Non-contact monitoring of the respiratory rate and active and sleeping phases of preterm infants using camera-based techniques

Pázmány Péter Catholic University

Roska Tamás Doctoral School

of Sciences and Technology

Ádám Nagy

Supervisor: Dr. Ákos Zarándy

A thesis submitted for the degree of *Doctor of Philosophy*

2023

Introduction

Physiological indicators, including pulse rate, respiratory rate, and blood oxygen level, are routinely monitored in intensive care units and neonatal clinics to assess the health of premature infants and adult patients. In the case of neonates, behavioral states such as quiet sleep, active sleep, transient, quiet alert, and active alert, which are classified according to the NIDCAP (Newborn Individualized Developmental Care and Assess) scale, are important parameters for describing sleep and activity patterns. These states are utilized to estimate the amount and quality of sleep and movement of the patient. Sleep is crucial for proper neonatal development, but it is also essential for adults to obtain adequate amounts and quality of sleep.

Traditionally, these parameters were measured using contact sensors (that in many cases were not installed) or relied on the senses of hospital staff. However, recent advancements in machine vision, image processing, and machine learning have led to a technological revolution in many applications. Generative algorithms (such as Dall·E 2 and ChatGPT) are becoming increasingly popular for various applications. Additionally, image processing algorithms have undergone significant development, including the emergence of new types of convolutional networks that are preferred in medical fields. Such new technologies give us the opportunity to reshape infant observation as well.

The present thesis also contributes to the advancement of the medical field through the development of novel machine vision and machine learning algorithms for monitoring premature infants. Specifically, three challenges were explored in this field including: contact-less monitoring of infants' breathing, automatic annotation of the data sets required by the previous algorithm, and classification of observed behavioral states of neonates. These challenges present significant difficulties in the

field and were addressed through the creation of innovative algorithms.

Physiological parameters such as respiration rate are routinely monitored in medical settings, particularly in neonatal wards, utilizing medical monitors such as the 'Phillips IntelliVue MP40/MP50'. However, the use of wired sensors in such monitors poses significant issues for both infants and medical staff. For example, 'contact' sensors can cause skin irritation and discomfort to neonates, while the wires can interfere with medical personnel's ability to carry out their duties effectively. As a result, non-contact monitoring methods have emerged as a promising alternative to wired sensors.

In accordance with 'Practical Skills in the Field of Family-Centered Development Support Care' guidelines approved by 'FINE' [17], medical and nursing personnel are required to continuously monitor the sleep and activity of neonates manually to administer appropriate developmental support care. Nevertheless, many medical institutions are often understaffed, and this level of continuous monitoring can be difficult to sustain. Consequently, there is an unmet need to automate the classification of behavioral states in neonates.

It is evident that in both scenarios, there is a pressing need to enhance existing methodologies, possibly through the replacement with machine vision and artificial intelligence-based algorithms. However, a common challenge in the utilization of AI-based algorithms is the requirement for extensive annotated datasets. Within this thesis, we address this challenge, particularly with respect to the non-contact monitoring of breathing, by presenting a viable solution for annotated data generation.

Applied methods and tools

To enable the further development and assessment of contactless algorithms for estimating the physiological and behavioral states of infants, a substantial volume of annotated data is required. However, due to ethical and other considerations, publicly accessible databases on this subject are lacking. Thus, we undertook the construction of our own database for this purpose.

The initial step in the process involved developing a data collection system capable of acquiring the image inputs and the expected outputs synchronized for our data-driven algorithms. For breathing monitoring, the task is comparatively simpler due to the current practice in hospitals, where ECG (electrocardiograph) is utilized for monitoring breathing, which we aim to replace with a non-contact approach. Consequently, image data was acquired through a color camera while the respiratory signal was obtained through the ECG, which was connected to a 'Phillips IntelliVue MP40/MP50' monitor. We constructed a data collection setup equipped with an infrared illuminator and a camera, capable of connecting to the aforementioned medical monitor. As a result, this system could collect image data facilitated by the infrared illumination even during nighttime, synchronized with the respiratory reference signal.

During the classification of the infant's behavioral states, a color camera equipped with an infrared illuminator was again used to record image data. However, in this case, the expected outcomes or behavioral states were recorded via a mobile phone application by the medical personnel who observed the preterm infants in question.

In terms of sources, the initial literature overview involved examining the relevant literature in the form of review articles that focused on contactless breathing monitoring of infants [9] and that focused on the non-contact monitoring sleep states

[19]. Subsequently, the literature on breathing monitoring techniques based on color camera motion analysis [8–10, 14] and the observation of infant behavioral states according to NIDCAP [2, 7, 17, 20] were comprehensively reviewed.In addition, the literature on infant observation was thoroughly reviewed from both contact monitoring and non-contact monitoring perspectives.

