Comparative Analysis of Motion-Based Algorithms for Estimating Infant Breathing Rates From an RGB-Camera

Demetra Carata-Dejoianu[1]
Supervisors: Jorge Martinez Castaneda[1], Kianoush Rassels[1]
[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
January 26, 2025

Name of the student: Demetra Carata-Dejoianu
Final project course: CSE3000 Research Project
Thesis committee: Jorge Martinez Castaneda, Kianoush Rassels, Cristoph Lofi

An electronic version of this thesis is available at http://repository.tudelft.nl/.

***Abstract*—Respiratory Rate (RR) is a vital health indicator, especially in infant monitoring, where early detection of abnormalities or variabilities in RR is crucial. Traditionally, the respiratory rate is extracted using contact-based methods, which, although reliable, can be quite intrusive and stressful for long-term monitoring. This study explores the potential of real-time remote RR monitoring on inexpensive hardware, by comparing three motion-based methods of extracting RR from RGB-camera feed: Pixel Intensity Changes (PIC), Optical Flow (OF), and Eulerian Video Magnification (EVM). The three algorithms were benchmarked using the public AIR-125 dataset, which features videos of infants in various positions, with a focus on their accuracy and computational intensity. The results show that the PIC algorithm slightly outperformed the other two algorithms in both accuracy and computational complexity. However, none of the algorithms managed to replicate the performance of the study which initially proposed the dataset as a benchmark.**

***Index Terms*—Respiratory rate, Motion-based algorithms, Infant monitoring, RGB-Camera, Pixel Intensity Changes, Optical Flow, Eulerian Video Magnification**

I. Introduction

Vital signs such as heart rate, oxygen saturation, and respiratory rate are important indicators of the health status of a person [1]. Changes in the respiratory rate (RR) are particularly indicative of life-threatening conditions [2], yet RR remains a sign that is often neglected in medical and sports environments [3], [4]. In cases like monitoring infants in the NICU, it is of crucial importance to intervene rapidly when abnormalities in breathing rate are detected to ensure optimal health outcomes [5].

Traditionally, RR is measured in hospitals either manually or with instruments such as capnographs, pulse oximeters and ECG-based monitoring systems. Although reliable, these require sensors attached to the skin, like nasal cannulas, chest belts or adhesive electrodes. Such contact-based methods can cause irritation and great discomfort, especially if applied for longer times. When it comes to infants, their sensitive skin can make this an even bigger problem. Furthermore, the presence of these sensors might interfere with the natural movement of babies, making them impractical in long-term monitoring in clinical or home settings. For these reasons, the possibility of measuring RR in a non-invasive, remote way has garnered increased interest.

The many physiological markers associated with respiration (chest movement, temperature changes in the air, blood irrigation) allow for experiments in extracting RR using a wide range of sensors. Methods involving simple RGB cameras have been studied extensively, due to the reduced cost of the needed equipment.

However, while methods of extracting RR from videos are continuously being developed, they are often bench-marked against datasets not available publicly, or on data collected in clinical, restricted conditions. This makes it difficult to draw a comprehensive comparison between the existing methods, especially when it comes to practical use scenarios like home monitoring.

The present paper aims to fill in that research gap and compare the performance of three motion-based RR estimation algorithms in the practical scenario of monitoring sleeping infants. A strong emphasis is put on the real-life and real-time aspects, meaning that the algorithms were tested on unrestricted subjects, in different positions, and were designed to output RR as quickly as possible. As such, this paper represents not only a comparison between the algorithms, but also a feasibility study. A brief overview of the types of algorithms considered for this task is presented in the section Background and Algorithm Selection.

To draw a conclusion about their performance, the proposed methods have been evaluated based on the following comparison criteria: accuracy and computational complexity. Given the restrictive condition of running real-time, these metrics have been chosen as an attempt to balance how fast an algorithm can be, while also outputting relevant data. Accuracy is measured in multiple ways: Root Mean Squared Error (RMSE) of the extracted RR compared to the ground truth RR, as well as the Pearson's correlation coefficient (ρ). The RMSE is mean to indicate how big the errors the algorithm makes are, while ρ measures whether the changes in RR are detected at the proper time. Mean Phase Coherence, an additional metric for measuring the quality of the signal, was also introduced to see how close the extracted respiratory signal is to the ground truth signal. This was done as only the working dataset only provided the respiratory signal, not ground truth values for the RR throughout time. Furthermore, computational complexity is measured in processing time per frame, as well as the CPU load. These metrics are especially relevant in the eventual deployment on embedded platforms, as the algorithms would have to run on less powerful processors. The results have been computed against the same pre-existing public benchmark dataset, the AIR-125 dataset [6], consisting of videos of sleeping infant subjects in different positions, manually annotated with information about the respiratory signal. Conditions such as differences in lighting or position are also taken into account when discussing the performance of the algorithms.

II. Background and Algorithm Selection

Remote vital sign extraction has been an active area of research for a few decades. Early work mostly focused on using Photophletysmography (PPG) to extract heart rate remotely [7], [8]. This technique, which measured the level of blood oxygenation using the subtle changes in skin colour, was later successfully applied for calculating the respiratory rate [9], [10]. There is, however, a variety of approaches that have been explored for remotely measuring RR, leveraging different types of sensors and underlying principles [11], [12], [13]:

1) **Thermography** - Studies focused on thermography have successfully measured RR by analysing the flow of hot air in the nasal or mouth area [14], [15]. Although effective, this method requires specialised sensors or thermal cameras, which restricts its practical uses.

2) **Acoustic Microphones** - Microphones have been used to determine RR by capturing respiratory sounds [16],

[17], [18]. However, this technique is susceptible to environmental noise, and therefore not suitable for all settings.

3) **Depth Cameras/Sensors** - By using cameras fitted with depth sensors, researchers were able to infer the respiratory rate by measuring the vertical movements of the chest [19], [20], [21], [22].

4) **Vibrometry** - Vibrometric methods utilise the subtle vibrations caused by inhaling and exhaling to determine the respiratory rate [23], [24], [25]. While accurate, they require specialised sensors, which can be quite costly.

