


## Article

# Examination of Potential of Thermopile-Based Contactless Respiratory Gating

Qi Zhan <sup>1</sup>, Wenjin Wang <sup>2,\*</sup>  and Xiaorong Ding <sup>3,\*</sup>

<sup>1</sup> College of Electrical and Information Engineering, Hunan University, Changsha 410082, China; zhanqi@hnu.edu.cn

<sup>2</sup> Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands

<sup>3</sup> School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 611731, China

\* Correspondence: wwang@tue.nl (W.W.); xiaorong.ding@uestc.edu.cn (X.D.)

**Abstract:** To control the spread of coronavirus disease 2019 (COVID-19), it is effective to perform a fast screening of the respiratory rate of the subject at the gate before entering a space to assess the potential risks. In this paper, we examine the potential of a novel yet cost-effective solution, called thermopile-based respiratory gating, to contactlessly screen a subject by measuring their respiratory rate in the scenario with an entrance gate. Based on a customized thermopile array system, we investigate different image and signal processing methods that measure respiratory rate from low-resolution thermal videos, where an automatic region-of-interest selection-based approach obtains a mean absolute error (MAE) of 0.8 breaths per minute. We show the feasibility of thermopile-based respiratory gating and quantify its limitations and boundary conditions in a benchmark (e.g., appearance of face mask, measurement distance and screening time). The technical validation provided by this study is helpful for designing and implementing a respiratory gating solution toward the prevention of the spread of COVID-19 during the pandemic.

**Keywords:** thermopile array; thermal imaging; respiratory rate; remote screening



**Citation:** Zhan, Q.; Wang, W.; Ding, X. Examination of Potential of Thermopile Based Contactless Respiratory Gating. *Sensors* **2021**, *21*, 5525. <https://doi.org/10.3390/s21165525>

Academic Editor: Maysam Abbod

Received: 28 June 2021

Accepted: 29 July 2021

Published: 17 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



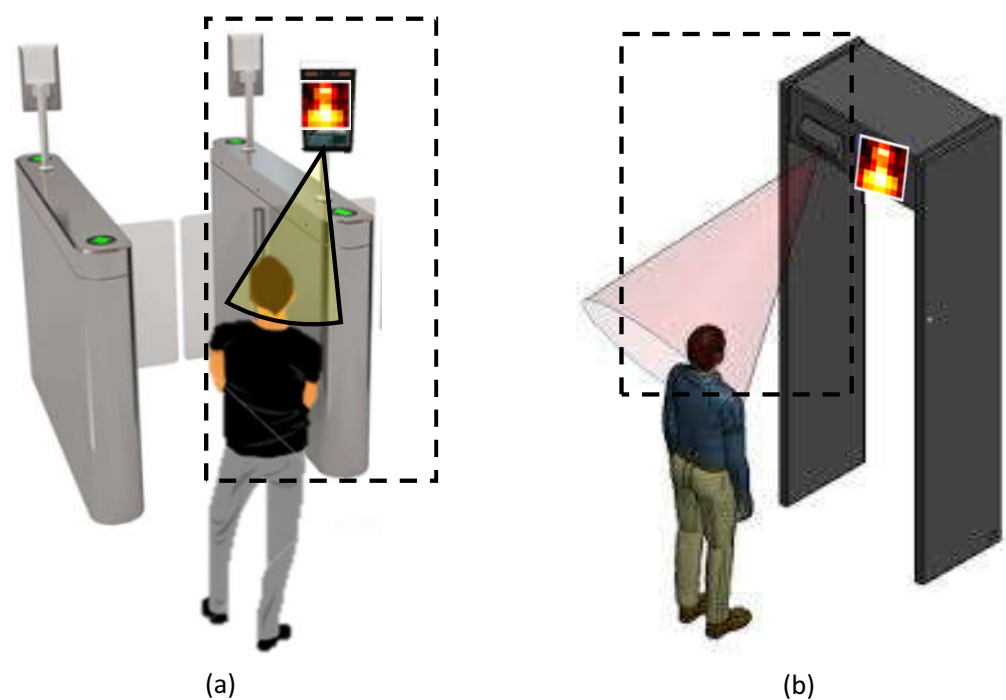
**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Coronavirus disease 2019 (COVID-19) is a novel coronavirus-induced respiratory disease, which has caused over 4,159,378 deaths as of 24 July 2021, according to [1]. It is urgent to take some measures to mitigate the outbreak of COVID-19. As difficulties in breathing is one of the major symptoms, the respiratory rate (RR) can be used as a critical physiological parameter to indicate the health deterioration or well-being of a person [2]. Since COVID-19 is a community-acquired pneumonia that is mainly transmitted through saliva droplets or discharge from the nose [3], people are therefore advised to wear face masks and have their body temperature measured at the entrance of public areas (e.g., train stations, airports, supermarkets, and libraries) to prevent the transmission of the virus [4]. However, body temperature only provides one dimension of information, and its accuracy is affected by environment (e.g., room temperature). The guidelines of National Institute for Health and Care Excellence (NICE) show that clinical features (e.g., body temperature  $\geq 38^\circ\text{C}$ , respiratory rate  $\geq 20$  breaths per min, pulse rate  $> 100$  min and crackles) could provide a rapid diagnosis of community-acquired pneumonia [5,6]. Adding one more variable (e.g., respiratory rate) can improve the accuracy of COVID-19 screening in public areas. Since COVID-19 is an infectious disease, we propose and examine the potential of a non-contact respiratory screening solution, called thermopile-based respiratory gating, that eliminates the risk of infection/contamination caused by sensing in a contact manner.