To manage the data and annotations, we created our own customized C++ based software, which allowed us to query and categorize the data from the database and display both the image data and the reference signals.

To address the aforementioned issues, machine learning and machine vision methodologies were applied. In recent years, convolutional neural networks have undergone significant advancements and have been successfully applied across multiple domains. Within the domain of motion estimation, various forms of optical flows have long been an established technique. A concise overview of the non-contact breathing monitor is depicted in Figure 1.

And RNN-s (recurrent neural network) such as LSTM (long short-term memory) and GRU (gated recurrent unit) have already been proven in many applications in both time series analysis and prediction. I have used convolutional networks as ROI (region of interest) detectors and 'dense' optical flow as motion estimators in my proposed breath monitoring algorithm. And for the classification of behavioural states, I used a GRU-based classifier with inputs generated using various machine learning and image processing techniques (see Figure 2). I also used machine learning and image processing techniques (e.g. simple morphological methods) and differential image analysis to generate automatic data annotation.

The custom algorithms were developed using Python programming language and the Pytorch and Pytorch-Lightning software packages. Subsequently, the algorithms were implemented and evaluated in C++. Data required for training and testing the machine learning models was acquired using the aforementioned data collection system in the Neonatal Intensive Care Unit at Semmelweis University II, Obstetrics and Gynecology Clinic. The collected data was then classified using our own software to determine the dataset suitable for training purposes.

The validation procedures were conducted using both Python and C++ on a

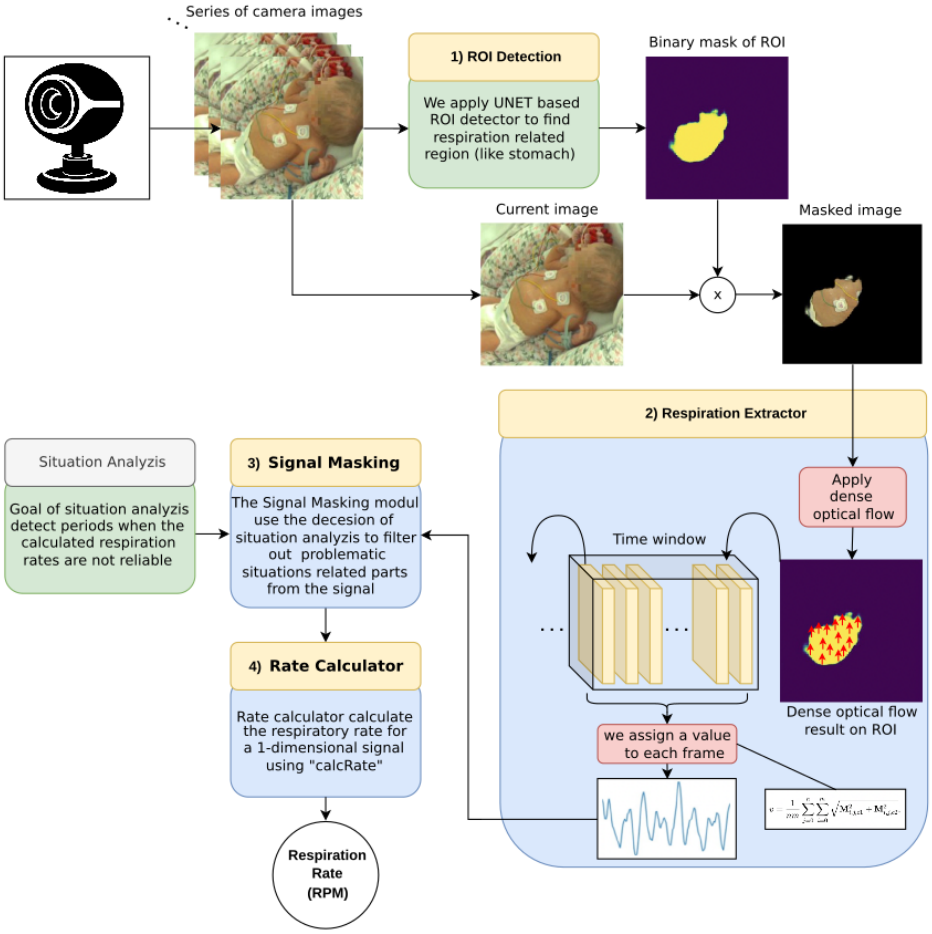

Figure 1: The present study proposes a novel non-contact breathing monitoring algorithm consisting of a UNET-based ROI detector (1) followed by a dense optical flow-based breathing extractor (2) and a peak detector-based breathing rate calculator (4). The algorithm also integrates a situation analysis module (3) that enables the identification of problematic situations, thereby facilitating reliable breathing rate estimation.

separated test dataset, comprising images of infants for which were not included in the training set. Furthermore, the proposed algorithms were deployed on resource-constrained embedded platforms, such as Raspberry Pi or Jetson Nano, and demonstrated real-time operation.

