Respiratory Rate Estimation Based on Detected Mask Area in Thermal Images

Natalia Kowalczyk and Jacek Rumiński
Gdańsk University of Technology
Gabriela Narutowicza 11/12, 80-233 Gdansk, Poland
{natalia.kowalczyk, jacek.ruminski}@pg.edu.pl

Abstract

The popularity of non-contact methods of measuring vital signs, particularly respiratory rate, has increased during the SARS-COV-2 pandemic. Breathing parameters can be estimated by analysis of temperature changes observed in thermal images of nostrils or mouth regions. However, wearing virus-protection face masks prevents direct detection of such face regions. In this work, we propose to use an automatic mask detection approach to select pixels within a mask region as a source of respiration information allowing efficient estimation of respiratory signals. We performed experiments with two important types of virus protection masks, i.e., FFP2 (N95) and surgical masks, for subjects while sitting, slowly walking from a short distance toward a camera, and slowly walking with moderate head movements. Experiments conducted with the adapted YOLO model have shown that detection of the mask area on the face allows for higher SNR values and reduces error in respiratory rate estimation in all analyzed scenarios. The Mean Absolute Error for respiratory rate estimation was below 1 bpm for sitting subjects for all types of masks. The error for walking subjects was 1.21 bpm for an FFP2 mask and about 2.1 bpm for a surgical mask. In the presence of head movements, while walking, the MAE was below 1.39 bpm and less than 1 bpm when only one outlier was removed.

1. Introduction

Remote assessment of vital signs is important in medical diagnostics, especially when contact sensors present a risk, are expensive, or can not be used due to skin/body diseases or infections. Contactless measurement has become very important in recent years due to the SARS-COV-2 virus pandemic. Using cameras and computer vision methods to analyze recorded images allows computing many physiological changes potentially useful for clinical or non-clinical applications [26]. The popular techniques include pulse rate estimation from visible light camera recordings (e.g., [30] [23] [31] [6]) and respiratory rate estimation from thermal camera recordings (e.g., [9] [27] [34] [28] [38]).

During the pandemic, automatic assessment of vital signs has become problematic in some situations due to wearing virus-protective masks. In respiratory rate analysis, detection of the position of the masked face in thermal images, particularly the location of the nostrils, is a challenge. However, a detected facial mask can be a source of valuable, respiratory-related information due to the temperature changes on a mask's surface caused by breathing.

This work aims to extract respiratory signals from people wearing different virus protection masks while sitting or walking. A sitting person in front of a camera resembles a scenario of a patient encounter in a clinic or testing point. A person walking toward the camera resembles another situation, when a person slowly moves in a queue at a virus testing point, airport security checkpoint, etc.

In particular, our contributions are:

- We designed a method to extract the respiratory signals 1) from the automatically detected mask regions and, for comparison, 2) from the lower half of the automatically detected face region from thermal images of people with virus protection masks.
- We compared and analyzed the extracted respiratory signals and related respiratory rates for subjects wearing different types of masks: surgical and FFP2 (N95).
- We compared and analyzed the extracted respiratory signals and related respiratory rates for subjects wearing different types of masks while sitting, slowly walking from a short distance toward a camera, and slowly walking with moderate head movements.
- We have shown that YOLO-based facial mask detection in thermal images allows extracting the respiratory signals with a higher value of Signal-to-Noise Ratio and reduces respiratory rate estimation error in all analyzed scenarios.

The rest of the work is structured as follows. In the following section, related work is described. Section III

presents this study's methodology, experimental design, and dataset. Results are shown in Section IV, followed by a discussion of results in Section V. Finally, Section VI concludes the study.

2. Related work

2.1. Respiratory rate estimation

Many previous studies addressed the problem of respiratory rate (RR) estimation from a sequence of thermal images recorded for a facial region. The most often used approach is based on the localization of the pixels that convey respiration information, i.e., thermal image pixels, whose values change due to the local temperature gradients caused by respiration. Such source areas or pixels are identified manually or automatically near nostrils and mouths (e.g., [27] [9] [2] [5] [24]). Next, values of source pixels are aggregated for each frame to obtain a time series (e.g., using a mean or other statistics [29] [35]). The extracted raw signals are post-processed using filtration (e.g., band-pass filters [25], moving-average filtration [29], etc.). Finally, the respiratory rate is estimated with peak detectors (e.g., [39]), wavelet analysis [9] or methods based on Fourier domain analysis (peak in the frequency domain, or using auto-correlation techniques [7]). Several deep learning methods were also proposed to improve the process of respiratory rate estimation (e.g., [21], [13] [22], [39]).

However, to our knowledge, no studies have been reported on respiratory signal estimation from faces covered by different virus protection masks in various activities of subjects. Only limited studies address the problem of facial images with masks in the related context. The authors of [32] proposed a synthetic dataset of thermal images of people with masks by putting masks on faces from the SpeakingFaces [1] set. The Cascade R-CNN was used to determine whether a mask was on the detected face and whether the mask's color indicated the inhalation or exhalation phase (the classification problem). The mask detection validated on the synthetic subset reached high average precision values (AP=0.879 for the best model). In addition, the authors recorded thermal sequences from 11 people at a distance of 1.5 m from the camera. Subjects wore facial masks (unknown types) and were asked to take breaths slowly, faster, and normally. This dataset was used to validate the model. The authors obtained MAE=3.76. In [15], authors investigated the effects of three types of masks (a surgical mask, a cloth mask, or an N95 respirator with an exhalation valve) on the thermal signatures of exhaled airflow when breathing, speaking, or coughing. In this preliminary study, with the participation of seven subjects, authors manually selected several ROIs on images with and without facial masks. They showed no significant difference between breathing rate estimation with or without a mask. However, the experiment is difficult to evaluate since comparing separate measurements in time (a person without a mask and later with a mask) is challenging. The problem of vital sign estimation from faces covered by masks was also addressed in [37]. However, in this study, the authors were interested in pulse rate from images recorded in visible light. The authors showed that it is possible to evaluate the heart rate with slightly worse performance for masked faces (e.g., using the forehead area as proposed in some of the earlier studies [23]).