The concept of respiratory gating is illustrated in Figure 1. Before entering a public space, subjects (with or without face mask) shall pass the gate, where a contactless sensor

is used to measure and check the RR. In this application scenario, the sensor selection and algorithm design are critical. Since it aims at screening the RR of a subject in the stand position, the motion-based methods [7–10] and photoplethysmography-based methods [11–13] are not suitable for this scenario. The reasons are as follows: (i) it is difficult to accurately measure the chest/belly movement (induced by inhaling and exhaling) from a subject in the stand position due to involuntary body motions; (ii) it is difficult to detect the skin pulsation from limited skin areas under a face mask; (iii) the uncertainty and variation of the on-site illumination condition pose an extra challenge (i.e., unknown factor) for the measurement that requires an active light source (e.g., RGB or near-infrared camera); and (iv) motion-based respiratory rate measurements do not resemble the true measurement of nostril airflow. Therefore, we consider a thermography-based modality as an appropriate option here.



**Figure 1.** Illustration of the proposed thermopile-based respiratory gating solution that contactlessly screen the RR of a subject at the entrance gate. (a) Entry gate, (b) security gate.

Thermal-based respiration monitoring has been proposed and demonstrated on both high and low resolution thermal cameras [14–20]. Based on our targeting scenario (i.e., a subject that may wear a face mask in the stand position at the entrance gate), we consider the low-resolution thermal camera as a feasible option, as the respiratory region of interest (ROI) is significantly decreased, due to the use of a face mask. Another consideration is that high-resolution thermal cameras are rather expensive such that they cannot be widely deployed in cost-sensitive areas. Therefore, we propose to use the low-cost thermopile array sensor to build the respiratory gating setup. The thermopile array sensor is comprised of a series of thermocouples, which detects the infrared radiation emitted by all objects within a certain temperature range. It has been widely used for human occupancy detection [21–23]. It is favored in contactless health monitoring applications because the thermopile array sensor evades privacy issues by its low-resolution property (e.g.,  $8 \times 8$  pixels), e.g., seizure detection during sleep [24], fall detection [25,26], sleep posture classification [27,28], household activities monitoring [29], bed-exit detection [30] and head/body posture detection [31].