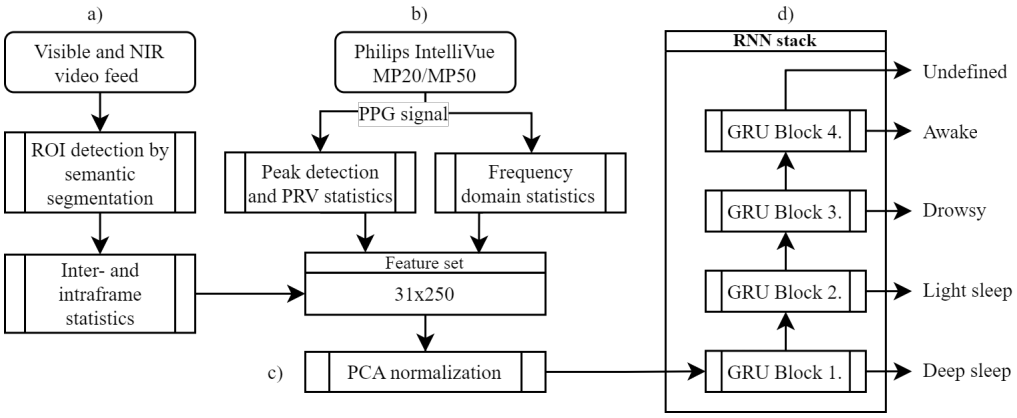

Figure 2: The diagram presents a concise overview of the algorithm for estimating behavioral states, with part a) depicting the video-based feature extraction modules, part b) displaying the computation of PPG signal-based features and their Principal Component Analysis (PCA), and part c) illustrating the classification module used for determining the prevailing behavioral state during the relevant time period.

New scientific results

The present doctoral research has comprehensively investigated numerous challenges arising from the monitoring of premature neonates within incubators.

We built a database and created software components that are able for database management and (even automatic) annotation.

In summary, data collection systems, data annotation tools and software have been developed, as well as algorithms that take a novel approach to the non-contact monitoring of certain vital signals and states of infants.

I have explored three challenging research topics in this area. Namely, (i) contactless monitoring of breathing in premature infants, (ii) automatic annotation of the data set used to train the required ROI detector, and (iii) contactless monitoring of infants' behavioural states.

I used machine vision and machine learning algorithms to solve the problems. Using these, I calculated the breathing of the infants in the image based on movement and calculated their behavioural states (according to the NIDCAP scale) based partly on movement and partly on certain vital signs of the infants.

In the following I present my new scientific findings in this thesis, which I have formulated in 3 thesis points.

The first problem we solved using camera-based algorithm is the non-contact respiratory monitoring of newborns. For its solution, a new method was applied involving a neural network-based ROI (Region of Interest) detector that identifies the infant's abdomen in the image, quantifies the motion in this area, and determines the respiratory rate. This process can be divided into four main parts:

1. ROI Detector

2. Respiratory Extraction Module

3. Respiratory Signal Masking Module

4. Rate Calculation Module

1) ROI Detector

The ROI Detector itself is a U-Net ([11]) applied for trunk segmentation, as this part of the images carries the most information related to respiration. By doing so, we can eliminate the interfering motion components of the limbs. The result of ROI detection is a binary mask, where pixels belonging to the trunk are assigned a value of 1, while pixels outside the trunk are assigned a value of 0. This binary image that represents the ROI serves as one of the inputs for the 'Respiratory Extraction Module.'

2) Respiratory Extraction Module

The task of the Respiratory Extraction Module is to generate a one-dimensional respiratory waveform from the video. The algorithm includes a sliding window that is populated with motion images calculated from the incoming camera frames as follows: From the new incoming frame and the preceding frame, the motion image is calculated (using Farneback's optical flow method [5]), and this resulting motion image is element-wise multiplied by the binary image obtained from the 'ROI Detector' according to the following equation:

$$\mathbf{M} = \text{U-Net}(\mathbf{I}(t)) \odot \text{Dense}(\mathbf{I}(t), \mathbf{I}(t-1)), \tag{1}$$

where $\mathbf{I}(t) \in \mathbb{R}^{\mathrm{m}\times\mathrm{n}}$ represents the current frame obtained from the camera at time t, where our recordings have been characterized by $m = 500$ and $n = 500$. The 'U-Net()' method is the application of the ROI detector on the current frame, while the 'Dense()' method calculates the dense optical flow using Farneback's method for the current and preceding frames. $\mathbf{M}$ denotes the motion image obtained by multiplying the binary ROI image with the optical flow.

The resulting motion image is inserted at the beginning of the sliding window. Since the sliding window should remain fixed in length, the last element is removed.

