Novel Methods for Video-based Respiration Monitoring of Newborn Babies

Ádám Nagy[1,2], Dmitry Chetverikov[3,1], and Ákos Zarándy[1,2]

[1] MTA SZTAKI, Hungary
csetverikov@sztaki.hu
[2] Pázmány Péter Catholic University, Hungary
[3] Eötvös Loránd University, Hungary

Abstract. We report on the recent progress in the study of video-based respiration monitoring of babies as a part of a national project[1]. Traditionally, vital parameters of premature babies kept in clinical incubators are continuously monitored by contact sensors that provide signals for measuring the parameters. In our previous paper [7], we presented an initial feasibility study of non-contact video-based monitoring of the respiratory rate (RR). In this paper, we propose two novel methods for RR monitoring and demonstrate their applicability by quantitative comparison to ground truth provided by a human observer.

1 Introduction

As discussed in [4, 7], potential brain lesions of *premature* babies (born before the 37th week of pregnancy) frequently lead to long-term disabilities. Such infants are kept in clinical incubators. The respiratory rate and other vital parameters, including heart rate and blood oxygen level, are continuously measured by monitoring systems that use electroencephalogram (EEG) and polysomnographic devices [2]. The input signals are provided by wired sensors attached to infant body.

Traditional monitoring systems are sensitive to motion artifacts, and their false alarm rates are relatively high. The moderately invasive contact sensors are still inconvenient for babies and stressful for parents. Our work is a part of a national R&D project that aims at substituting contact sensing by non-contact video-based one. We are developing and testing novel methods for video-based measurement of respiratory rate (RR) defined as the number of *breaths per minute* (bpm) [10].

The results of our initial feasibility study on video-based RR monitoring are presented in [7]. The paper provides a brief overview of related previous work on non-contact RR measurement in adults and infants. It is emphasized that the methods developed for adults are not necessarily applicable to premature babies because of essential differences in respiratory rates and patterns. The respiratory signals of the babies are less regular, and their RRs are significantly higher, typically 30–40 bpm. In addition,

[1] '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', grant VEKOP 2.2.1-16-2017-00002

the respiratory motion of infant body is less visible making the motion estimation more difficult.

In the initial study [7], several cues for video-based RR measurement were considered, including feature point tracking, background subtraction, sparse optical flow, and temporal changes of intensity in region of interest (ROI). (See section 2 for technical details.) The main conclusion of the feasibility study was that video-based respiration monitoring of premature babies was possible, and all of the considered cues provided information about respiration. However, in a lack of any quantitative comparison to ground truth, it remained unclear how robust and precise these cues were under varying conditions, and how their data had to be combined.

The novel methods for RR estimation presented in our current paper were developed in order to better understand the process of respiration and its relation to the visual cues, improve the robustness, and provide quantitative comparison to ground truth. The main contributions of this study are the two novel methods and their experimental evaluation based on manual ground truth obtained by a human observer.

The rest of this paper is organized as follows. Section 2 is a discussion of previous work on video-based respiration monitoring of newborn babies, with the accent on our feasibility study [7]. Section 3 presents the two novel methods for RR estimation we developed after the feasibility study. Test results are shown and discussed in section 4. Finally, conclusions and future work are presented in section 5.

2 Previous Work on Respiration Monitoring of Newborn Babies

Our initial paper [7] contains a discussion of related work by other researchers. In this section, we briefly discuss those studies that are most closely related to our approach in terms of the image processing tools they use.

In [6], the authors address the problem of feature point selection and tracking for adult respiration monitoring, mainly in telemedicine. They use the original single resolution implementation of the well-known Kanade-Lucas-Tomasi (KLT) feature point tracker [8]. Tracked points exhibiting large persistent motion are selected as breath indicators. The selected breath feature points are compared to manually selected points as the ground truth. The study [6] does not provide any RR measurement.

In our initial study [4, 7], we applied a modern multiresolution implementation of the KLT tracker [1] for RR measurement based on feature point selection and tracking. In the initialization (learning) phase, a given number of distinct feature points was selected and tracked. Their horizontal displacements were calculated and the point with the largest persistent motion was selected to monitor the respiration activity. After the learning period, the signal provided by the selected point was processed and its relevant maxima were counted to obtain the respiratory rate. The operation of this initial method is illustrated in figure 1. The box shows the selected representative feature point. A part of the obtained respiratory signal is also displayed.