Though low resolution is an attractive property for thermopile array sensors in terms of privacy protection, it remains a main challenge for image and video processing. It is impossible to perform face detection or facial landmark detection [14,15] on a thermopile image. Different approaches have been proposed to address the issue of ROI detection. Pereira [16] computed the signal quality index (SQI) of each ROI and selected suitable ROIs to extract the respiratory signal based on SQI values. Since the ROI to be selected contains breathing-induced motion and respiratory flow, the extracted signal is unreliable, especially when apnea events are present. Lorato [17] used nostril temperature changes to define the ROI, but since the nostril is a very small area, the distance between the subject and sensor needs to be very close (e.g., 10 cm). The method used in [17] is not suitable for our gating scenario, because it is difficult to require people to keep a certain distance from the sensor in public. To detect the RR from low-resolution thermal videos in our gating scenario, we built a setup in the lab and investigated different image processing and signal extraction approaches, which include the following: (i) full video processing that uses very simple spatial statistics of  $8 \times 8$  pixels to generate the respiratory signal; (ii) ROI-base processing that performs a rough segmentation of respiratory and non-respiratory regions to refine the extraction of the respiratory signal. The feasibility of thermopile-based respiratory gating was demonstrated and its limitations and boundary conditions that may appear in real applications were fairly discussed. Fast Fourier transform (FFT) and inter-beat-intervals (IBI) are commonly used for respiratory rate calculation (averaged and instantaneous rates) [32]. However, the difference of these two respiratory rate calculations has not been thoroughly explored. We evaluated their differences in this work in the context of fast respiratory screening.

The main contribution of this paper is that we propose a novel concept for contactless respiratory gating that uses a cost-effective thermopile sensor and a simple image- and signal-processing method to screen the RR of a subject at the entrance of public areas, where the subject is in the standing position (with or without face mask). It targets a new application scenario in COVID-19 that helps in controlling the spread of the pandemic, using contactless health sensing technology. The proposed respiratory gating is cost effective and easy to deploy in practice. For image-processing algorithms, we present different options where an automatic ROI selection-based method is highlighted in our benchmark. We also report the limitations and boundary conditions of this proposal by quantifying the effect of the factors included (e.g., with/without face mask measurement distance and screening time). The obtained insights improve the understanding toward real applications. The remainder of this paper is organized as follows. In Section 2, we introduce the measurement setup for respiratory gating. In Section 3, we present and analyze six benchmark methods based on either the full video processing or ROI selection. Section 4 shows the experimental results and discussions. Finally, in Section 5, we draw the conclusions.