Next, for each frame in the window, a value (v) is assigned using Equation 2, resulting in a 1-dimensional vector of length N=200. This vector contains elements of type 'float-64'. In the subsequent step, the 1-dimensional waveform is filtered with a bandpass Butterworth filter (20–120 RPM) [3] to obtain an output waveform whose frequency falls within the physiological frequency range of respiration and correlates with the infant's respiration.

$$v = \frac{1}{nm} \sum_{j=0}^{n} \sum_{i=0}^{m} \sqrt{\mathbf{M}_{i,j,c1}^2 + \mathbf{M}_{i,j,c2}^2}, \tag{2}$$

where $\mathbf{M}_{i,j,c1}$ represents pixel (i, j) of the current motion image calculated as described above, on channel c. The motion image has two channels: $c1$ represents horizontal motion, while $c2$ represents vertical motion. The spatial dimensions of the motion image are denoted by n and m. The sequence of assigned values v for each masked motion frame defines the 1-dimensional motion signal that correlates with the respiration: $\mathbf{s} = [v_1, v_2, ..., v_N]$.

3) Respiratory Signal Masking Module

From the 'situation analysis' module, based on a Recurrent Neural Network (RNN), we obtain a decision for each incoming frame, determining whether calm respiration occurs in the current situation based on the computed features from the current and preceding frames. The decisions can be binary values of 1 or 0, which are also stored in a sliding window (decision window). This window contains 1s and 0s, where 1s correspond to time periods when the non-contact respiratory monitor can reliably provide the respiratory rate, and 0s indicate problematic situations. The Respiratory Signal Masking Module filters out the motion signals associated with problematic situations by element-wise multiplication between the vector containing the obtained waveform and the vector containing the elements of the decision window. This process zeros out the segments falling within the time range of problematic situations.

4) Rate Calculation Module

This is the final module of the non-contact respiratory monitor. Its task, as the name suggests, is to determine the respiratory rate. Its input is the respiratory wave-

form (**s**), which is a 1-dimensional vector of type 'float-64' with a length of 200. The output is the respiratory rate in breaths per minute (BPM) represented by a 'uint8' value ranging from 20 to 120 BPM. (This range corresponds to the physiological range of infant respiration according to our literature review and experiments.)

For rate calculation, we developed a peak detection-based algorithm called CalcRate. This algorithm starts with signal inversion (1) and applies a bandpass filter (2) that allows frequencies between 20 and 120 to pass. This is followed by an adaptive thresholding step, where the threshold level is the adaptive (calculated based on the current signal's local environment) mean value (3). This yields a binary signal, where a value of 1 represents breaths (positive peaks), and the onset of breaths is indicated by rising edges (4). By counting the rising edges (5), we can determine the number of breaths in a given 'time window'. In the final step, breaths within 0.3 seconds of each other are merged (6), and the respiratory frequency is calculated by averaging the time intervals between peaks and converting the result to frequency value (7).

I. Thesis point: *I propose a novel contactless camera-based respiration monitoring algorithm that estimates the respiration rate of premature infants in the image by analysing a sequence of frames captured by a RGB camera and that works reliably in real hospital conditions, as it is able to detect situations where the estimation of the respiration rate is not possible, such as when treatment/feeding is taking place or the infant is removed from the incubator.*
Related publications: [J1]

The second algorithm I designed, which I would like to present in more detail, is closely related to the first non-contact respiratory monitoring algorithm. Specifically, it performs the automatic annotation of the dataset required for the training of the first module of the previous algorithm, the ROI detector. Manual annotation, especially for semantic segmentation with pixel-level annotation, is a laborious task. Furthermore, when working with a limited number of annotators, subjectivity can pose a problem. The annotation of the dataset can vary based on individual anno-

tators' judgments. With automatic machine annotation, this problem is mitigated, while saving us from tedious annotation work.

This automatic annotation algorithm can be divided into two main parts:

- Detection module
- Tracking module

The detection module's task is to locate the region associated with respiratory motion based on the extent and frequency of motion. The tracking module aims to track the detected region's potential motion to avoid running the detection module continuously.