2.2. Mask detection

Several papers addressed the mask detection problem for visible and thermal light images. Most of them use visible light images. In [36], authors used the transfer learning method and the ResNet-50 [12] model, which was trained using the MAFA dataset. They achieved a mask detection accuracy of over 98%. Automatic mask detection based on IoT in public transport has been proposed in [20]. A hybrid model combining deep learning and machine learning was created and validated on the created dataset (over 1,600 images) and publicly available datasets to determine whether a person is wearing a face mask. The detection accuracy was over 99%. Another interesting problem was the detection of the position of the mask on the face. Probably only one of the publicly available sets contains annotations about the location of the mask area on the face - a face mask detection dataset (FMD) [19]. It has over 52,000 images and annotations of location and classes - face with and without a mask, incorrect mask, and mask area. The authors used Yolov4 [4] model, achieving an average precision of 87.05%. The same group has also created another solution - ETL-YOLO v4 [18] using the FMD set, reaching the Average Precision of mask location detection of about 87% and mean Average Precision (mAP) of about 67.6%.
Only limited works addressed the detection of a mask or a face with a mask for thermal images. The examples include the work [32] described in the previous section. In [10], authors analyzed the face detection problem of people with and without masks. A new private set of over 7,900 thermal images was used. Several popular classifiers were analyzed, and some were also pre-trained on RGB images. The best results were obtained for the adapted Yolov3 [33] model. The achieved mAP was 99.3%, while the precision was limited and equal to 66.1%. In [17], authors proposed detecting a face mask area in thermal images. They created a dataset of nearly 9,400 thermal images of people wearing three types of masks. The annotated dataset was used to train the adapted Yolov5 [14] model in the "nano" version. The obtained mAP was over 97% with a precision of about 95%. In addition, they proposed a CNN-based model to classify face mask types into three classes: surgical, FPP2, and cloth mask. The obtained accuracy for the best model

was 91%.

3. Methods

In Figure 1, the overall pipeline of the proposed method is shown. Mask and face detection are performed for each frame of the thermal video recorded for each subject. A respiratory signal is extracted based on the average value of pixels located in the region of the mask or the lower half of the face. A binary respiratory signal (inhalation) is also extracted from each frame. In the next stage, the post-processing of breathing signals takes place. The last step is the quantitative analysis of the extracted signals providing estimated respiratory rates, SNR values, etc.

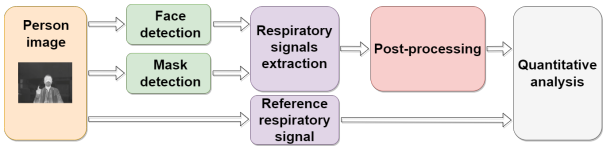


Figure 1. Method pipeline overview.

3.1. Experimental setup and data acquisition

Two experiments were conducted to collect a proprietary dataset of thermal recordings to determine respiration rate. The first was for a sitting subject and the second for a slowly walking subject. All sequences were recorded with the FLIR A665SC thermal camera with 50 frames per second. The frame's resolution was 640x480 pixels.

In the first experiment (sitting subjects), 15 participants were asked to wear a certified FFP2 (N95) mask, and one-minute recordings were recorded. The subject was looking at the camera and breathing. During a series of recordings, to obtain a reference signal, the subject had to signal inhalation (upward movement) and exhalation (downward movement) with a finger movement. After a short break, the same procedure was used for subjects wearing surgical masks. In this part of the study, 30 recordings were collected, 15 for each type of mask. The subject was about 1 m away from the camera. The temperature in the room was 22.5 degrees Celsius. The average age of the study participant was 35.66 ± 12.56 years. An example of two images from the recordings (for two different types of masks) is presented in Figure 2.

The second experiment consisted of recording the breathing process in masks while slowly moving toward the camera (walking subjects). This experiment resembled real-world conditions, such as moving in a queue at airports, testing points, or building entrances. Each of the 15 participants (age 41.47 ± 11.51 years) moved towards a camera set 6m from the starting point. It must have taken 1 minute to walk this road. As in the first experiment, each

(a) (b)

Figure 2. Examples of images from recordings obtained during experiment 1: (a) a person with a surgical mask and (b) a person with an FFP2 mask.

participant had to signal the moment of inhalation and exhalation. As this experiment is practically more interesting, more measurements were taken. The first recording for each participant was made in a surgical mask. Next, it was repeated. The third recording was made for an FFP2 mask, and it was also repeated. The fifth recording was also performed for subjects wearing FFP2 masks. But the participants were asked to move their heads from side to side (right-left) along the way. In total, 75 files (about 1.5GB each) were recorded, each with a sequence of images lasting 1 minute. Unfortunately, after the experiments, two files were found corrupted. However, the corrupted files were for repeated measurements, so for each experiment, at least one recording was available for each subject. The temperature where the experiments were performed (a building corridor) was 19 degrees Celsius. Sample images from the recordings obtained in the second experiment are shown in Figure 3.

Figure 3. Examples of images from recordings registered during the second experiment - for two distances from the camera for a person with an FPP2 mask.

The experiments were performed with permission of the local Committee for Ethics of Research with Human Participants on 02.03.2021. Informed consent was obtained from all subjects involved in the study.