5) **RGB-Cameras** - The prospect of using RGB cameras for measuring respiratory rate has garnered increased interest, as hardware is relatively affordable and widely-accessible compared to other sensors. (A simple standard definition (720p) camera can be purchased for under 100 USD by any user.) In literature, two main paradigms can be observed when it comes to extracting RR this way:
   - PPG-based algorithms - As mentioned above, these take advantage of the subtle changes in the colour of the skin caused by the blood flow associated with respiration [9], [10]. However, given the very subtle differences that they aim to detect, they are quite sensitive to noise, and tend to not perform that well in non-standardised conditions.
   - Motion-based algorithms - These examine movements in areas such as the chest or the pit of the neck to measure respiratory rate [26], [27], [28]. They are prevalent in literature, and they are more robust to environmental changes than the PPG-based methods. For this reason, they will be the focus of this paper.

In the realm of motion-based RR extraction, an entire spectrum of algorithms has been developed, relying on different computational techniques. In a review [12], Massaroni et al. classified the algorithms in three main categories: Pixel Intensity Changes, Optical Flow, and Motion Magnification.

Algorithms that focus on computing the changes in the pixel intensities have been studied in-depth by [28], [29], as well as [30], [31].

Optical Flow has been studied by [32] as well [33], and [27].

Finally, the reportedly more robust, but more computationally complex method of Eulerian Motion Magnification has appeared in the works of [34], [26], [35], as well as [36].

Recent studies have built upon the foundation of these algorithms and introduced deep-learning based methods of measuring RR [37], [38], [39]. Although showing promise, methods involving Convolutional Neural Networks have not yet been shown to add improvements in performance significant enough to justify the increased computational burden [38]. For this reason, such methods are outside the scope of this research. Instead, the main focus will be on lightweight implementations of the motion-based algorithms, with an accent on trying to balance computational efficiency and reliability, a crucial aspect for real-time applications.

Although a wide range of variations of these main algorithms have been proposed, direct comparisons - especially in real-life contexts - remain limited. As such, this paper strives to perform a comparative analysis of the methods of Pixel Intensity Changes(PIC), Optical Flow(OF), and Eulerian Motion Magnification(EVM), and understand their applicability in real-time, real-context respiratory rate measurement.

III. Methodology

The entirety of the project was realised in Python3 [40], with the use of the OpenCV [41] library for processing the video. Additionally, the SciPy [42] and Matplotlib [43] libraries were used for processing the extracted signal, respectively for plotting the graphs. The choice of Python as a programming language was made because of its versatility and for its extensive support regarding image processing libraries. Furthermore, in the eventual deployment on embedded systems, the cross-platform compatibility of Python would ease the process. Despite other languages such as C++ offering better performance for CPU-intensive tasks, the performance of optimised Python code was deemed sufficient for the scope of this research.

The algorithms were implemented and tested on the same machine, an upper mid-range laptop, described in more detail in the Results section.

For the experiment, a comparison pipeline was set up, which can be visualised in fig. 1. A more in-depth description of each algorithm can be found in their respective subsections. Preprocessing and RR calculation from the signal are the same for all three algorithms, so the algorithms are only assessed on the quality of the motion signal they extract.

Since the RR estimation is meant to be done in real-time, a sliding window was introduced. 10 initial seconds for calibration are allowed, time in which only the movement signal is recorded, with no calculations regarding the RR. A window smaller than 10 seconds would not be sufficient to detect enough inhaling-exhaling cycles for the calculation to be relevant. Introducing the calibration period does cause a slight delay in outputting the first reading, but it is short enough that it should not affect practical use. After calibration, the signal is processed to obtain the RR. The estimation is updated every 1 second, and is based on the last 10 seconds of the video feed.

A. ROI Selection

The first step in the processing pipeline after extracting the video feed is choosing the area of the image to focus on. Picking a suitable Region of Interest (ROI) is crucial to the accuracy, as certain areas of the body, such as the face or the chest, visibly change when inhaling and exhaling. Furthermore, restricting the area of calculations can greatly improve performance. In this study, all algorithms use the chest and abdomen area as a ROI, since babies have been shown to be particularly abdominal breathers [44].

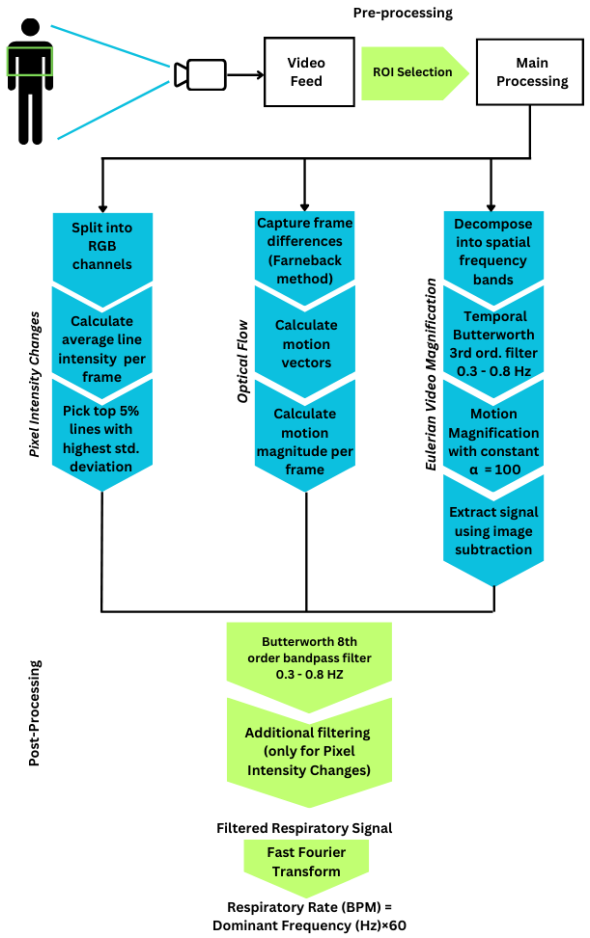

Fig. 1: Workflow for comparing the three algorithms

For this experiment, the ROI was selected manually before running the algorithms. Automatic selection and tracking of the ROI were considered, but were ultimately deemed unrealistic in the given time frame, considering how varied the positions of the subjects would be in a real-life context. The ROI selection screen can be seen in fig. 2.

B. Pixel Intensity Changes

The Pixel Intensity Changes algorithm quantifies the differences in individual pixels across the video sequence. The code used in this research is based on studies by Massaroni et al. [28], with slight changes when it comes to filtering out the noise.