The more complex part is the detection module, which can be summarized in four main steps:

- 1) Motion estimation
- 2) Area-based filtering
- 3) Frequency-based filtering
- 4) Displacement-based filtering

1) Motion Estimation The goal of this step is to generate a motion image that indicates the intensity of motion around each pixel between the current frame at time t and the preceding frame at time $t-1$. Several algorithms exist for creating such motion images, such as dense optical flow, block-matching algorithms [4], or 'DeepFlow' [18], which was applied by Yue Sun et al. for estimating respiratory motion [15]. However, the simplest solution is to calculate the difference image between the two mentioned frames. This procedure does not provide information about the direction of motion, but it captures the intensity. Since we do not require direction information for this application, and the squared difference image ($\mathbf{D}$) effectively reveals the contrast of motion, enabling the detection of large abdominal movements (especially the motion at the diaper and skin junction), this method is perfect for our application. The difference image alone would be noisy. However, squaring the difference image ($\mathbf{D}^2$) enables the detection of the largest areas of

motion, and it can be computed quickly and easily (see Equation (3)). Squaring enhances stronger motions and suppresses noise, improving the signal-to-noise ratio. This way, we obtain a single-channel rotation-invariant motion image.

$$\mathbf{D}(x, y, t) = (\mathbf{I}(x, y, t) - \mathbf{I}(x, y, t-1))^2, \tag{3}$$

where $\mathbf{I}(x, y, t)$ represents the intensity of the pixel at coordinates (x,y) at time t, and $\mathbf{D}(x, y, t)$ denotes the intensity of the squared difference image at coordinates (x,y) for time t.

2) Area-based Filtering

After calculating the aforementioned motion image, if the infant exhibits calm respiration without any other motion in the image, we can locate the region around the abdomen where respiratory motion occurs. However, if larger connected motion areas are visible in the $\mathbf{D}(x, y, t)$ image or if they are unrealistically small, we can determine that these do not stem from abdominal respiratory motion. Hence, if the camera-to-infant distance is fixed, we can establish lower and upper thresholds for the pixel count of the largest connected component in the $\mathbf{D}(x, y, t)$ image. If the pixel count of the largest component does not fall within this range, the detection is considered unsuccessful.

To compute this, we first need to apply thresholds to the $\mathbf{D}(x, y, t)$ motion image. Subsequently, on the resulting binary image, we can find the connected regions using the spaghetti algorithm [1]. The spaghetti algorithm returns a list of pixel clusters, where a cluster contains pixels that are connected to each other. The algorithm assigns a cluster label to each pixel in the output image, where the background is treated as a separate cluster and is not considered further. We then select the largest connected component that differs from the background (see Equation (4)).

$$g(x, y) = \begin{cases} 1, & \text{if } l(x, y) = \phi \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $l(x, y)$ represents the group number assigned by the spaghetti algorithm to the pixel at coordinates (x,y), and ϕ denotes the group number of the cluster that

contains the most pixels.

In this way, we have selected the largest connected component (g) in which the image contains only 0s and 1s. Next, we examine whether the pixel count of g falls within the allowed range: $Th_{lower} < \sum_{x,y=1,1}^{M,N} g(x,y) < Th_{upper}$. These are empirically set threshold values. If the pixel count of the largest connected component is too small, it likely does not correspond to the respiratory motion in the abdominal region. If it is too large, it likely represents intense limb motion or other interfering moving objects. If the pixel count falls within the allowed range, the algorithm can proceed to the next step of the detection. Otherwise, the algorithm returns to the first step, which applies motion estimation to the next incoming frame.

3) Frequency-based Filtering

Area-based filtering alone is not sufficient to find the region associated with respiratory motion in the image, as there may be other motions (e.g., limb movements) whose extent in the motion image is similar to that of abdominal respiratory motion. Therefore, additional filtering based on the frequency attribute of respiratory motion is necessary. To achieve this, we analyze the temporal pattern of motion. However, before that, we need to transform the sequence of motion images into a one-dimensional signal.

Similar to previous approaches, we apply a sliding window, but instead of image frames, the motion images are appended to the window. The window contains 300 motion images calculated according to Equation (3). These images are then fattened into one-dimensional vectors, referred to as samples, with a length of H = N x M, where N represents the horizontal and M the vertical extent of the motion images. Thus, in the sliding window, instead of 2D motion images, we have 1D-transformed motion images, resulting in a matrix of size (300 x H). We can then apply Principal Component Analysis (PCA) [16] to this matrix, which maps the samples to the eigenvector space. By selecting the first component, we perform dimensionality reduction, reducing the samples to one dimension. As a result, we obtain a 1D motion signal, which is 300 units long and describes the most emphasized motion identified by PCA among the motions present in the motion image.

After extracting the signal, we can examine whether its frequency falls within the

range corresponding to the physiological frequency of respiration. Healthy newborns typically breathe in the frequency range of 25 to 68 breaths per minute [6]. However, based on our measurements, in certain (even abnormal) cases, higher respiratory frequencies may occur. Therefore, we expanded the possible range of respiratory frequencies to 20-120 breaths per minute. To assess whether the frequency of the 1D signal associated with the current content of the sliding window falls within the physiological range of respiration, we calculate the Fast Fourier Transform (FFT) spectrum of the signal and select the frequency corresponding to the highest peak in the spectrum.

$$f = \mathbf{b}[\max_{f} \{FFT(\mathbf{s}_M)\}], \tag{5}$$

where $\mathbf{s}_M$ represents the motion wave associated with the sliding window mentioned earlier, and $\mathbf{b}$ is a vector containing the frequencies of the FFT spectrum.