3.2. Face and mask detection

Respiratory signal extraction requires identifying the pixels that change their values in time due to respiratory activity. Two methods were used in this study. The first uses the adapted model of the face with mask detection from [10]. Based on the detected region of the face, the lower half of the region is used to calculate the average

value of the pixels in this area for each frame. As a result, a raw respiratory-related signal is extracted. The choice of this region is closely related to the location of the mask on the face, i.e., the covered area of the nose and mouth, based on which RR can be determined.

The second method consists in detecting the mask area on the face in thermal images using the adapted Yolov5 model [14], specifically created for mask detection in thermal images [17]. The trained model achieved very high values of mAP and precision (>95%), so it was used in this study. The respiratory-related signal is extracted similarly to the previous method - by obtaining the average value in the detected region of the mask. In both methods, in the event of a temporary lack of face or mask detection in a given frame (50fps), the predicted area of interest is approximated using the area extracted in the previous frame.

For each recording, reference signals were created based on the finger movements of the subject. These signals are binary and represent the start of inhalation. It should be noted that the moment of inhalation determined by the study participant is subjective; therefore, it may differ from person to person.

3.3. Extraction of respiratory signals and post-processing

Each extracted respiratory signal from the respective ROI of each frame was processed with a baseline removal filter implemented using asymmetrically reweighted penalized least squares smoothing [3]. Next, each signal was normalized to a [0,1] range. Noise peaks were identified as values outside the 1.5 X interquartile range (IRQ). Identified peaks were smoothed with an upper or a lower limit value:

$$v_{ul} = \mu \pm 1.5 * IQR \tag{1}$$

Next, the square root of each sample was calculated to amplify low amplitude changes. Signals were further standardized about mean and smoothed using the asymmetric least squares smoothing function proposed in [8]. Finally, the values of each signal were inverted to obtain positive values for the inhalation phase that corresponds to the binary reference signal.

Fig. 4 presents examples for the selected post-processing steps.

3.4. Quantitative analysis

The SNR values were calculated for each raw, extracted respiratory signal (i.e., before post-processing). The modified version of the SNR metric from [11] was used:

$$SNR = 10log_{10}\frac{\sum_1^{60}(U_t(f) * \hat{S}(f))^2}{\sum_1^{60}((1 - U_t(f)) * \hat{S}(f))^2} \tag{2}$$

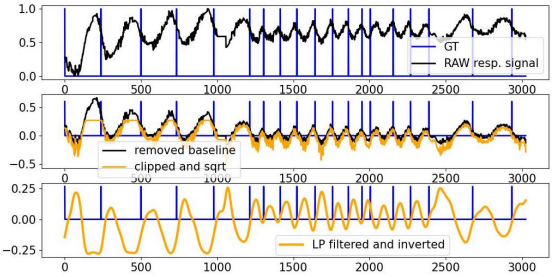


Figure 4. Example of the results of raw respiratory-related signal post-processing functions.

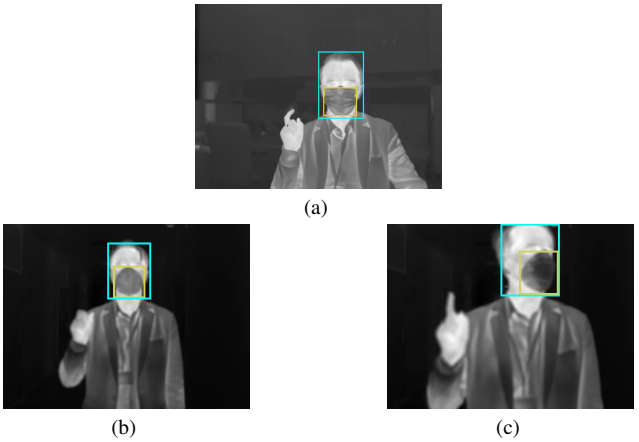
where $\hat{S}(f)$ is the spectrum of the extracted, raw respiratory signal, S, f is the frequency in breaths per minute (BPM), and $U_t(f)$ is a binary template window moving through the spectrum. The 10-bin window was used in this study.

Finally, the respiration rate as breaths per minute was estimated as the mean of peak distances in a signal, which was returned by the autocorrelation function.

All the methods were implemented in Python and are available in a GitHub repository [16]. Additionally, example video recordings are included for all experiments. This should allow future testing of the method and also for future video recordings.

4. Results

Figure 5 presents examples of face (in blue) and mask (in orange) detection in thermal images. Despite the different quality of recordings and facial or mask features blurring, masks are detected with high accuracy.



(a) (b) (c)

Figure 5. Selected frames with detected face and mask bounding boxes from sitting subject with surgery mask (a) and walking subject with an FFP2 mask in the 60th second of the recording: (b) without head movements and (c) with head movements.