In the study [2], the authors present methods for newborn baby monitoring in order to detect neonatal clonic seizure (sustained rhythmical jerking) and apnea (suspension of breathing). The well-known background modelling and motion detection

Fig. 1. Breath counting by the method [4].

approach [9] is used to extract motion areas. No RR estimation is provided, and it is not clear if the obtained motion data can be used for this purpose.

In our initial study [7], we also used the background modelling approach [9]. However, our method was based on the following three different signals (cues) rather than a single one:

1. Motion detection using background subtraction [9].
2. Sparse optical flow by OpenCV [5] based on [8].
3. Brightness monitoring in the motion area.

The prior signals were normalized and smoothed, then fused into the final combined signal. The peaks and valleys of the fused signal were used to count breaths and calculate the respiratory rate. Figure 2 shows an example of RR monitoring based on the fused signal.

In the previous paper [7], we also discussed *conditions* for efficient video-based breath monitoring. The major conditions can be summarized as follows:

– The monitoring system should not disturb the work of nurses.
– The camera should view the baby from side at a proper angle.
– The baby should be centered and fill up a majority of the viewfield.
– The clothing of the baby should be textured to reveal motion.
– The lighting should be adequate for the visibility of the baby.

Even when the above conditions are fulfilled, the task of respiration monitoring of newborn babies remains *challenging*. The babies can be placed on back or belly and have different levels of activity, including varying rhythmicity and intensity of reaction. Environmental changes can include initial start up, temporary occlusions, and varying lighting conditions, camera angle, position and magnification. Emergency conditions, such as too rapid breathing, too long period without a breath, or anomaly in measurements, must be identified and indicated at low false alarm rates.

Fig. 2. Breath counting in the fused signal [7].

3 Novel Methods for RR Estimation

The methods proposed in the previous study [7] were efficient enough to demonstrate feasibility of video-based respiration monitoring. However, it was clear that further development was needed to improve the performance in terms of precision and robustness. In particular, the method [4] based on feature point tracking was not rotation-invariant as only horizontal displacements were used. In addition, it relied on a single feature point while robust estimation must be based on many points.

The other method based on signal fusion also contained weak points. In particular, signal normalization could amplify noise and even indicate false respiration activity in periods without a breath. Signal fusion could blur extrema of the combined signal leading to imprecise breath counting biased towards signal minima. The method relied on a pre-set ROI while adaptive ROI selection was needed when baby was active.

Below, we present two methods for forming respiration signals. The first one is based on sparse *optical flow* (OF). It resembles the optical flow component of the fused signal described in the previous section, but differs from it in essential details. The second method is based on an efficient implementation of the Singular Values Decomposition (SVD) which is an alternative to the background subtraction method [9], another component of the fused signal. We do not fuse the OF and SVD signals: each signal is treated separately, and two breath counts are obtained and compared to manual ground truth.

3.1 Respiration Signal by Optical Flow

The method based on sparse optical flow uses algorithms provided by OpenCV [5]. Greyscale version of input colour video is used. Optical flow is a set of 2D displacement vectors showing motion between current and previous frames. It assumes no point tracking in frame sequences and it can be calculated in all pixels (dense OF) or selected pixels (sparse OF). In this study, we considered two different ways of selecting the pixels where motion is observed: (1) a set of distinct corner-like feature points as suggested in [8]; (2) a set of points on a regular location grid covering most of the area where respiration activity can occur.

After preliminary tests, the second option was selected. An advantage of the first option is that it uses the locations where optical flow can be reliably calculated. However, it prefers strongly textured parts of image which are not guaranteed to cover the respiratory motion area: it may happen that many of the selected points are static and non-informative. The calculation of displacement vectors on regular grid of 10×10 locations is illustrated in figure 3. The vectors are enlarged for better visibility. Note the dominant direction of the vectors due to breath activity, except for the independent motion of the right leg.

Fig. 3. An example of sparse optical flow on regular grid of image locations.