Next, we examine whether this frequency is within the specified range (20 RPM - 120 RPM). If the frequency (f) corresponding to the highest peak in the computed frequency spectrum does not fall within this range, the detection is not successful on the current motion image, as we cannot guarantee that the detected motion is related to respiration. If it falls within the range, the algorithm can proceed to the next step. Otherwise, the algorithm returns to the first step, which applies motion estimation to the next incoming frame.

4) Displacement-based Filtering

The aforementioned filters may already be suitable for detecting the region containing respiratory motion. However, in certain cases, there might be motion with the appropriate frequency visible in the sequence of motion images, and the extent of the largest connected component in the last image might also be appropriate, yet this connected component is not a consequence of respiratory motion in the motion image. In such cases, calm respiration can be observed in the image, but sudden low-intensity motion occurs in a region other than the abdominal area, such as around the limbs. In rare cases, the extent of this limb motion might coincide with the search range, resulting in the detection of the region around the limb as

the source of respiratory motion. To avoid this, we introduced displacement-based filtering, which monitors the motion of the geometric center of the largest connected component and considers the detection unsuccessful if a sudden, unrealistically large displacement is observed.

To achieve this, we again apply a sliding window, which, in this case, contains coordinates $\mathbf{c}_i = (\mathbf{x}_i, \mathbf{y}i)$ instead of image frames. If the distance between the centers within the sliding window never exceeds the empirically set threshold Th_{st}, the detection is successful.

$$\forall \mathbf{c}_i \in \mathbf{c}, d(\mathbf{c}_i, \mathbf{c}_0) < Th_{st} \tag{6}$$

where $i < N$ represents the ith index, N is the length of the sliding window containing the center points, $d()$ denotes the Euclidean distance between two input points, and $\mathbf{c}$ is a vector containing $N = 300$ peaces of 2D points, where $\mathbf{c}_i \in \mathbb{R}^2$.

Tracking Module

The task of the tracking module is to track the centroid of the detected region using 'sparse' optical flow. During tracking, the module saves a binary image that contains the detected region (annotation) and also saves the corresponding current input image. Thus, it generates an annotated dataset. Tracking continues until something occludes the tracked point or it moves out of the image, or the algorithm switches to a new video. In these cases, we need to return to the running of detection module. If the detection module re-finds the region related to respiration, automatic annotation can resume.

Thesis point II: *I showed that by using a neural network-based segmenter to find the expected location of breathing, it is possible to automatically generate an annotated database using continuous video recordings of premature infants. I have proposed an algorithm that detects and tracks the location of respiratory motion based on motion extent and frequency to annotate frames for the segmenter from continuous video recordings of premature infants.*

Related publications: [C1]

The third algorithm (Figure 2) I would like to present performs classification into behavioral states according to NIDCAP using video recordings of infants and pulse variability data obtained from the Philips IntelliVue MP20/MP50 medical monitor.

The operation of this algorithm can be divided into three main steps:

- 1) Extraction of video-based features
- 2) Extraction of pulse variability-based features
- 3) Application of the classification module

1) Extraction of video-based features

The extraction of video-based features involves detecting predefined informative regions in the image, such as the infant's abdomen, the image borders, or the area around the abdomen overlapping with the limbs. Various motion signals are extracted from these regions, such as the difference image within a given area, dense optical flow, or the average of pixels within the region in the HSV-transformed image. We calculate eight types of motion signals for three informative regions, and these signals are combined with the respiratory signal, resulting in a set of 25 one-dimensional waveforms.

2) Extraction of pulse variability-based features

Since every newborn is continuously monitored with a pulse oximeter but not with an EKG, we decided to use the pulse signal from the contact sensor and calculate metrics based on pulse rate variability (PRV). The most straightforward approach is to apply time-domain metrics. If the peaks are detected, the durations between them can be calculated. From such a sequence of intervals, more complex metrics can be computed for comparing different phases. The computation of the complex features we used is described by equations 7 and 8.

$$RMSSD = \sqrt{\frac{\sum_{i=1}^{N}(\mathbf{PP}_{i+1} - \mathbf{PP}_i)^2}{N}} \tag{7}$$

$$SDNN = \sqrt{\frac{\sum_{i=1}^{N}(\mathbf{PP}_i - \mathbf{mPP})^2}{N-1}} \tag{8}$$

where **PP** represents the time difference between two successive pulse peaks, $\mathbf{PP}_i$ is the time difference corresponding to the i-th pair of pulse peaks, and **N** is the number of pulse peaks. **mPP** denotes the average of the PP intervals, **SDNN** ([13]) is the standard deviation of the aforementioned PP intervals from **mPP**, which is the square root of the variance, while **RMSSD** is the square root of the mean of squared differences between successive normal heartbeats. The calculation is performed in a 5-minute sliding window, following the HRV metrics overview [12].