The first step in processing is splitting the video frame into its red, green, and blue components. Each of the channels is analysed individually in order to obtain the intensity of the pixels for each line. The formula used is:

$$I_{\text{line}} = \frac{\sum_{j=1}^{N}(R_j + G_j + B_j)}{N}$$

where I_{line} represents the intensity of a row of pixels in the frame, N represents the number of pixels in the frame, and

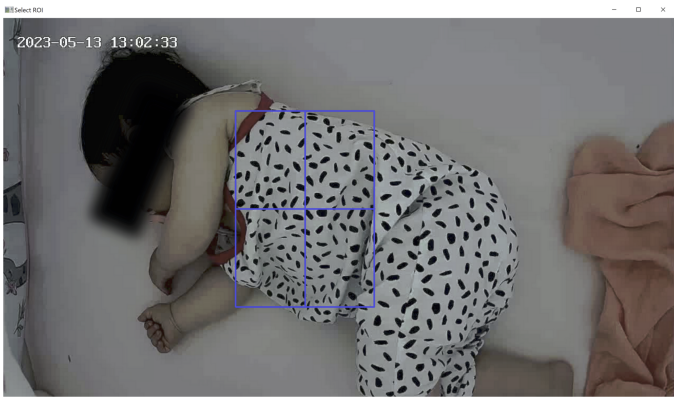

Fig. 2: ROI Selection on a video of Subject 5, shown at the start of any algorithm. The blue rectangle represents the ROI selected by the user. The video was taken from the AIR-125 dataset.

R_j, G_j and B_j represent the intensity of the jth pixel of the row in the red, green, and blue channels.

An array for the I_{line} values is stored for each frame. In order to produce the motion signal, the standard deviation of the rows' intensities is calculated across the 10-second window. The top 5% rows with the highest standard deviation are chosen, as these are the most likely to contain data regarding movement. The means of these rows' intensities represent the motion signal.

Given that this algorithm analyses pixel intensities, it is prone to noise caused by sudden movements or glitches in the video. To combat this, an extra mechanism has been introduced, which drops the frames where too much sudden movement is detected. This is done by keeping an array of the absolute motion differences between frames. A motion threshold is calculated with formula $T = \mu_{\text{d}} + k \cdot \sigma_{\text{d}}$, where d is the array of differences, and k is a constant with the value of 3, determined experimentally. Any frame that has an absolute difference compared to the previous frame above T is dropped.

C. Optical Flow

Optical flow is a method of quantifying movement in a video by analysing the displacement of pixels across multiple frames [45]. It was first introduced as a concept by James J. Gibson in 1950 [46], but only adopted in computer vision decades later, in works such as the one of Horn and Schunck [47].

The method used for calculating the optical flow in this research is the Farneback method [48], available in the OpenCV library. This has been chosen because, as opposed to frameworks like the Lucas-Kanade method [49], [50], which focuses on certain feature points, it tracks displacement across the entire image. Since the selected ROI, the chest area, does not present any distinctive points to track, this method seemed favourable. However, calculating the displacement using this dense flow method increases the computational cost. Deep-learning based optical flow frameworks like FlowNet [51] and PWC-Net [52] were not considered, also for the reason of high computational intensity.

The magnitude of the motion is calculated both on the horizontal and vertical vectors of movement, as the positions

of the infants in the videos vary. The formula for calculating the motion magnitude is therefore $M = \sqrt{f_x^2 + f_y^2}$, where f_x is the horizontal motion vector and f_y is the vertical motion vector. These changes in movement intensity represent the outputted signal.

D. Eulerian Video Magnification

Eulerian Video Magnification is a method of enhancing particular movements in a video, which was introduced in 2012 by Wu et al. [53]. It is suitable for the task of detecting respiratory movements as it can be focused precisely on the frequencies at which respiration occurs.

In the tested version of the algorithm, the video is first converted to greyscale, to reduce the complexity. After this, the video is decomposed into Gaussian pyramids, a series of images where each image is a progressively downscaled version of the original frame. For this experiment, the number of pyramid levels was 3. This allows the algorithm to focus on different levels of spatial details, as the motion at different scales could have different frequencies. Each of the levels of the pyramids goes through a Butterworth bandpass filter of level 3 between 0.3-0.8 Hz. The motion is then amplified by adding back the filtered signal to the original frame, multiplied by a constant α, which has been chosen as 100 experimentally. This makes the motion in the video more noticeable.

Following this process, the respiratory movement has been enhanced, making it easier to extract the signal. This extraction is performed using the frame differences method. This is a basic video processing method, which analyses the pixel-wise differences between two consecutive frames, and then appends that value to the motion signal.

E. Post-Processing

The three algorithms all produce the same output: a signal representing the intensity of the movement of the ROI at a certain time. In order to extract the respiratory rate in Breaths Per Minute (BPM), there are a few steps left in post-processing. First, to remove the noise caused by unwanted movements, a Butterworth bandpass filter of order 8 between 0.3 and 0.8 Hz is applied to the signal. These frequencies correspond to 18-48 BPM, in the range of normal breathing for an infant [54].

Additional to the bandpass filter, the PIC algorithm has two more filters in place, as the algorithm was quite noisy. First, after the bandpass filter, the signal is normalised by subtracting the mean and dividing by the standard deviation. Then, an exponential moving average filter with $\alpha = 0.4$ is applied to the window. This reduces the number of small peaks caused by unwanted movement or problems with the video.

After the filtering, a Fast Fourier Transform is applied to the resulting signal, to determine the dominant frequency in Hertz (Hz), f. The respiratory rate is then calculated with the formula $RR = f \times 60$.

IV. Ethical Considerations

When performing studies in such a sensitive field as healthcare, paying attention to the ethical aspects is crucial. Throughout this project, the principles of Honesty, Scrupulousness, Transparency, Independence, and Responsibility, extracted from The Netherlands Code of Conduct for Research Integrity 2018 [55], were used as guidelines. In the following subsections, two of the most important ethical considerations of this project are detailed.