After the calculation of displacement vectors in the selected locations, the mean displacement vector is obtained and its magnitude is used as the signal value for the current frame. Contrary to the tracking-based method [4] used in the initial study, this solution is rotation-invariant. The obtained magnitude is usually small (subpixel), except for the cases of large motion due to nurse intervention.

Figure 4 shows a sample respiration signal obtained by the method based on sparse optical flow. The plot covers 1000 frames of a test video (40 seconds at 25 fps). The signal is characterized by a series of double peaks, each corresponding to a single breath event (air in/out) as indicated in the signal. The deep minima between breath events are the pauses between breaths. The method we applied to detect the breath events is described later in this section.

3.2 Respiration Signal by Singular Value Decomposition

Our second method aims at detecting and counting pauses between breath motions rather then the breaths themselves. As illustrated later, the pauses are more distinct

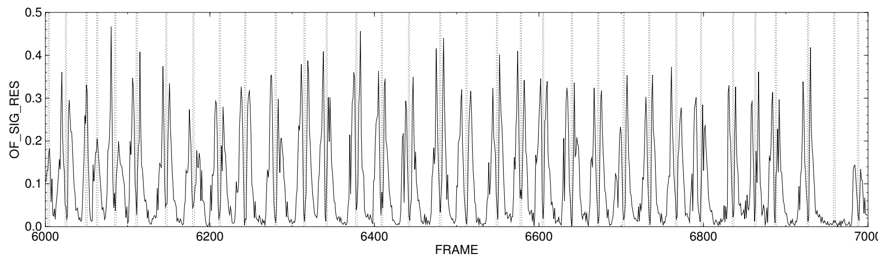

Fig. 4. A sample signal by the OF-based method with the detected breath events indicated.

and easier to detect in respiration signals provided by the fast incremental (running) implementation of SVD [3] we developed earlier. This implementation avoids computing SVD each time from scratch as its temporal window moves forward. Instead, it modifies the previous result by including the entering frame, then excluding the leaving one.

In the SVD-based method for respiration monitoring, ROI is the entire (greyscale) image except pre-defined margins. Its resolution is reduced by a factor of 3–5 in order to speed up the algorithm. The resulting ROI is processed by a median filter to reduce noise and remove artifacts.

The incremental SVD is applied to each 25 consecutive pre-processed ROI images. The 25 images are stacked in a large matrix whose singular vector with the largest singular value provides the (static) background. For each frame, the value of the respiration signal is the mean absolute deviation of the pre-processed ROI from the background. Figure 5 shows a sample SVD signal and the detected pauses between breath motions. Note that each breath is detected twice.

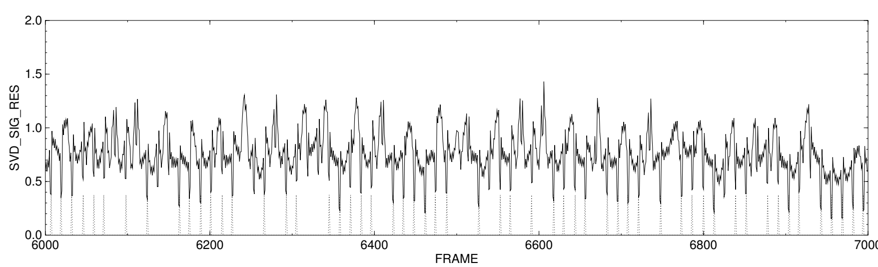

Fig. 5. A sample signal by the SVD-based method with the detected pauses indicated.

3.3 Respiration Signal Processing and Breath Counting

A major difference between OF and SVD respiration signals is that the former has a stable reference level, the zero one, while that of the latter varies and should be extracted adaptively. Otherwise, respiration signal processing and breath counting in the OF and SVD methods are similar. Below, we describe the main steps of the SVD version, then point out the differences in the OF version which is simpler.

The SVD version has the following main steps illustrated in figure 6.

1. Apply large median, then mean filter to obtain the reference level (figures 5, 6a).
2. Subtract the reference level from the signal and negate the result (figure 6b) .
3. Smooth the obtained signal in a similar way to remove its reference level.
4. Adaptively threshold the signal using Otsu's algorithm [5].
5. Remove peaks of unit width, then fill in narrow gaps (figure 6c).
6. Use the end of each run (connected component) to indicate a motion pause.

