ supported by the grant VEKOP 2.2.1-16-2017-00002. The study was also supported by the European Union, co-financed by the European Social Fund (EFOP-3.6.3-VEKOP-16-2017-00001).

References

1. F.Q. Al-Khalidi, R. Saatchi, D. Burke, H. Elphick, and S. Tan. Respiration rate monitoring methods: A review. *Pediatric Pulmonology*, 46:523–529, 2011.

2. Ufuk Bal. Non-contact estimation of heart rate and oxygen saturation using ambient light. *Biomedical Optics Express*, 6:86–97, 2015.

3. K.D. Benkstein, B. Raman, C.B. Montgomery, C.J. Martinez, and S. Semancik. Microsensors in dynamic backgrounds: Toward real-time breath monitoring. *IEEE Sensors Journal*, 10:137–144, 2010.

4. Stan Birchfield. KLT: An Implementation of the Kanade-Lucas-Tomasi Feature Tracker. `http://cecas.clemson.edu/~stb/klt/`, 2007.

5. 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. *Computers in Biology and Medicine*, 80:158–165, 2017.

6. Dániel Egyed. Implementation of video analysis algorithms in WPF environment for continuous monitoring of prematurely born babies. Master’s thesis, Eötvös Loránd University, 2017. In Hungarian.

7. A.R. Guazzi, M. Villarroel, J. Jorge, J. Daly, M.C. Frise, P.A. Robbins, and L. Tarassenko. Non-contact measurement of oxygen saturation with an rgb camera. *Biomedical Optics Express*, 6:3320–3338, 2015.

8. G.W. Hunter, J.C. Xu, A.M Biaggi-Labiosa, et al. Smart sensor systems for human health breath monitoring applications. *Journal of Breath Research*, 5:037111, 2011.

9. Intel Corporation. OpenCV: Open Source Computer Vision Library. `https://opencv.org/`, 2017.

10. L. Kong, Y. Zhao, L. Dong, Y. Jian, et al. Non-contact detection of oxygen saturation based on visible light imaging device using ambient light. *Optics Express*, 21:17464–17471, 2013.

11. J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek. Non-contact heart rate and heart rate variability measurements: A review. *Biomedical Signal Processing and Control*, 13:102–112, 2014.

12. Z. Li and F. Wang. Respiratory Monitoring Based on Video Processing. In *International Conference on Health Informatics*, pages 35–38, 2014.

13. M.-Z. Poh, D. J McDuff, and R.W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, 18:10762–10774, 2010.

14. H. Qi, Z. Guo, X. Chen, Z. Shen, and Z.J. Wang. Video-based human heart rate measurement using joint blind source separation. *Biomedical Signal Processing and Control*, 31:309–320, 2017.

15. D. Shao, Y. Yang, C. Liu, F. Tsow, H. Yu, and N. Tao. Noncontact monitoring breathing pattern, exhalation flow rate and pulse transit time. *IEEE Transactions on Biomedical Engineering*, 61:2760–2767, 2014.

16. J. Shi and C. Tomasi. Good features to track. In *Conf. on Computer Vision and Pattern Recognition*, pages 593–600, 1994.

17. C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In *Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 246–252, 1999.

18. L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D.A. Clifton, and C. Pugh. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiological Measurement*, 35:807–831, 2014.

19. Wikipedia. Respiratory rate. `https://en.wikipedia.org/wiki/Respiratory_rate`, 2018.

20. H.-Y. Wu, M. Rubinstein, E.e Shih, J. Guttag, F. Durand, and W. Freeman. Eulerian Video Magnification for Revealing Subtle Changes in the World. *ACM Transactions on Graphics*, 31:1–8, 2012.

21. F. Zhao, M. Li, Y. Qian, and J.Z. Tsien. Remote measurements of heart and respiration rates for telemedicine. *PloS one*, 8(10):e71384, 2013.