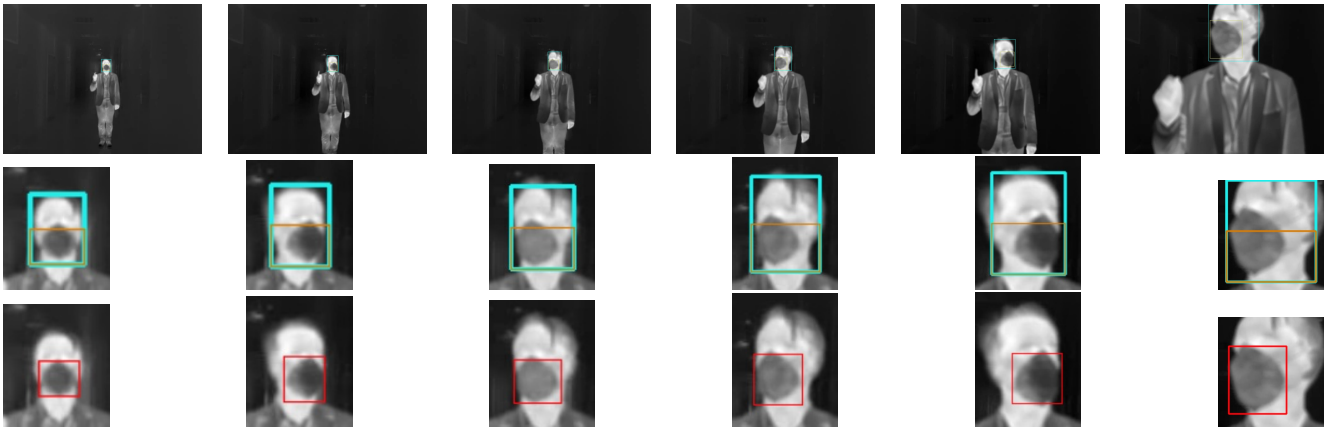


Figure 6. Selected frames with detected face and mask bounding boxes from a walking, W00 subject wearing an FFP2 mask, in the 0, 10, 20, 30, 40, and 60th second of the recording with head movements: original image resolution (top), cropped and enlarged image with detected face - blue, and half of the face - orange (middle), and cropped and enlarged image with detected mask area (bottom).

Other examples of the face/mask detection phase are presented in Figure 6. It shows frames from 0, 10, 20, 30, 40, and 60th seconds of recordings from the walking subject with head movements. The detected regions are presented in images with the original resolution (top) and as the cropped and enlarged versions with face area (bottom). Automatically detected face and mask regions are presented with blue and red rectangles. Additionally, the extracted lower half of the detected face used in the further analysis is presented as an orange rectangle. The face and mask area are detected well at different angles of the face to the camera and at different distances from it. The difference between the mask region and the lower half of the face region is visible for rotations with higher angles (e.g., images in the last columns). Figure 7 shows examples of reference signals and the raw signals extracted from the recordings collected during the "walking experiment". The presented signals were extracted for the participants wearing an FFP2 mask. The graphs show the moments of inhalation - a decrease in the signal value due to the decreasing temperature in the detected area and exhalation - an increase in the signal value by heating the mask area with exhaled air.

Due to limited space, the results for all experiments are available at the GitHub repository [16]. Examples of detailed results for all sitting subjects in the FFP2 mask are in Table 1 and for walking subjects in the FFP2 mask in Table 2.

The cumulative results for all experiments for sitting and walking subjects (without head movements) are presented in Table 3.

The extracted respiratory signals for each sitting participant are shown in Figures 8 - for the lower half of the detected face region and 9 - for the detected mask region.

Similarly, filtered signals for the walking subjects without head movements are shown in Figure 10. The mask

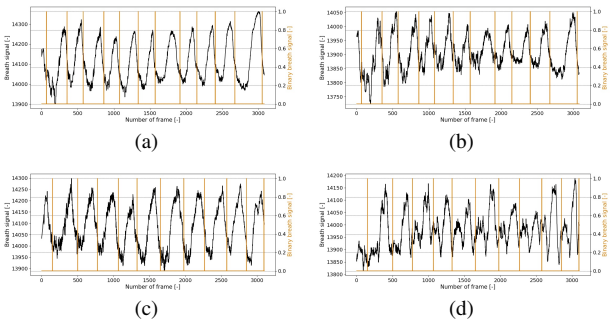


Figure 7. Examples of raw respiratory signals and reference signals (start of inhalation) for the walking, W00 subject: (a) and (b) without head movements, (c) and (d) with head movements. Signals (a) and (c) are extracted for the detected mask region, and (b) and (d) for the lower half of the detected face.

Example	gt_bpm	det_bpm_mask	det_bpm_face	SNR_mask	SNR_face
S00	21	19.523	19.523	26.836	13.372
S01	16	14.543	14.563	10.695	7.173
S02	11	10.292	10.300	24.803	13.455
S03	10	10.545	10.610	16.266	9.712
S04	14	13.720	13.544	11.196	11.596
S05	11	11.070	11.268	7.989	10.442
S06	15	15.025	15.038	35.459	32.370
S07	10	9.202	9.202	21.097	14.707
S08	13	13.550	13.587	19.446	10.033
S09	19	18.265	18.265	27.613	27.870
S10	12	10.753	10.811	17.503	15.683
S11	15	14.901	14.975	17.267	14.872
S12	16	16.601	16.588	24.966	21.676
S13	11	10.408	10.399	26.277	27.523
S14	19	14.936	14.894	13.199	8.579

Table 1. Comparison of the ground truth ('gt_') value of the number of breaths with the number automatically detected ('det_') for mask and face regions for sitting subjects wearing an FFP2 mask

regions were used to calculate average values.