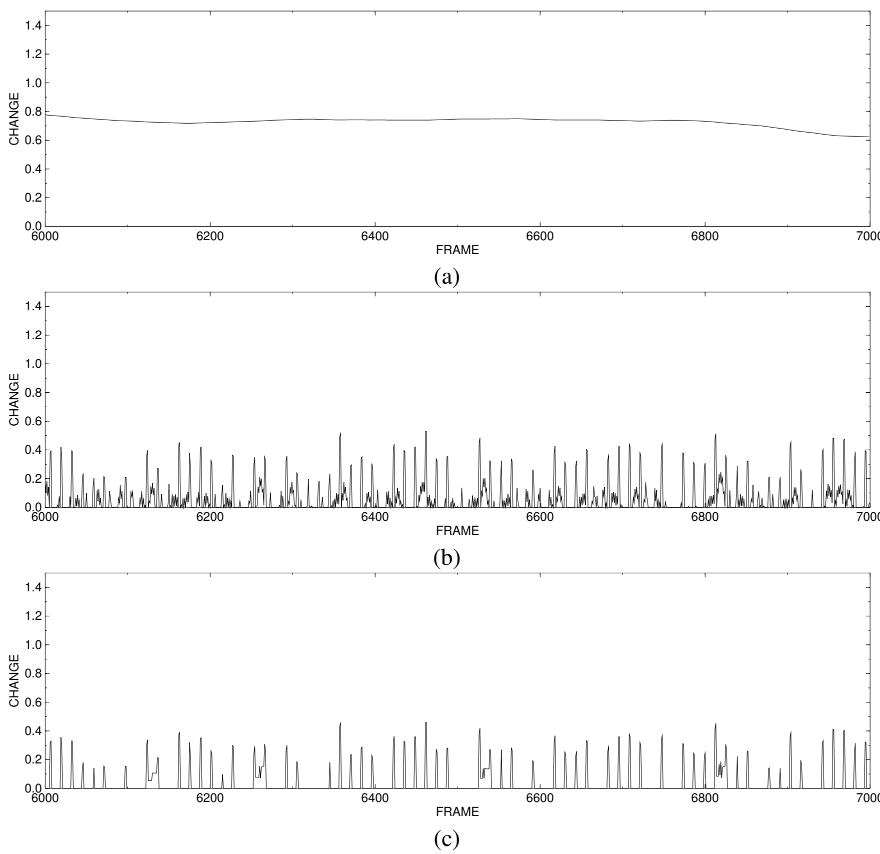

Fig. 6. (a) Reference level of the signal. (b) Signal after step 2. (c) Signal after step 5.

The main parameters of the algorithm are the size of the filters and the minimal width for a gap not to be filled in. The OF version does not need reference level removal as its level is known and stable. We only apply smoothing by a mean filter whose size is much smaller than for SVD signals. The rest of the steps are identical except for the last one where the centres of runs are used instead of the ends.

4 Test Results

The two proposed methods were tested on a 14-minute video (21099 frames) recorded at 25 fps. Ground truth (GT) was obtained manually by a human observer who had to click the mouse each time breath motion (air in or out) was observed. The video was processed by the two methods, and the resulting respiration rates were compared to the ground truth. The results of the comparison are summarized below.

Overall, both methods provided results very similar to the ground truth. For the whole video, the average respiration rates are given in table 1, where 'mean' is the mean of the results by OF and SVD. The result of SVD is a bit closer to the GT than that of OF.

OF	SVD	mean	GT
43.5	42.0	42.8	42.4

Table 1. Average respiration rates for the whole video (bpm).

However, the overall respiration rate only indicates long-term correlation between GT and our results. Figure 7 provides a deeper insight into the behavior of the two methods. It shows the variations of GT and our results on finer temporal scale: for intervals of 40 seconds, i.e., 1000 frames. (The first interval was excluded as the learning period.)