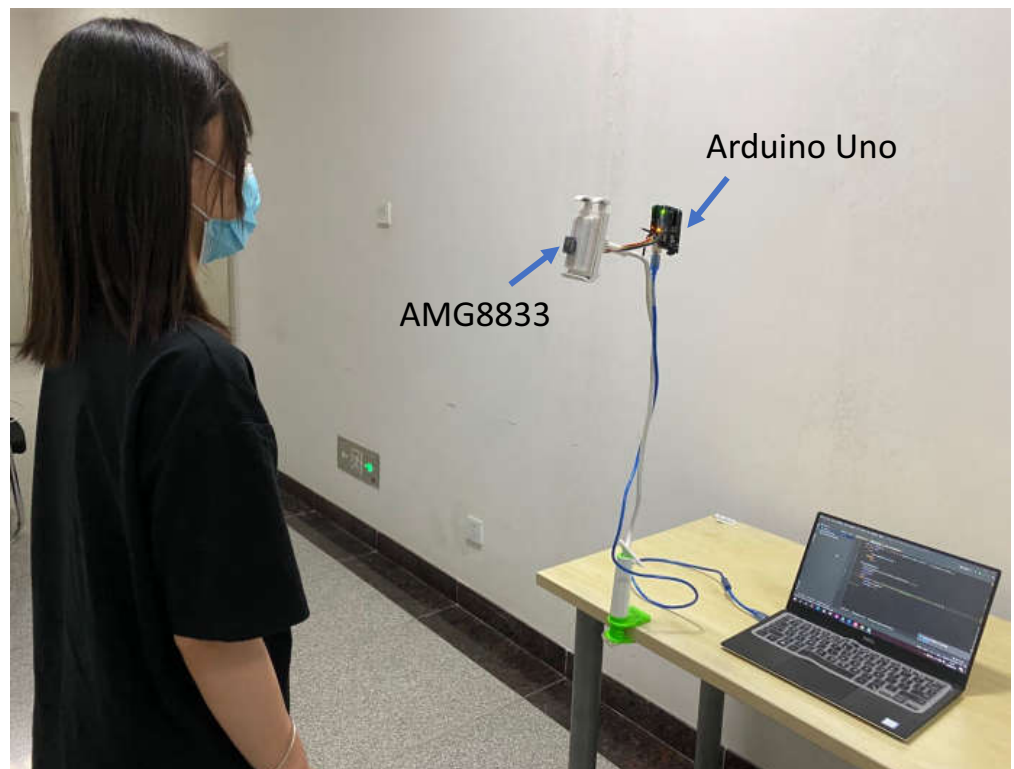
## 2. Setup and Measurement

This section introduces the measurement setup for thermopile-based respiratory gating, which was used to collect the benchmark dataset.

### 2.1. Experimental Setup

To explore the feasibility of using a thermopile array sensor for respiratory gating, we built an experimental setup (see Figure 2) that consists of a Grid-Eye thermopile array AMG8833 from Panasonic, an Arduino Uno and a laptop. The thermopile sensor was placed in front of the subject with an angle of view of 60 degrees, roughly aiming at the nostril/mask area for nasal flow measurement. The thermal videos were recorded at a constant frame rate of 10 frames per second (fps). The spatial resolution is  $8 \times 8$  pixels with absolute temperature distribution. During the recording, the thermopile array sensor was connected to an Arduino Uno, and the acquisition was performed through Python on a

laptop with an Intel Core i5 processor (2.30 GHz). A moving mean filter was applied on the raw data of each video with a sliding window (with 1 s length) to reduce quantization noise.



**Figure 2.** The experimental setup for recording the thermopile videos of a standing subject with a face mask.

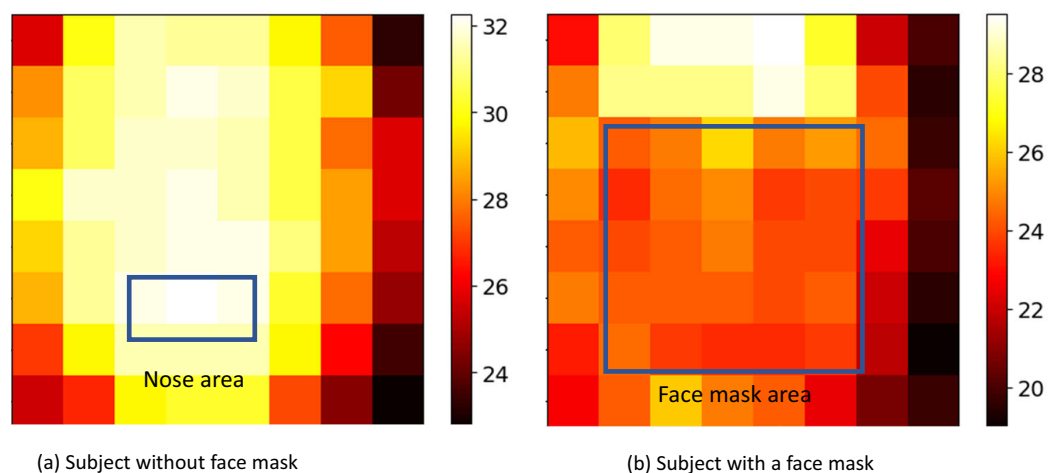
## 2.2. Benchmark Dataset

A total of 75 videos were recorded from 12 healthy subjects (3 males and 9 females, aged from 22 to 68 years) with different configurations. Unless specified otherwise, each subject stood still in front of the thermopile array sensor and was guided to mimic the sinusoidal breathing pattern displayed on the frontal screen during the recording. The guided breathing signal had a duration of four minutes and the breathing frequency was changed from 10 to 30 breaths per minute (bpm). Specifically, the breathing frequency in the first and fourth minute was 20 bpm, and in the second and third minute, 10 bpm and 30 bpm, respectively. For the recording with a face mask, each subject was required to wear a surgical mask. This study was approved by Hunan University, and written informed consent was obtained from each subject.