In the "walking experiment," participants were also

Example	gt_bpm	det_bpm_mask	det_bpm_face	SNR_mask	SNR_face
W00_0	13	12.245	12.336	24.964	19.799
W00_1	12	11.696	11.742	22.138	16.365
W01_0	13	11.848	11.691	7.181	8.214
W01_1	12	12.058	11.914	7.584	11.554
W02_0	17	17.529	17.515	24.360	16.815
W02_1	16	16.575	16.522	22.673	14.125
W03_0	10	9.514	9.554	24.918	22.621
W04_0	8	7.673	7.595	13.380	12.257
W05_0	13	12.712	12.842	24.035	12.208
W05_1	12	10.733	10.753	23.882	11.375
W06_0	11	13.286	17.013	16.951	7.010
W06_1	12	12.205	12.255	19.788	9.678
W07_0	16	15.652	15.693	11.760	7.129
W07_1	16	15.332	15.280	21.841	5.581
W08_0	13	11.820	21.641	7.622	4.147
W08_1	12	10.909	13.357	9.292	9.209
W09_0	19	11.936	12.521	13.660	9.388
W09_1	29	18.605	18.100	4.47	6.189
W10_0	14	14.184	13.953	16.793	10.528
W10_1	14	13.624	13.636	15.743	9.765
W11_0	16	16.304	14.041	4.765	8.926
W11_1	16	15.986	16.304	9.071	3.341
W12_0	6	5.484	5.386	26.096	14.139
W12_1	6	5.747	5.964	24.815	15.287
W13_0	8	7.042	7.220	19.274	14.535
W13_1	8	8.555	8.483	24.960	15.784
W14_0	7	6.383	6.472	30.482	15.468
W14_1	7	5.917	5.814	25.120	21.744

Table 2. Comparison of the ground truth value of the number of breaths with the number automatically detected for mask and face regions for walking subjects wearing an FFP2 mask

Experiment	Mask type	Region	RMSE	MAE	minSNR	mSNR
sitting	ffp2	face	1.314	0.908	7.173	15.938
		mask	1.303	0.883	7.989	20.041
	surgery	face	0.849	0.721	5.591	18.296
		mask	0.84	0.703	6.368	20.839
walking	ffp2	face	3.195	1.664	3.341	11.899
		mask	2.496	1.208	4.47	17.789
	surgery	face	6.966	4.314	2.191	10.492
		mask	3.209	2.107	2.265	14.453

Table 3. Estimation error values for all experiments (without head movements, mSNR - mean SNR)

asked to perform head movements. This type of experiment is similar to natural conditions (people turn around and look at the sides). Table 4 shows the results obtained for each walking participant wearing the FFP2 mask, with head movements.

Table 5 shows the calculated values of the performance measures for the walking subjects with head movements. It presents results for the mask's detected area and the face's lower half. One outlier was identified (W04, the lowest SNR - see also in Figure 11). The presented results were calculated with and without the outlier.

Examples of post-processed signals and ground truth respiratory signals obtained for walking subjects with head movements are presented in Figure 11. The signals extracted from the detected mask area and face area are compared. Additionally, the signal with the lowest SNR is presented (W04).

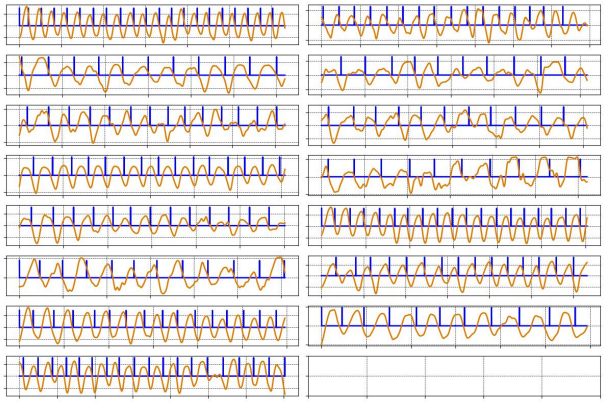


Figure 8. Extracted (filtered) respiratory signals for sitting subjects without head movements. The lower half of the face region was used as the ROI for pixels aggregation.

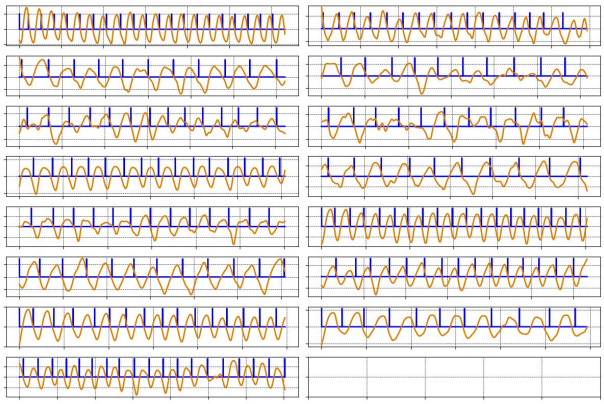


Figure 9. Extracted (filtered) respiratory signals for sitting subjects without head movements. The mask region was used as the ROI for pixels aggregation.

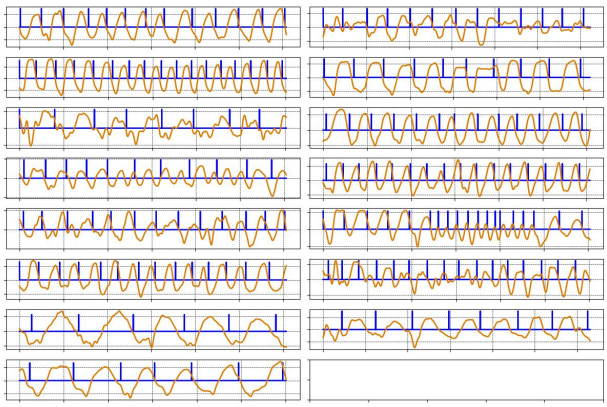


Figure 10. Extracted (filtered) respiratory signals for walking subjects without head movements. The mask region was used as the ROI for pixels aggregation.