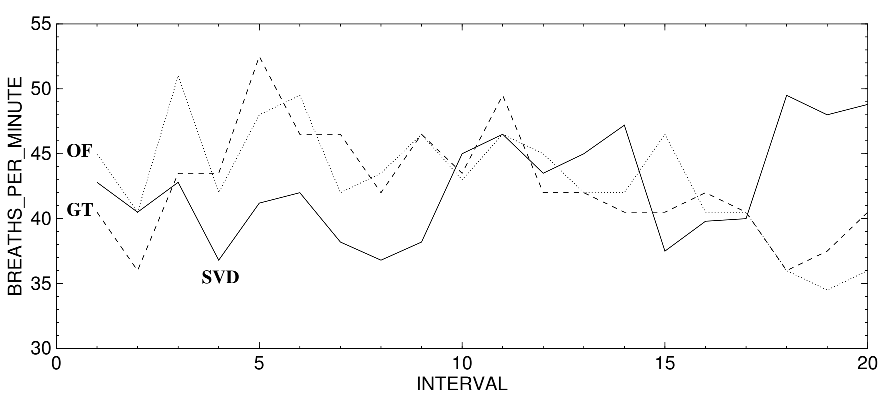

Fig. 7. Average respiration rates for intervals.

One can see that OF better correlates with GT than SVD. The difference is especially distinct in the end of the video when both GT and OF rates decrease while the SVD rate increases. The total absolute difference between OF and GT is 6.4% of the ground truth, the value for SVD is 12.5%. The normalized cross-correlation between GT and OF is 0.67, while for SVD it is only -0.28. In our future work, we have to understand the reason of these deviations.

We also processed two further, shorter videos. One of them had no ground truth, so we could only compare the results of the two proposed methods. We experienced a limited but persistent difference between the two results, with the OF one apparently being closer to the visually observed respiration. In absence of GT, no quantitative conclusion can be drawn, but it is clear that camera-based respiration monitoring is possible here, as well. However, the source of the persistent difference between the two measurements also needs clarification.

Finally, we tried to estimate respiration rate in the third video where the baby lies on his belly. The baby is very active, and his permanent motion as well as frequent interventions by nurse make this video very hard to process properly. In the rare periods of lower activity and no intervention, our results are reasonably close to the manual ground truth. Otherwise, respiration rate here is difficult, sometimes impossible, to estimate. Even the ground truth cannot be reliable since the respiratory motion is often hardly visible by human eye.

5 Conclusion and Outlook

We have presented two novel methods for video-based respiration monitoring on newborn babies. Quantitative comparison of the methods to manual ground truth convincingly demonstrates that reasonably precise, continuous measurement of respiratory rate is possible when activity level of the baby is low and nurse interventions are rare and short.

However, the following important issues need further investigation to make our solutions applicable and acceptable in clinical environment:

- Better understanding of the nature of our measurements.
- Detailed analysis of robustness to varying conditions and parameter settings.
- Large-scale tests with different babies and comparison to traditional sensor data.
- Detecting emergency conditions at low false alarm rates.

We plan to address these issues in near future.

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. Stan Birchfield. KLT: An Implementation of the Kanade-Lucas-Tomasi Feature Tracker. http://cecas.clemson.edu/~stb/klt/, 2007.
2. 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.
3. D. Chetverikov and A. Axt. Approximation-free running svd and its application to motion detection. *Pattern Recognition Letters*, 31:891–897, 2010.
4. 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.
5. Intel Corporation. OpenCV: Open Source Computer Vision Library. https://opencv.org/, 2017.
6. Z. Li and F. Wang. Respiratory Monitoring Based on Video Processing. In *International Conference on Health Informatics*, pages 35–38, 2014.
7. G. Morse, D. Chetverikov, and D. Egyed. Video-based respiration monitoring of newborn babies: a feasibility study. In *Ninth Hungarian Conference on Computer Graphics and Geometry (Grafgeo)*, pages 209–215, 2018.
8. J. Shi and C. Tomasi. Good features to track. In *Conf. on Computer Vision and Pattern Recognition*, pages 593–600, 1994.
9. 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.
10. Wikipedia. Respiratory rate. https://en.wikipedia.org/wiki/Respiratory_rate, 2018.