### 2.2.1. Dataset A: Guided Breathing with and without Face Mask

In real applications, the subject standing at the gate may not wear a face mask, and the distance between the subject and sensor may vary. We included these challenges in our experiments. Five scenarios with different subject-to-sensor distances were created, including the cases with and without a face mask. For the recordings with a face mask, the thermal videos were recorded at three different distances of 10, 30 and 50 cm. For the recordings without a face mask, we only performed recordings at distances of 5 cm (N-5 cm) and 10 cm (N-10 cm). Based on a pilot measurement, we found that the thermopile cannot measure the nostril temperature changes beyond the distance of 10 cm, due to the large quantization noise when the measurement is performed on the small nostril ROI with a few pixels [17]. Figure 3 exemplifies the image areas with respiratory flow induced heat

exchange for a subject with and without a face mask. The waiting time of the subject at the gate is a critical factor that needs to be considered in practical applications. Therefore, we defined four different sliding window lengths of 5, 10, 20, and 30 s and analyzed the effects of different sliding window lengths to measure the respiratory rate on this dataset.



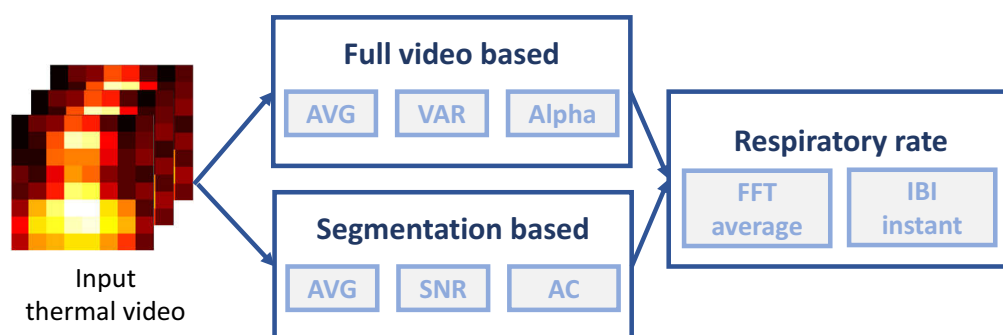
**Figure 3.** Examples of a subject face (with and without a face mask) captured by thermopile array sensor at the distance of 10 cm.

### 2.2.2. Dataset B: Guided Breathing at Different Subject-to-Sensor Distances (with Face Mask)

To explore the boundary conditions for measurement distance (i.e., the maximum distance allowed between the sensor and subject for a valid measurement), the recordings were performed on a single subject with a face mask at multiple discrete distances, ranging from 10 to 150 cm with an interval of 10 cm.

## 3. Methods

To extract the RR from low-resolution videos acquired by the thermopile array sensor, we explored different image processing approaches for respiratory signal extraction and different methods for respiratory rate calculation (see the overview in Figure 4). The image and signal processing methods are detailed in this section.



**Figure 4.** Overview of our algorithmic benchmark system. It consists of two image processing methods (full video based and segmentation based) for respiratory signal extraction and two methods for respiratory rate calculation. AVG—signal selection based on averaging; VAR—signal selection based on variation; Alpha—signal selection based on Alpha tuning; SNR—signal selection based on signal-to-noise ratio; AC—signal selection based standard deviation.

### 3.1. Full Video Processing-Based Methods (FVP)

In this subsection, we introduce three full video processing-based methods that use simple image statistics to create a respiratory signal.

#### 3.1.1. Averaging

Given a thermal video, we use  $I(x, y, t)$  to denote the temperature of a pixel at location  $(x, y)$  of a  $t$ -th thermal image. When the subject wears a face mask or the distance between the subject and sensor is very close (e.g., 5 cm), the respiratory region has a relatively large area in the thermal image. Therefore, we can temporally concatenate the spatially averaged pixel values to approximate a time signal that includes respiratory rhythm, denoted as AVG. For a thermal image with height  $H$ , and width  $W$ , the  $t$ -th respiratory signal  $AVG(t)$  calculated by AVG can be expressed as follows:

$$AVG(t) = \frac{1}{HW} \sum_{x=1}^H \sum_{y=1}^W I(x, y, t) \quad (1)$$

#### 3.1.2. Variation

When the subject does not wear a face mask or the subject-to-sensor distance is large (e.g., 50 cm), the respiratory region will be small in the image. Hence, the spatially averaged signal resembles temperature variations of non-respiratory areas rather than the respiratory flow. According to [33], spatial mean and spatial standard deviation have a complementary effect in qualifying pixels. So, we use the spatial standard deviation of pixel values as an alternative to generate a time signal, denoted as VAR, i.e., the mean of the non-respiratory area is removed from the standard deviation representation. Since the standard deviation is calculated on the second order statistics, it does not reflect the polarity of the signal (e.g., values are all positive). To preserve the inhaling and exhaling phases in the standard deviation signal, we use the third power instead of the second power to compute the spatial variance. A comparison of the spatially averaged signal, 2nd-order and 3rd-order variation signals is shown in Figure 5. The  $t$ -th respiratory signal  $VAR(t)$  calculated by VAR can be expressed as follows:

$$VAR(t) = \sqrt[3]{\frac{\sum_{x=1}^H \sum_{y=1}^W (I(x, y, t) - \mu)^3}{HW}} \quad (2)$$

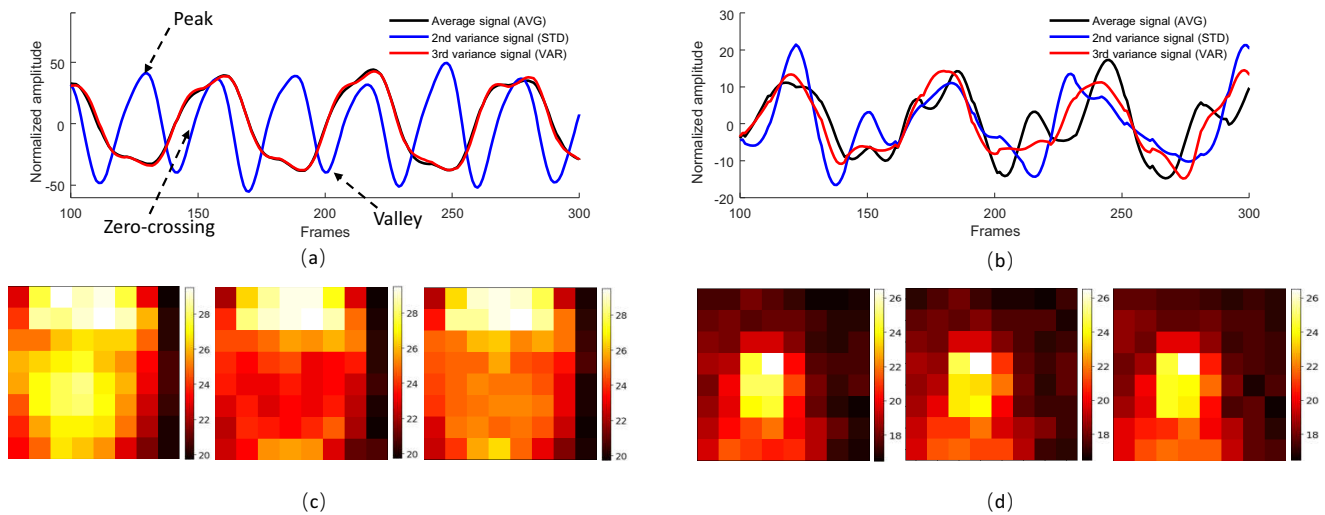
where  $\mu$  denotes the spatially averaged pixel values in the thermal image.

#### 3.1.3. Alpha Tuning

As described above, AVG and VAR have complementary temporal behaviors, i.e., if respiratory modulation is stronger in one signal, it will be weaker in another signal. Therefore, as the third approach, we propose to combine the AVG and VAR signals such that the respiratory component could be enhanced. In addition, the dependency on the respiratory area size will be lessened in a combined version. Similar to [34,35], we use alpha-tuning to combine the two signals with a positive sign in between (i.e., additive relationship). The rationale is that during exhaling, both the averaged value of the respiratory area (AVG) and the contrast between respiratory and non-respiratory areas (VAR) increase simultaneously, and vice versa for inhaling (see examples in Figure 5). Thus, the respiratory components in AVG and VAR signals should be in-phase, so adding two signals shall boost the strength of respiration. Therefore, the  $t$ -th respiratory signal  $Alpha(t)$  calculated by alpha tuning can be expressed as follows:

$$\text{Alpha}(t) = \text{AVG}(t) + \frac{\sigma(\text{AVG}(t))}{\sigma(\text{VAR}(t))} \cdot \text{VAR}(t) \quad (3)$$