The computed PRV-based metrics (which are also one-dimensional signals of a given length) are concatenated with the previously created video-based features. Then, a PCA normalization is applied to the resulting feature set. This feature set serves as the input to the classification module, implemented as an RNN cascade-based module.

3) Classification module

The classification into NIDCAP-based behavioral classes is performed by an RNN (specifically, a GRU - Gated Recurrent Unit) cascade-based module. To enhance performance, the GRU classifiers (blocks) are organized into a cascaded structure resembling a decision tree. The blocks differentiate behavioral states in the order of deep sleep, active sleep, transitional, and wakefulness. The order is important because a subsequent block is only make classification for the given period's feature set if the preceding blocks did not yield a positive classification. The most accurate binary classifiers are listed above. The accuracy of the cascade is thereby increased for the entire class set. The binary classifier blocks consist of GRU layers, linear layers with ReLU activation functions, and a sigmoid activation function at the end.

Thesis point III: *I propose an algorithm to classify the behavioral states of infants (NIDCAP scale) by extracting 1D temporal motion signals from a sequence of consecutive frames (visual actigraph) and to input the resulting signal sequence combined with pulse variability-based features to a recurrent-neural-network-based classification cascade. It's output is one of the behavioural states and, to the best of our knowledge, it is the first AI-based classifier that classifies into these NIDCAP scale states using only image data from the camera and pulse variability-based data*

as input.

Related publications: [C2], [P1], [P2]

Journal publications of the thesis

[J1] **A. Nagy**, P. Földesy, I. Jánoki, D. Terbe, M. Siket, M. Szabó, J. Varga, and A. Zarándy, "Continuous camera-based premature-infant monitoring algorithms for nicu," *Applied Sciences*, vol. 11, no. 16, pp. 1–24, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/16/7215 10)

[J2] P. Földesy, I. Jánoki, **A. Nagy**, M. Siket, and A. Zarándy, "Temperature measurement with photodiodes: Application to laser diode temperature monitoring," *Sensors and Actuators A: Physical*, vol. 337, p. 113441, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924424722000796 19)

[J3] P. Földesy, M. Siket, **A. Nagy**, and I. Jánoki, "Correction of overexposure in laser speckle contrast imaging," *Opt. Express*, vol. 30, no. 12, pp. 21 523–21 534, Jun 2022. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-30-12-21523 19)

[J4] P. Földesy, M. Siket, I. Jánoki, K. Demeter, and **A. Nagy**, "Ensemble averaging laser speckle contrast imaging: statistical model of improvement as function of static scatterers," *Opt. Express*, vol. 29, no. 18, pp. 29 366–29 377, Aug 2021. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-29-18-29366 19)

Conference publications of the thesis

[C1] **A. Nagy**, P. Földesy, I. Jánoki, M. Siket, and A. Zarándy, "Automatic annotation to train roi detection algorithm for premature infant respiration monitoring in nicu," in *2023 In: Képfeldolgozók, és Alakfelismerők Társasága Képfeldolgozók és Alakfelismerők Társaságának 14. Országos Konferenciája*, no. 24, 01 2023, pp. 1–27. 15)

[C2] I. Jánoki, **A. Nagy**, P. Földesy, A. Zarándy, M. Siket, J. Varga, and M. Szabó, "Neonatal activity monitoring by camera-based multi-lstm network," 10 2022, pp. 1–4. 18)

[C3] A. Zarándy, P. Földesy, **A. Nagy**, I. Jánoki, D. Terbe, M. Siket, M. Szabó, and J. Varga, "Multi-level optimization for enabling life critical visual inspections of infants in resource limited environment," in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2020, pp. 1–5. 19)

Other publications of the author

[Au1] **A. Nagy**, D. Chetverikov, and A. Zarándy, "Novel methods for video-based respiration monitoring of newborn babies," in *2019 In: Képfeldolgozók, és Alakfelismerők Társasága Képfeldolgozók és Alakfelismerők Társaságának 12. Országos Konferenciája*, no. 22, 08 2019, pp. 1–10. 19)

Poster publications of the thesis

[P1] P. Földesy, **A. Nagy**, Z. Róka, Zarándy, I. Janoki, J. Varga, and M. Szabó, "Automatic nidcap sleeping stage monitoring by deep learning network using video actigraphy and heart rate variability," in *Poster presented at 32nd Annual Meeting of the European Society of Paediatric and Neonatal Intensive Care*. ESPNIC, 2023, p. 1. 18)