Example	gt_bpm	det_bpm	SNR
W00	11	9.967	25.723
W01	10	9.975	18.922
W02	14	14.516	22.699
W03	11	10.230	25.517
W04	9	22.422	4.156
W05	12	11.080	19.523
W06	17	16.949	16.907
W07	15	14.822	13.562
W08	12	11.353	9.841
W09	25	24.773	10.947
W10	11	10.629	22.006
W11	12	11.299	18.583
W12	5	4.082	19.343
W13	12	12.552	9.189
W14	7	6.452	21.721

Table 4. Comparison of the ground truth value of the number of breaths with the number automatically detected for mask region for walking subjects in FFP2 mask

Experiment	Mask type	RMSE	MAE	minSNR	mSNR
walking	mask_all	3.517	1.392	4.156	17.243
	mask_no_1_outlier	0.618	0.533	9.189	18.177
	face_all	6.079	4.466	5.321	9.952
	face_no_1_outlier	5.773	4.116	5.321	10.153

Table 5. Estimation error values for walking subjects with head movements (mSNR - mean SNR)

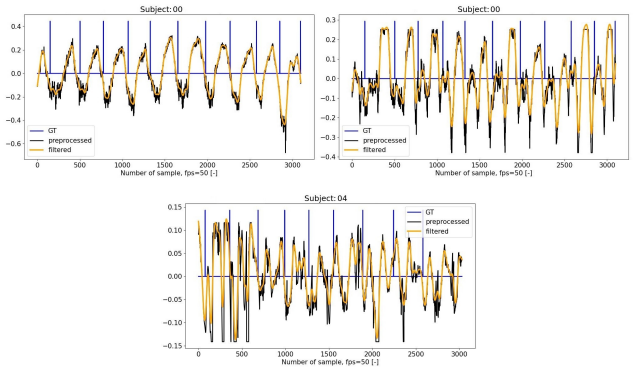


Figure 11. Extracted (filtered) respiratory signals for walking subjects with head movements. Waveform obtained for the subject S00 from the detected FFP2 mask ROI (top), from the lower half of the detected face ROI (middle), the signal with the lowest SNR value for subject W04 (bottom).

Figure 12 presents all extracted respiratory signals for walking subjects with head movements (without the W04 outlier) when the mask region was used as the source ROI.

5. Discussion

When automatic mask detection is used to identify pixels that convey respiration information, lower RMSE and MAE values and higher SNR values are observed for both experiments (for sitting and walking subjects). The lower part of

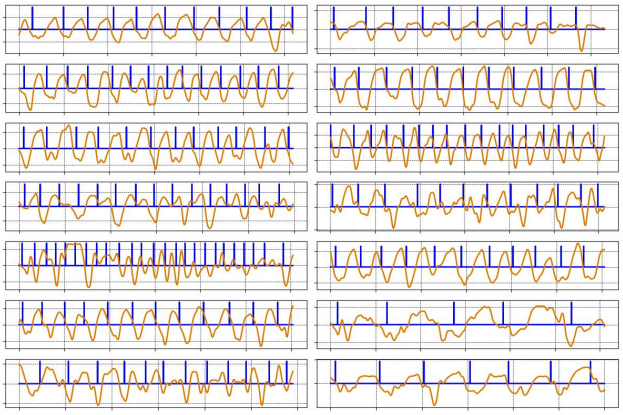


Figure 12. Extracted (filtered) respiratory signals for walking subjects with head movements (without the W04 outlier). ROI = mask region.

the detected face area is an approximation of the location of the mask area, and in most cases, it is larger than the mask area. This can cause inaccuracies and disturbances in the extracted signal.

The difference in the values of the obtained RMSE, MAE, and SNR measures is also visible for walking and sitting subjects without head movements. Results for walking conditions are less accurate. The factors that affect the obtained results are probably the distance from the camera and the image quality (e.g., related to proper respiratory-related source identification).

The accuracy of the respiration rates estimated for signals obtained for the sitting subjects is less than 1 bpm (MAE) and is within the Fourier-based method accuracy. For N=3000 and fs=50Hz, the resolution in the frequency domain is $\delta f = fs \div N = 50 \div 3000 = 0.0167Hz, 0.0167Hz * 60s = 1bpm$. The results are much better than in [32] where MAE=3.76 was obtained for a similar "sitting" experiment (but with unknown types of facial mask). However, different test datasets were used, so the obtained MAE values cannot be directly compared. This study used different mask types, and more difficult "walking" experiments were performed. The respiratory signal is also estimated, allowing for a much better explainability of the results. The worst results were obtained for the walking experiment for the subjects with the surgery masks and when the lower half of the detected face area was used as a source ROI. This is a consequence of improperly identifying pixels that do not convey respiration information. The surgery mask detection in thermal images is less accurate for the used method. Similarly, the lower half of the detected face region can include many, not respiratory-related pixels. This can be especially present in the first phase of the walking, when the subject is far away from the camera (small resolution) and when a head is rotated. Nevertheless,

the RR was accurately estimated for all experiments when automatic mask detection was used. The time cost for the detection of the mask with Yolov5 or for the face with mask detection with Yolov3 model is approximately 15.6 ms per frame, so they can be certainly used in real-time scenarios.

For the W04 example, the obtained SNR value is significantly lower. The high noise content of the signal extracted for this person may be due to fast and large-angle movements, and the mask detection model could not correctly detect its exact location. In future studies, the value of SNR could be used to filter reliable signals (or other methods should be used to improve the quality of extracted signals).