A. Reproducibility

Multiple steps have been taken to ensure transparency and reproducibility of the results. Firstly, the algorithms that have been studied were developed based on peer-reviewed papers, cited in the bibliography. Secondly, the program code used for the experiments is open-source and available in its entirety on GitHub [56]. All functions that were not implemented by the author are part of open libraries. Lastly, the dataset that the results are based on is publicly available. Therefore, a third party should easily be able to reproduce the results of this study.

B. Data Sensitivity

Health indicators such as respiratory rate are considered sensitive personal data under the regulatory framework of the European Commission [57]. Their collection, processing, and storage are subject to strict legal and ethical requirements to ensure privacy and security. This becomes even more imperative when it comes to data collected from vulnerable subjects such as infants, as they are unable to provide informed consent. To address these concerns, several precautions have been taken. First of all, the working dataset was acquired from an open scientific source, having been previously used in performing research [6]. The dataset was published with the intention of a public benchmark, and contains videos collected by the Northeastern University clinical team, under the ethical approval of their Institutional Review Board. A Data Management Plan regarding the chosen dataset has also been sent for approval to the Data Stewards of TU Delft. At the time of writing, the approval was still pending. In the dataset, videos of three of the subjects had been collected from YouTube. However, due to uncertainties regarding the source and consent of these videos, they have been discarded from the present study.

Secondly, besides the videos and the ground truth values of the respiration, the dataset does not feature any information regarding the identity of the subjects. As an additional measure, the faces of the subjects used as examples in this paper have been censored, to further protect their anonymity.

Lastly, all aggregated data derived from the dataset was stored and processed in a secure environment, and will not be shared further than the scope of this study.

V. Results

The results of the study were obtained by comparing the three algorithms against the AIR-125 dataset, based on their accuracy and computational complexity. The dataset originally consisted of 125 videos of 8 subjects, but due to aforementioned ethical concerns, only the videos from the first 5

subjects were used, resulting in a working dataset of 99 videos. The resolution of the videos was fixed at 1280×720 pixels, while the refresh rate was either 10 or 15 fps. Each algorithm was run once on each of the videos.

For measuring accuracy, RMSE and Pearson's correlation coefficient (ρ) were used, both comparing the extracted RR to the one of the ground truth files. However, it is worth mentioning that the ground truth RR also had to be computed, as the dataset only contained the respiratory impulse as a signal. Given this circumstance, Mean Phase Coherence (R) was introduced as an additional metric to compare the extracted signal after filtering to the ground truth respiratory impulse. This was inspired by phase coherence, which has been used to measure the synchronization rate between oscillating signals [58]. Given that the amplitude of the signals is not relevant, but their oscillation rate is, this seemed like a suitable way to measure their correlation. The metric is computed by averaging the phase coherence between the two signals calculated every 10 seconds, when the sliding window changes entirely. The Mean Phase Coherence can be described by the following formula:

$$R = \left| \frac{1}{T} \sum_{t=1}^{T} e^{j\Delta\phi(t)} \right|,$$

where: T is the total number of time samples, $\Delta\phi(t) = \phi_x(t) - \phi_y(t)$ is the phase difference between the two signals at time t, and $e^{j\Delta\phi(t)}$ represents a unit vector on the complex plane for the phase difference. The value of R can be between 0 and 1, with 0 meaning that the signals are not correlated at all, and 1 meaning that they are perfectly synchronised.

As for measuring computational intensity, the chosen metrics were CPU Load and processing time per frame.

A. Accuracy

The results of the three algorithms are comparable in terms of accuracy, with PIC having a slight advantage over the others. Table I represents the overall performance of the three algorithms, as well as the performance of the AirFlowNet method, which was introduced in the same paper as AIR-125 [6], and the performance of another motion-based method tested on the dataset [59]. The results show that all of the algorithms were significantly outperformed by the AirFlowNet method. The method proposed by Guo et al., while still outperforming the algorithms of this paper, also obtained much lower scores than AirFlowNet. It is worth mentioning that both of these methods are based on convolutional neural networks, and were both trained and tested on the AIR-125 dataset.

TABLE I: RMSE, Pearson's Correlation Coefficient, and Mean Phase Coherence for the algorithms

Method	**RMSE**	ρ	R
PIC	7.78	0.13	0.40
OF	7.90	0.01	0.37
EVM	7.59	0.02	0.39
Guo et al. [59]	6.74	0.32	-
AirFlowNet. [6]	5.40	0.72	-

Although the ρ values for all algorithms are small, the higher R values indicate that they are able to extract a motion signal at least similar to the ground truth. In fig. 3, fig. 4, and fig. 5 the respiratory signal extracted by each of the algorithms on the same subject in the same window, can be seen.

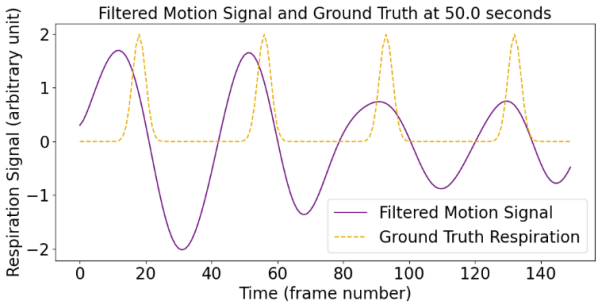

Fig. 3: Motion Signal extracted using PIC - Subject 5, Video 4

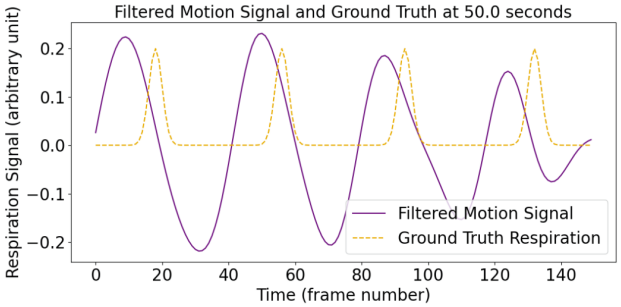

Fig. 4: Motion signal extracted using OF - Subject 5, Video 4

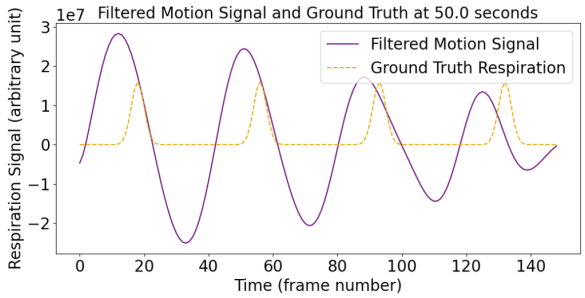

Fig. 5: Motion signal extracted using EVM - Subject 5, Video 4

In table II and table III, the performance of the algorithms can be seen in different conditions. From the first table, it can be observed that videos with the subject laying on their back produced better results than those with them laying on their stomach or the side. Seeing that this position allows for a better view of the entirety of the thoracic and abdominal area, the results are not surprising. As for the colour of the video, in the case of OF and EVM, the results are inconclusive. However, for PIC, the coloured videos produced better results as opposed to the greyscale ones. Again, this is unsuprising, considering that the first two algorithms convert the video to black-and-white before processing, while PIC analyses the pixel intensities on the RGB channels individually.