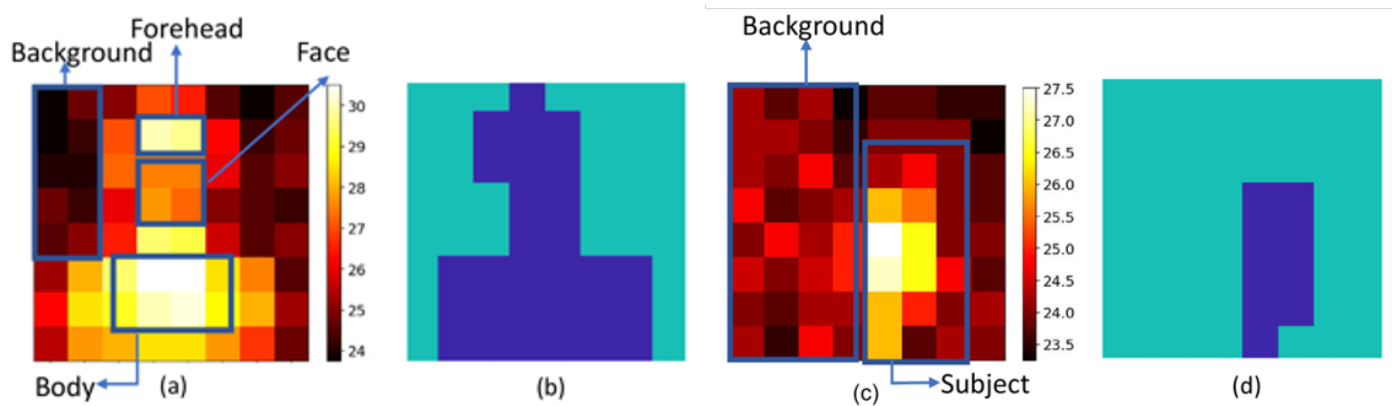
where  $\sigma$  denotes the standard deviation operator. Note that the AVG and VAR signals are centered to zero (i.e., zero-mean) before the alpha-tuning combination.



**Figure 5.** The spatially averaged, 2nd-order variation signal (standard deviation) and 3rd-order variation signal extracted from the thermopile video where the subject wears a face mask and the distance between the subject and sensor is (a) 10 cm and (b) 50 cm, respectively. The left column, middle column and right column of thermal images (c) are taken at the peak (exhaling), valley (inhaling) and zero-crossing of the respiratory signal (a). The left column, middle column and right column of thermal images (d) are taken at the peak (exhaling), valley (inhaling) and zero-crossing of the respiratory signal (b).

### 3.2. Segmentation-Based Methods (Seg)

As the respiratory signal is extracted from low-resolution thermal images, facial landmark-based ROI detection cannot be applied. In our application scenario, the temperature of the background can be assumed to be lower than the temperature of the human body (as shown in Figure 6). Thus, we can separate the thermal image into foreground and background areas based on the DC-temperatures, where the DC-temperatures refer to temporally averaged temperature values in a time interval. First, we calculate the DC-temperature of each pixel in a thermal image sequence. Next, K-means clustering [36,37] is applied to these DC-temperature features to cluster the pixels into two groups, denoted as the foreground and background (the maximum and minimum DC-features are used as the initial centroids of K-means clustering, and the distance from each centroid to pixels is computed by squared Euclidean distance.) After that, we calculate the mean of each cluster and choose the one with the higher temperature as the foreground. We note that the foreground/background clustering is updated in a sliding window process in real-time, which will be introduced later.



**Figure 6.** (a,c) Thermal images of subject with a face mask at the distance of 30 cm and 150 cm away from the sensor, respectively. (b,d) Automatically segmented foreground (navy blue) and background (green) regions by K-means clustering.

### 3.2.1. Averaging

Given the foreground that includes the respiratory area (face mask or nostril), we propose to use the spatially averaged pixel values of the foreground and concatenate them into a temporal trace in the similar way as AVG used for full video processing.