Many previous studies on estimating respiratory signals used the detection of nostrils, mouth regions, or related pixels (e.g., [9], [22]) to locate the best sources representing the highest contribution to the respiratory signal-to-noise ratio. It is much easier for subjects who are wearing a mask. A mask is directly heated or cooled by the respiratory-related heat flow. This study showed that it was possible to accurately extract respiratory signals for subjects wearing different mask types. The facial masks used in this study are probably the most popular and effective protection against airborne viruses. Although many different factors were considered in this work, the study has several limitations. First, other mask types (and from other producers) could be used to analyze the influence of the mask type on the results. In this study, the FFP2 and surgery masks were used, and the results show no significant difference in the chosen metric values (similar mean SNR values and MAE $<$ 1 bpm).
The subject's self-observation of inhalation/exhalation phases was used. Each breathing phase beginning was signaled by rising or lowering a pointing finger. This reference method has limited accuracy since some subjects can show intention to start a given breathing phase while others can point to these two phases with different delays. The other option would probably be to use the pressure belt [34], which could be easily used in stationary experiments. Using a pressure belt as a reference for the walking experiments would be more difficult due to movement-related noise requiring additional filtering. Also, related data synchronization would be a challenge. Nevertheless, it is an interesting option to use in the future.
Further studies should also consider different walking paths, head movements, etc. Probably the best validation option would be to observe the subject in real situations (e.g., at the airport), but this should require many data privacy aspects to be addressed. The volunteer group has only 15 healthy subjects. However, even in this group, different breathing behavior was observed. Some volunteers were breathing with a high frequency of breaths (29 bpm), others with a low rate (5-6 bpm) or with irregular rhythm (e.g., Fig. 4). Nevertheless, observing the results for a larger group, including patients, would be very interesting for future studies.

Another limitation of the study is related to the local measurement environment. The experiment was performed in the corridor with no sources of reflections and for typical room (hall) temperature. Other environment options related to different ambient temperatures, moisture levels, the influence of air conditions, etc., could be considered in future studies. It is also important to mention that this study's traditional data processing method was based on signal processing. Many other methods have been recently proposed for camera-based physiological signal extraction [26] that could be used in the future. However, this study has many advantages, as concluded in the next section.

6. Conclusion

This paper presents probably the first study on estimating respiratory signal and rate from automatically detected facial mask regions in thermal videos for sitting and walking subjects. Different experiments were designed, and the results show high reliability of respiratory rate estimation in the considered scenarios. We also demonstrate that pixel values from the automatically detected mask region improve the results compared to data obtained from the detected facial sub-region. The proposed method produces similar results for two popular virus protection masks: FFP2 (N95) and surgical masks. We analyze the extracted respiratory signals and related respiratory rates for subjects wearing different types of masks while sitting, slowly walking from a short distance toward a camera, and slowly walking with moderate head movements. The values of quality metrics were lower for "walking experiments," which was expected. However, the results could be acceptable for many applications. They could be improved using more sophisticated data processing methods (e.g., identification of best sources within mask region).
The proposed method is highly explainable. It is relatively easy to match each point at the extracted respiratory signal with a mask detected from a given video frame showing the changes in temperature within the mask area. The examination documentation could contain the original sequence of frames with detected mask ROIs or, to improve data privacy, only a sequence of mask ROIs.
The presented work is part of a comprehensive study focused on the complex, remote examination of people wearing virus protection masks using thermal imaging. The complex examination considers previously known methods for face detection and body temperature estimation using detected characteristic facial regions (the inner canthus of the eyes) and new methods for mask detection [17], mask type classification [17], respiratory signal and respiratory rate estimation (this study, and future works on respiratory patterns analysis), identification if a mask is appropriately adjusted and covering nose, mouth, and chin (work in progress).

References

[1] Madina Abdrakhmanova, Askat Kuzdeuov, Sheikh Jarju, Yerbolat Khassanov, Michael Lewis, and Huseyin Atakan Varol. Speakingfaces: A large-scale multimodal dataset of voice commands with visual and thermal video streams. *Sensors*, 21(10), 2021. 2

[2] Farah Q Al-Khalidi, Reza Saatchi, Derek Burke, and Heather Elphick. Tracking human face features in thermal images for respiration monitoring. 2

[3] Sung-June Baek, Aaron Park, Young-Jin Ahn, and Jaebum Choo. Baseline correction using asymmetrically reweighted penalized least squares smoothing. *Analyst*, 140:250–257, 2015. 4

[4] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020. 2

[5] Ronan Chauvin, Mathieu Hamel, Simon Brière, François Ferland, François Grondin, Dominic Létourneau, Michel Tousignant, and François Michaud. Contact-free respiration rate monitoring using a pan–tilt thermal camera for stationary bike telerehabilitation sessions. *IEEE Systems Journal*, 10(3):1046–1055, 2016. 2

[6] Qiong Chen, Yalin Wang, Xiangyu Liu, Xi Long, Bin Yin, Chen Chen, and Wei Chen. Camera-based heart rate estimation for hospitalized newborns in the presence of motion artifacts. *BioMedical Engineering OnLine*, 20(1):1–16, 2021. 1

[7] Youngjun Cho, Simon J. Julier, Nicolai Marquardt, and Nadia Bianchi-Berthouze. Robust tracking of respiratory rate in high-dynamic range scenes using mobile thermal imaging. *Biomedical Optics Express*, 8, 2017. 2

[8] Paul H.C. Eilers and Hans F.M. Boelens. Baseline correction with asymmetric least square smoothing. *Leiden Univ.Med.Cent.Rep.*, 1:1–24, 2005. 4

[9] Jin Fei and Ioannis Pavlidis. Thermistor at a distance: Unobtrusive measurement of breathing. *IEEE Transactions on Biomedical Engineering*, 57, 2010. 1, 2, 8

[10] Natalia Głowacka and Jacek Rumiński. Face with mask detection in thermal images using deep neural networks. *Sensors*, 21(19), 2021. 2, 3

[11] Gerard De Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60, 2013. 4

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[13] Preeti Jagadev, Shubham Naik, and Lalat Indu Giri. Contactless monitoring of human respiration using infrared thermography and deep learning. *Physiological Measurement*, 43(2):025006, mar 2022. 2