[P2] **A. Nagy**, I. Jánoki, A. Zarándy, P. Földesy, M. Szabó, and J. Varga, "Computer vision based neonatal behavior stage detection and classification," in *Poster presented at 9th Congress of the European Academy of Paediatric Societies*. EAPS, 2022, pp. 1–2. 18)

Bibliography

[1] Federico Bolelli, Stefano Allegretti, Lorenzo Baraldi, and Costantino Grana. Spaghetti labeling: Directed acyclic graphs for block-based connected components labeling. *IEEE Transactions on Image Processing*, PP:1–1, 10 2019. 12)

[2] Aude BRIS, Nadia Mazille-Orfanos, Pauline SIMONOT, Maude LUHERNE, Cyril Flamant, Geraldine Gascoin, Gearóid ÓLaighin, Richard Harte, and Patrick Pladys. Parents' and healthcare professionals' perceptions of the use of live video recording in neonatal units: a focus group study. *BMC Pediatrics*, 20, 04 2020. 4)

[3] Stephen Butterworth et al. On the theory of filter amplifiers. *Wireless Engineer*, 7(6):536–541, 1930. 9)

[4] G. de Haan, P.W.A.C. Biezen, H. Huijgen, and O.A. Ojo. True-motion estimation with 3-d recursive search block matching. *IEEE Transactions on Circuits and Systems for Video Technology*, 3(5):368–379, 388, 1993. 11)

[5] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. volume 2749, pages 363–370, 06 2003. 8)

[6] Susannah Fleming, Matthew Thompson, Richard Stevens, Carl Heneghan, Annette Plüddemann, Ian Maconochie, Lionel Tarassenko, and David Mant. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. *The Lancet*, 377(9770):1011–1018, 2011. 14)

[7] Deanna Hanson-Abromeit. The newborn individualized developmental care and assessment program (nidcap) as a model for clinical music therapy interventions with premature infants. *Music Therapy Perspectives*, 21:60–68, 01 2003. 4)

[8] João Jorge, Mauricio Villarroel, Sitthichok Chaichulee, Alessandro Guazzi, Sara Davis, Gabrielle Green, Kenny McCormick, and Lionel Tarassenko. Non-contact monitoring of respiration in the neonatal intensive care unit. In *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, pages 286–293, 2017. 4)

[9] Lalit Maurya, Pavleen Kaur, Deepak Chawla, and Prasant Mahapatra. Non-contact breathing rate monitoring in newborns: A review. *Computers in Biology and Medicine*, 132:104321, 03 2021. 3, 4)

[10] Á Nagy, D Chetverikov, and Á Zarándy. Novel methods for video-based respiration monitoring of newborn babies. In *2019 In: Képfeldolgozók, és Alakfelismerők Társasága Képfeldolgozók és Alakfelismerők Társaságának 12. Országos Konferenciája*, number 22, pages 1–10, 2019. 4)

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. 8)

[12] Fred Shaffer and J. P. Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5, September 2017. 17)

[13] Phyllis K Stein, Matthew S Bosner, Robert E Kleiger, and Brooke M Conger. Heart rate variability: a measure of cardiac autonomic tone. *American heart journal*, 127(5):1376–1381, 1994. 17)

[14] Yue Sun, Wenjin Wang, Xi Long, Mohammed Meftah, Tao Tan, Caifeng Shan, Ronald M. Aarts, and Peter H. N. de With. Respiration monitoring for premature neonates in nicu. *Applied Sciences*, 9(23), 2019. 4)

[15] Yue Sun, Wenjin Wang, Xi Long, Mohammed Meftah, Tao Tan, Caifeng Shan, Ronald M Aarts, et al. Respiration monitoring for premature neonates in nicu. *Applied Sciences*, 9(23):5246, 2019. 11)

[16] Michael E. Tipping and Christopher M. Bishop. Mixtures of Probabilistic Principal Component Analyzers. *Neural Computation*, 11(2):443–482, 02 1999. 13)

[17] Inga Warren, Ezam Mat-Ali, Mark Green, and Dumisani Nyathi. Evaluation of the family and infant neurodevelopmental education (fine) programme in the uk. *Journal of Neonatal Nursing*, 25:93–98, Nov. 2018. 2, 4)

[18] Philippe Weinzaepfel, Jerome Revaud, Zaid Harchaoui, and Cordelia Schmid. Deepflow: Large displacement optical flow with deep matching. *2013 IEEE International Conference on Computer Vision*, 2013. 11)

[19] Jan Werth, Louis Atallah, Peter Andriessen, Xi Long, Elly Zwartkruis-Pelgrim, and Ronald M. Aarts. Unobtrusive sleep state measurements in preterm infants – a review. *Sleep Medicine Reviews*, 32:109–122, 2017. 4)

[20] Björn Westrup. Newborn individualized developmental care and assessment program (nidcap): Family-centered developmentally supportive care. *Early human development*, 83:443–9, 08 2007. 4)