TABLE II: Comparison of metrics by position

Position	Algorithm	RMSE	ρ	R
Back	PIC	0.12	7.02	0.42
	OF	0.08	5.32	0.43
	EVM	0.03	5.73	0.43
Stomach	PIC	0.13	7.32	0.40
	OF	-0.03	6.58	0.36
	EVM	0.05	6.86	0.36
Side	PIC	0.13	8.35	0.39
	OF	0.01	9.68	0.35
	EVM	0.00	8.76	0.38

TABLE III: Comparison of metrics by video colour

Position	Algorithm	RMSE	ρ	R
Coloured	PIC	0.18	7.64	0.45
	OF	-0.03	7.58	0.43
	EVM	-0.04	7.54	0.44
Greyscale	PIC	0.11	7.83	0.38
	OF	0.03	8.00	0.35
	EVM	0.03	7.62	0.36

B. Computational Complexity and Number Crunching

The experiment was run on a machine equipped with an 8-core Intel(R) Core(TM) i7-10870H CPU running at 2.20GHz, with 32 GB of RAM. This is an upper-mid range laptop processor, so it is expected that the algorithms would perform poorer on an embedded platform. Examples of the computational intensities for each of the algorithms, compared on the same video, can be seen in fig. 6, fig. 7, and fig. 8. The CPU load percentage represents the mean load over the 8 processors, to get an estimate of how much of the total available power is being used. In the graphs for the frame processing times, the threshold is calculated with the formula $\frac{1}{fps}$ second, representing the maximum time a frame can be processed to maintain real-time estimation.

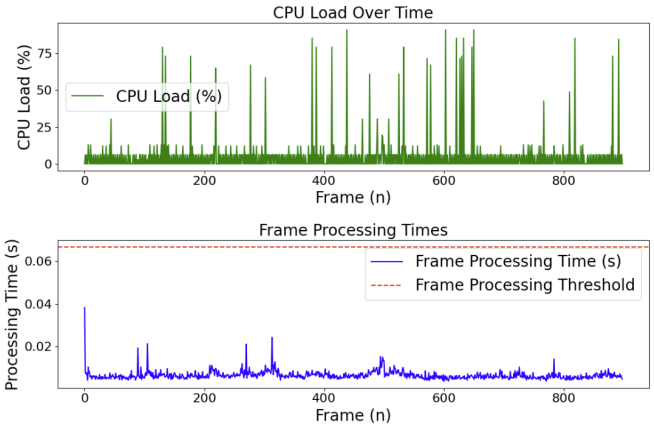

Fig. 6: Computational Complexity for PIC - Subject 1, Video 4

On average, the Pixel Intensity Changes algorithm was the fastest, while the Eulerian Video Magnification algorithm was the slowest. An overview of the frame processing times depending on the FPS of the video can be seen in fig. 9. While in the case of PIC and OF this time remains almost constant,

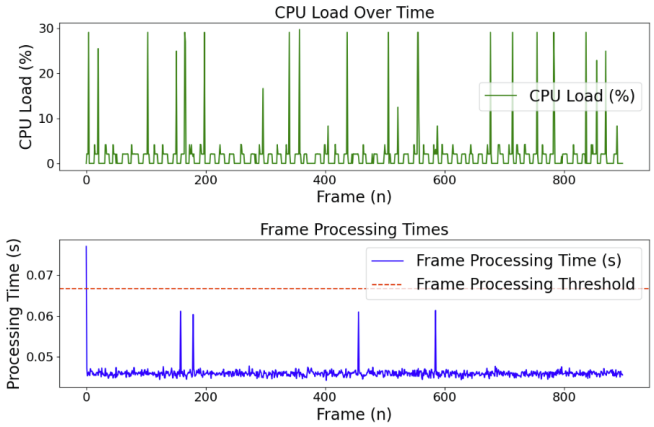

Fig. 7: Computational Complexity for OF - Subject 1, Video 4

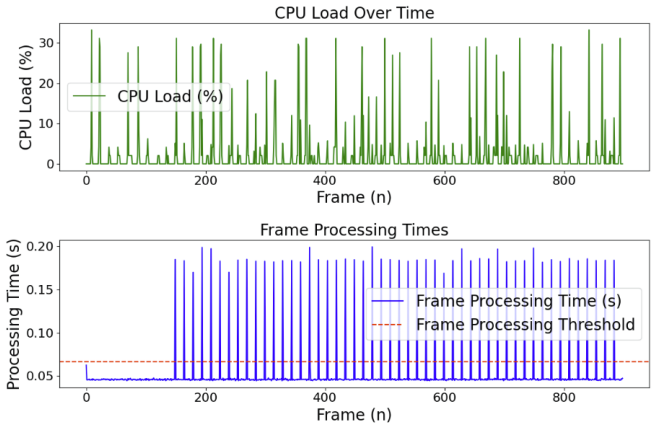

Fig. 8: Computational Complexity for EVM - Subject 1, Video 4

the processing time for EVM spikes occasionally. This is due to the fact that EVM processes the frames in chunks, rather than consecutively.