[14] Glenn Jocher. YOLOv5 by Ultralytics, 5 2020. 2, 4

[15] Ekaterina Koroteeva and Anastasiya Shagiyanova. Infrared-based visualization of exhalation flows while wearing protective face masks. *Physics of Fluids*, 34(1):011705, 2022. 2

[16] Natalia Kowalczyk and Jacek Rumiński. Respiratory rate estimation based on detected mask area in thermal images. https://github.com/natkowalczyk/Respiratory-Rate-Estimation-Based-on-Detected-Mask-Area-in-Thermal-Images. 4, 5

[17] Natalia Kowalczyk and Jacek Rumiński. Mask detection and classification in thermal face images. *arXiv preprint arXiv:2304.02931*, 2023. 2, 4, 8

[18] Akhil Kumar, Arvind Kalia, and Aayushi Kalia. Etl-yolo v4: A face mask detection algorithm in era of covid-19 pandemic. *Optik*, 259:169051, 2022. 2

[19] Akhil Kumar, Arvind Kalia, Kinshuk Verma, Akashdeep Sharma, and Manisha Kaushal. Scaling up face masks detection with yolo on a novel dataset. *Optik*, 239:166744, 2021. 2

[20] Tamilarasan Ananth Kumar, Rajendrane Rajmohan, Muthu Pavithra, Sunday Adeola Ajagbe, Rania Hodhod, and Tarek Gaber. Automatic face mask detection system in public transportation in smart cities using iot and deep learning. *Electronics*, 11(6):904, 2022. 2

[21] Alicja Kwasniewska, Jacek Ruminski, and Maciej Szankin. Improving accuracy of contactless respiratory rate estimation by enhancing thermal sequences with deep neural networks. *Applied Sciences*, 9(20):4405, 2019. 2

[22] Alicja Kwasniewska, Maciej Szankin, Jacek Ruminski, Anthony Sarah, and David Gamba. Improving accuracy of respiratory rate estimation by restoring high resolution features with transformers and recursive convolutional models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3857–3867, 2021. 2, 8

[23] Magdalena Lewandowska, Jacek Ruminski, Tomasz Kocejko, and Jedrzej Nowak. Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity. In *2011 Federated Conference on Computer Science and Information systems (FedCSIS)*, pages 405–410, 2011. 1, 2

[24] Gregory F. Lewis, Rodolfo G. Gatto, and Stephen W. Porges. A novel method for extracting respiration rate and relative tidal volume from infrared thermography. *Psychophysiology*, 48, 2011. 2

[25] Lalit Maurya, Reyer Zwiggelaar, Deepak Chawla, and Prasant Mahapatra. Non-contact respiratory rate monitoring using thermal and visible imaging: A pilot study on neonates. *Journal of Clinical Monitoring and Computing*, 2022. 2

[26] Daniel McDuff. Camera measurement of physiological vital signs. *ACM Comput. Surv.*, 55(9):40, 2023. 1, 8

[27] Jayasimha N Murthy, Johan van Jaarsveld, Jin Fei, Ioannis Pavlidis, Rajesh I Harrykissoon, Joseph F Lucke, Saadia Faiz, and Richard J Castriotta. 1, 2

[28] Toshiaki Negishi, Shigeto Abe, Takemi Matsui, He Liu, Masaki Kurosawa, Tetsuo Kirimoto, and Guanghao Sun. Contactless vital signs measurement system using rgb-thermal image sensors and its clinical screening test on patients with seasonal influenza. *Sensors*, 20(8):2171, 2020. 1

[29] Carina Barbosa Pereira, Michael Czaplik, Vladimir Blazek, Steffen Leonhardt, and Daniel Teichmann. Monitoring of

cardiorespiratory signals using thermal imaging: A pilot study on healthy human subjects. *Sensors (Switzerland)*, 18, 2018. 2

[30] Ming-Zher Poh, Daniel McDuff, and Rosalind W. Picard. Advancements in noncontact, multiparameter physi- ological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, 58(1):7–11, 2010. 1

[31] Jaromir Przybyło. A deep learning approach for remote heart rate estimation. *Biomedical Signal Processing and Control*, 74:103457, 2022. 1

[32] Leonardo Queiroz, Helder Oliveira, and Svetlana Yanushkevich. Thermal-mask–a dataset for facial mask detection and breathing rate measurement. In *2021 International Conference on Information and Digital Technologies (IDT)*, pages 142–151. IEEE, 2021. 2, 7

[33] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 2

[34] Jacek Ruminski. Analysis of the parameters of respiration patterns extracted from thermal image sequences. *Biocybernetics and Biomedical Engineering*, 36(4):731–741, 2016. 1, 8

[35] Jacek Ruminski and Alicja Kwasniewska. *Evaluation of Respiration Rate Using Thermal Imaging in Mobile Conditions*, pages 311–346. Springer Singapore, Singapore, 2017. 2

[36] Shilpa Sethi, Mamta Kathuria, and Trilok Kaushik. Face mask detection using deep learning: An approach to reduce risk of coronavirus spread. *Journal of biomedical informatics*, 120:103848, 2021. 2

[37] Jeremy Speth, Nathan Vance, Patrick Flynn, Kevin Bowyer, and Adam Czajka. Remote pulse estimation in the presence of face masks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2086–2095, 2022. 2

[38] Fan Yang, Shan He, Siddharth Sadanand, Aroon Yusuf, and Miodrag Bolic. Contactless measurement of vital signs using thermal and rgb cameras: A study of covid 19-related health monitoring. *Sensors*, 22(2):627, 2022. 1

[39] Fan Yang, Shan He, Siddharth Sadanand, Aroon Yusuf, and Miodrag Bolic. Contactless measurement of vital signs using thermal and rgb cameras: A study of covid 19-related health monitoring. *Sensors*, 22(2), 2022. 2