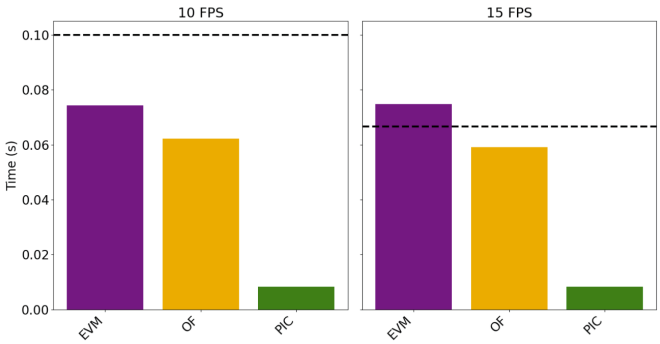

Fig. 9: Average Frame Processing Times for the algorithms. The black line represents the delay threshold.

When it comes to the average CPU load throughout processing, PIC required the most processing power. However, the value is still quite small, and considering that the PIC algorithm runs much faster than the other two, it can be said that the overall CPU Load is comparable for all three algorithms.

TABLE IV: Average CPU Load for the algorithms throughout video processing

	PIC	OF	EVM
Average CPU Load	6.42 %	2.79 %	2.37 %

VI. Discussion

A. Accuracy

The results show that PIC performed slightly better than the other two algorithms, yet still considerably below an optimal performance. While the R values of the algorithms suggest they are able to extract a motion signal slightly correlated to the real respiratory signal, the low ρ values indicate that all of the algorithms fail at properly calculating the RR from this motion signal. One of the reasons could be that the window size and the time between updates are quite short, causing the RR to oscillate quite a lot between readings. Nonetheless, the main reason remains the fact that the algorithms are not able to fully remove noise, causing the Fourier Transform to detect unwanted dominant frequencies when calculating the RR.

Considering the poor results of other methods on the same dataset, though, there is also the possibility that the data was too difficult to work with in the first place. The many different positions of the subjects in the dataset, as well as the varying lighting conditions represent a tough challenge for RR extracting methods, even for machine learning-based ones trained and tested on the dataset, like the Guo et al. method. This indicates that there remains research to be done before deploying remote RGB-camera based methods for RR extraction in such unrestricted scenarios.

B. Real-Time performance

From the experimental results, it can be seen that the Pixel Intensity Changes algorithm has the fastest performance, while the Eulerian Video Magnification algorithm performs the slowest. This was to be expected, given the individual complexities of each of the algorithms. However, for a real-life, real-time implementation, these running times need to be compared to an established standard. In this paper, the concept of "real-time" was defined based on the 10-second sliding windows - for an algorithm to run in real-time, it needed to be able to process a 10 second video every 1 second. From this standpoint, only the Pixel Intensity Changes and the Optical Flow performed well enough at both 10 fps and 15 fps. EVM was not much slower, also managing to stay under the threshold for videos running at 10 fps. Given that the frame processing time metric was taken as an average over all of the runs, there were also instances where OF and EVM exceeded the delay threshold, even at 10 fps. This happened in cases where the selected ROI was larger, as there were more pixels to be processed.

This model of real-time computation is however, quite restrictive, and an implementation where the RR is calculated every 10 seconds could also be considered real-time, as it would still allow for sufficient time to detect an abnormal breathing rate. In such a context, all algorithms could conceivably be modified to allow for real-time performance.

The machine that the experiment was run on is also of note. By using an upper mid-range laptop processor, it is expected that the results would be much more favourable than by using an embedded processor. In the case of the PIC algorithm, the processing time margin is large enough that porting the code to an embedded system should still allow for good performance. For the other two algorithms, more experiments are needed.

C. Limitations

In the process of this study, there were certain factors which affected the methods used, and subsequently the results.

Firstly, the selection of the algorithms to compare was a lengthy one, due to the vast amount of research done in this field in recent years. For each of the three methods presented in this study, there are variations in the parameters and filtering methods used. For the current comparison, the algorithms were applied based on well-established implementation mentioned in previous research. Therefore, other, newly-developed implementations may perform better than the presented methods.

Secondly, the datasets available for health-related research are scarce, even more so when it comes to infant subjects. The chosen dataset was deemed the most compatible, given that it has been used in similar research, yet it only featured a small number of subjects, which could have an impact on the reproducibility of the results on other datasets. The quality of the videos also varied throughout the dataset, with some of the samples featuring periodic glitches - this significantly affected some of the calculations. Furthermore, the dataset was only annotated with the respiratory impulse, not the calculated respiratory rate. This meant that extra calculations had to be done to determine the ground truth values, which introduced a level of uncertainty in the results. Given the strict time constraints, collecting new data for this research was impossible.

Lastly, the ROI had to be selected manually. This meant that the results for the same video could differ slightly, based on the selected area. In future research, this could be addressed by implementing automatic ROI detection, for example by using machine learning.

VII. Conclusion And Future Work

The poor results obtained by all three of the algorithms suggest that classical video-based RR extraction methods are not ready for deployment in real-life scenarios, where movement and lighting conditions may vary greatly. These findings highlight the need for further refinement and optimisation to improve robustness and reliability in diverse settings.

Having achieved better results in both accuracy and complexity than the other two, the PIC algorithm shows more promise. Before such a method of extracting RR is implemented in real-life scenarios, though, there remain many steps to be taken for improving its performance. Signal filtering needs to be further refined to ensure that the extracted signal is as close as possible to the real respiratory motion. For this, more experimental work is needed, to investigate what filter chain yields the best results. Another step that could be

taken in improving accuracy is a more sophisticated method of detecting unwanted movement.

Another aspect which remains to be studied is the optimal window size for performing the calculations. The results of this study suggest that the 10 seconds sliding window might have been too short of an interval to smoothly detect RR. Updating the calculated value less often than once every second could also cause less fluctuation in the RR estimations, bringing the results closer to the real-life value.

Implementing an automatic way of selecting the ROI would have a significant impact on reliability, seeing that currently the manually selected ROI can vary. Tracking the ROI would also represent an important step, which would enable focus on the same area even in the case of movement. For tasks like this, neural network based methods such as YOLOv3 [60] have shown promise, although they introduce great computational requirements.

As for the computational complexity, multiple optimisations could be made to ensure that the algorithms run on an embedded platform as well. Firstly, downscaling the videos to 480p would reduce the number of pixels by 55.5 %, which would significantly improve the performance, while still remaining clear enough to detect the movement changes. Secondly, a programming language with better memory management capacities, such as C++, could be used. Lastly, other techniques such as GPU acceleration or multithreading could be implemented, depending on the hardware capabilities. This would allow for more video frames to be processed in parallel, reducing the needed computation time.

In conclusion, while all three of the algorithms performed sub-optimally, the PIC algorithm does hold potential for improvement in real-time, real-life applications. Filtering methods, as well as other optimisations remain to be studied for the future development of such algorithms. These efforts could bring video-based RR extraction closer to practical real-world application.

REFERENCES

[1] A. Sapra, A. Malik, . Priyanka, and B. Affiliations, "Vital sign assessment," 2023.

[2] C. P. Subbe, R. G. Davies, E. Williams, P. Rutherford, and L. Gemmell, "Effect of introducing the modified early warning score on clinical outcomes, cardio-pulmonary arrests and intensive care utilisation in acute medical admissions," *Anaesthesia*, vol. 58, pp. 797–802, 8 2003.

[3] C. M. A, "Respiratory rate: The neglected vital sign," *Medical Journal of Australia*, vol. 188, pp. 657–659, 6 2008.

[4] A. Nicolò, C. Massaroni, and L. Passfield, "Respiratory frequency during exercise: The neglected physiological measure," *Frontiers in Physiology*, vol. 8, 12 2017.

[5] M. O. Edwards, S. J. Kotecha, and S. Kotecha, "Respiratory distress of the term newborn infant," pp. 29–37, 3 2013.

[6] S. K. R. Manne, S. Zhu, S. Ostadabbas, and M. Wan, "Automatic infant respiration estimation from video: A deep flow-based algorithm and a novel public benchmark," in *Perinatal, Preterm and Paediatric Image Analysis*, D. Link-Sourani, E. Abaci Turk, C. Macgowan, J. Hutter, A. Melbourne, and R. Licandro, Eds. Cham: Springer Nature Switzerland, 2023, pp. 111–120.

[7] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics Express*, vol. 16, no. 26, pp. 21 434–21 445, 2008. [Online]. Available: https://doi.org/10.1364/oe.16.021434

[8] 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*, vol. 18, pp. 10 762–74, 5 2010.

[9] C. Wang, Z. Li, and X. Wei, "Monitoring heart and respiratory rates at radial artery based on ppg," *Optik*, vol. 124, pp. 3954–3956, 10 2013.

[10] Y. D. Lin, Y. H. Chien, and Y. S. Chen, "Wavelet-based embedded algorithm for respiratory rate estimation from ppg signal," *Biomedical Signal Processing and Control*, vol. 36, pp. 138–145, 7 2017.

[11] L. Maurya, P. Kaur, D. Chawla, and P. Mahapatra, "Non-contact breathing rate monitoring in newborns: A review," 5 2021.

[12] C. Massaroni, A. Nicolo, M. Sacchetti, and E. Schena, "Contactless methods for measuring respiratory rate: A review," pp. 12 821–12 839, 6 2021.

[13] W. Wang and A. C. D. Brinker, "Algorithmic insights of camera-based respiratory motion extraction," *Physiological Measurement*, vol. 43, 7 2022.

[14] R. Murthy, I. Pavlidis, and P. Tsiamyrtzis, "Touchless monitoring of breathing function," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, 2004, pp. 1196–1199.

[15] T. Negishi, S. Abe, T. Matsui, H. Liu, M. Kurosawa, T. Kirimoto, and G. Sun, "Contactless vital signs measurement system using rgb-thermal image sensors and its clinical screening test on patients with seasonal influenza," *Sensors*, vol. 20, no. 8, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/8/2171

[16] A. Azarbarzin and Z. M. K. Moussavi, "Automatic and unsupervised snore sound extraction from respiratory sound signals," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 5, pp. 1156–1162, 2011.

[17] E. C. Larson, M. Goel, G. Boriello, S. Heltshe, M. Rosenfeld, and S. N. Patel, "Spirosmart: using a microphone to measure lung function on a mobile phone," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ser. UbiComp '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 280–289. [Online]. Available: https://doi.org/10.1145/2370216.2370261

[18] Y. Nam, B. A. Reyes, and K. H. Chon, "Estimation of respiratory rates using the built-in microphone of a smartphone or headset," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 6, pp. 1493–1501, 2016.

[19] E. A. Bernal, L. K. Mestha, and E. Shilla, "Non contact monitoring of respiratory function via depth sensing," in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2014, pp. 101–104.

[20] S. Kumagai, Y. Takahashi, M. Hasegawa, Y. Fujita, K. Nakamura, and T. Kubo, "Markerless respiratory motion tracking using single depth camera," *Open Journal of Medical Imaging*, vol. 6, no. 1, p. 20, 2016. [Online]. Available: https://doi.org/10.4236/ojmi.2016.61003

[21] M. Martinez and R. Stiefelhagen, "Breathing rate monitoring during sleep from a depth camera under real-life conditions," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 3 2017, pp. 1168–1176. [Online]. Available: http://ieeexplore.ieee.org/document/7926718/

[22] P. S. Addison, A. Antunes, D. Montgomery, P. Smit, and U. R. Borg, "Robust non-contact monitoring of respiratory rate using a depth camera," *Journal of Clinical Monitoring and Computing*, vol. 37, pp. 1003–1010, 8 2023.

[23] L. Scalise, I. Ercoli, P. Marchionni, and E. P. Tomasini, "Measurement of respiration rate in preterm infants by laser doppler vibrometry," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, May 2011, pp. 657–661.

[24] H. Aygun and A. Apolskis, "Tracheal sound acquisition using laser doppler vibrometer," in *ACOUSTICS 2018 - ACWSTEG 2018*. Cardiff, City Hall, Cathay's Park: London South Bank University, April 2018, pp. 23–24, 23 - 24 Apr 2018.

[25] K. Kroschel and J. Metzler, "Contactless measurement of the respiration frequency by vibrometry," in *Proc. Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung*, 2018, pp. 310–317.

[26] A. Al-Naji and J. Chahl, "Remote respiratory monitoring system based on developing motion magnification technique," *Biomedical Signal Processing and Control*, vol. 29, pp. 1–10, 8 2016.

[27] Y. Sun, W. Wang, X. Long, M. Meftah, T. Tan, C. Shan, R. M. Aarts, and P. H. de With, "Respiration monitoring for premature neonates in nicu," *Applied Sciences (Switzerland)*, vol. 9, 12 2019.

[28] C. Massaroni, D. L. Presti, D. Formica, S. Silvestri, and E. Schena, "Non-contact monitoring of breathing pattern and respiratory rate via rgb signal measurement," *Sensors (Switzerland)*, vol. 19, 6 2019.

[29] C. Massaroni, A. Nicolo, M. Sacchetti, and E. Schena, "Contactless methods for measuring respiratory rate: A review," pp. 12 821–12 839, 6 2021.

[30] B. A. Reyes, N. Reljin, Y. Kong, Y. Nam, S. Ha, and K. H. Chon, "Towards the development of a mobile phonopneumogram: Automatic

breath-phase classification using smartphones," *Annals of Biomedical Engineering*, vol. 44, pp. 2746–2759, 9 2016.

[31] B. A. Reyes, N. Reljin, Y. Kong, Y. Nam, and K. H. Chon, "Tidal volume and instantaneous respiration rate estimation using a volumetric surrogate signal acquired via a smartphone camera," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, pp. 764–777, 5 2017.

[32] R. Janssen, W. Wang, A. Moço, and G. D. Haan, "Video-based respiration monitoring with automatic region of interest detection," *Physiological Measurement*, vol. 37, pp. 100–114, 12 2015.

[33] M. Mateu-Mateus, F. Guede-Fernández, V. Ferrer-Mileo, M. A. García-González, J. Ramos-Castro, and M. Fernández-Chimeno, "Comparison of video-based methods for respiration rhythm measurement," *Biomedical Signal Processing and Control*, vol. 51, pp. 138–147, 5 2019.

[34] Y. Zhang and F. Shang, "Noncontact extraction of breathing waveform," 2015.

[35] V. Mattioli, D. Alinovi, G. Ferrari, F. Pisani, and R. Raheli, "Motion magnification algorithms for video-based breathing monitoring," *Biomedical Signal Processing and Control*, vol. 86, 9 2023.

[36] S. Ahani, N. Niknafs, P. M. Lavoie, L. Holsti, and G. A. Dumont, "Video-based respiratory rate estimation for infants in the nicu," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 12, pp. 684–696, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10738847/

[37] J. Jorge, M. ViIllarroel, S. Chaichulee, K. McCormick, and L. Tarassenko, "Data fusion for improved camera-based detection of respiration in neonates." SPIE-Intl Soc Optical Eng, 2 2018, p. 36.

[38] Q. Zhan, J. Hu, Z. Yu, X. Li, and W. Wang, "Revisiting motion-based respiration measurement from videos," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2020-July. Institute of Electrical and Electronics Engineers Inc., 7 2020, pp. 5909–5912.

[39] S. K. R. Manne, S. Zhu, S. Ostadabbas, and M. Wan, "Automatic infant respiration estimation from video: A deep flow-based algorithm and a novel public benchmark," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14246 LNCS. Springer Science and Business Media Deutschland GmbH, 2023, pp. 111–120.

[40] P. S. Foundation, "Python: A programming language for general-purpose programming," 2023, version 3.x. [Online]. Available: https://www.python.org

[41] G. Bradski and A. Kaehler, "Opencv library: Open source computer vision," 2023, version 4.x. [Online]. Available: https://opencv.org

[42] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and S. . Contributors, "Scipy 1.0: Fundamental algorithms for scientific computing in python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[43] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[44] I. Coyne, F. Neil, and F. Timmins, "Essential skills," *Clinical Skills in Children's Nursing*, pp. 81–114, 2010.

[45] A. Alfarano, L. Maiano, L. Papa, and I. Amerini, "Estimating optical flow: A comprehensive review of the state of the art," *Computer Vision and Image Understanding*, vol. 249, p. 104160, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314224002418

[46] J. J. Gibson, *The Perception of the Visual World*. Houghton Mifflin, 1950.

[47] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1, pp. 185–203, 1981. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0004370281900242

[48] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," 2003. [Online]. Available: http://www.isy.liu.se/cvl/

[49] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of Imaging Understanding Workshop*. Pittsburgh, Pennsylvania: Carnegie-Mellon University, Computer Science Department, 1981, pp. 121–130.

[50] D. Patel and S. Upadhyay, "Optical flow measurement using lucas kanade method," *International Journal of Computer Applications*, vol. 61, pp. 6–10, 01 2013.

[51] A. Dosovitskiy, P. Fischer, E. Ilg, P. Häusser, C. Hazırbaş, V. Golkov, P. v.d. Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," in *IEEE International Conference on Computer Vision (ICCV)*, 2015. [Online]. Available: http://lmb.informatik.uni-freiburg.de/Publications/2015/DFIB15

[52] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume," 2018. [Online]. Available: https://arxiv.org/abs/1709.02371

[53] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.*, vol. 31, no. 4, Jul. 2012. [Online]. Available: https://doi.org/10.1145/2185520.2185561

[54] S. Fleming, M. Thompson, R. Stevens, C. Heneghan, A. Plüddemann, I. MacOnochie, L. Tarassenko, and D. 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*, vol. 377, pp. 1011–1018, 2011.

[55] A. of Universities in the Netherlands (VSNU), "The netherlands code of conduct for research integrity 2018," 2018, accessed: 2024-12-27. [Online]. Available: https://www.vsnu.nl/files/documents/Netherlands%20Code%20of%20Conduct%20for%20Research%20Integrity%202018.pdf

[56] D. C. Dejoianu, "Motion-based respiratory rate extraction algorithms," 2025, accessed: 2025-01-22. [Online]. Available: https://github.com/demicarata/motion_based_rr_estimation

[57] E. Union, "Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (general data protection regulation)," pp. 1–88, 2016. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj

[58] M. Rosenblum, A. Pikovsky, and J. Kurths, "Phase synchronization of chaotic oscillators," *Physical review letters*, vol. 76, pp. 1804–1807, 04 1996.

[59] T. Guo, Q. Lin, and J. Allebach, "Remote estimation of respiration rate by optical flow using convolutional neural networks," *Electronic Imaging*, vol. 33, no. 8, pp. 267–1–267–1, 2021. [Online]. Available: https://library-imaging-org.tudelft.idm.oclc.org/ei/articles/33/8/art00004

[60] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018. [Online]. Available: https://arxiv.org/abs/1804.02767