### 3.2.2. SNR

To further exclude outliers from the foreground, such as forehead, neck and body (see Figure 6), we use the signal-to-noise ratio (SNR) as the quality metric to assess the quality of thermal signals measured from foreground pixels and select the ones with stronger respiratory energy as the output. The SNR is calculated as a ratio of the inband (e.g., [10, 50] bpm) and outband energies of the signal. Finally, the selected pixels are averaged in the temporal domain. Specifically, the SNR is calculated in the same sliding window used for K-means clustering.

### 3.2.3. AC

In addition to SNR, another quality metric to select the respiratory regions is by AC, which refers to the standard variations of temperature values in the sliding window [18]. We compute the standard deviation of each pixel in the foreground and select the one with the highest standard deviation as the respiratory region.

## 3.3. Respiratory Rate Calculation

For all benchmarked respiratory signal extraction methods, a sliding window based process is applied in the time domain to measure and overlap/add the signals in shorter time intervals. Since different sliding window lengths mean different time latency for respiratory signal generation, we define four sliding window lengths (e.g., 5, 10, 20 and 30 s) to extract the respiratory signal. To suppress distortions, a bandpass filter with a low cut-off frequency of 0.167 Hz and a high cut-off frequency of 0.833 Hz is applied in the sliding window to eliminate signal components outside the respiratory band.

We investigated two different methods for the respiratory rate calculation (averaged rate and instantaneous rate). For each measurement, we have different evaluation metrics to assess its performance.

### 3.3.1. Averaged Respiratory Rate

It is calculated in the frequency domain by taking the frequency index of the maximum spectrum peak within the respiratory band ([10, 50] bpm) [13,35]. The frequency spectrum is derived within a short time interval by a sliding window (with 10 s length and 0.1 s sliding step). The averaged respiratory rates estimated in the sliding window are concatenated into a long rate trace.



We use mean absolute error (MAE) to measure the difference of averaged respiratory rates obtained by the thermopile array sensor and reference, Pearson correlation coefficient to evaluate their correspondence, and coverage to evaluate the percentage of correctly measured rates with an absolute error smaller than 3 bpm. MAE is defined as follows:

$$MAE = \frac{\sum_{i=1}^N |RR_{pre}(i) - RR_{ref}(i)|}{N} \quad (4)$$

where  $RR_{pre}$  and  $RR_{ref}$  indicate the RR extracted from the thermopile array sensor and reference RR signal, respectively. The Pearson correlation coefficient is defined as follows:

$$Pearson = \frac{\sum_{i=1}^N (RR_{pre}(i) - \overline{RR_{pre}})(RR_{ref}(i) - \overline{RR_{ref}})}{\sqrt{\sum_{i=1}^N (RR_{pre}(i) - \overline{RR_{pre}})^2 \sum_{i=1}^N (RR_{ref}(i) - \overline{RR_{ref}})^2}} \quad (5)$$

where  $\overline{RR_{pre}}$  and  $\overline{RR_{ref}}$  are the mean values of RR estimated from the thermopile array sensor and reference RR signal, respectively. Coverage is defined as follows:

$$Coverage = \frac{C}{N} \quad (6)$$

where  $C$  represents the number of  $RR_{pre}(i)$  in the range of  $[RR_{ref}(i) - 3, RR_{ref}(i) + 3]$ .

### 3.3.2. Instantaneous Respiratory Rate

It is derived in the time domain by taking the inverse of inter-breaths-intervals between the detected respiratory peaks (due to inhaling) [12,38], which is, therefore, more sensitive to spontaneous respiratory changes. To quantify the beat-to-beat accuracy of the measurement, we use the following two metrics to assess the detected respiratory peaks: (i) precision, which denotes the percentage of valid camera measurement w.r.t. the total number of detected camera peaks (e.g., accuracy); and (ii) recall, which denotes the percentage of valid camera measurement w.r.t. the total number of reference peaks (e.g., sensitivity or retrieval rate). We define the  $i$ -th respiratory peak detected by the thermopile array sensor as  $P_{pre}(i)$  and that detected by the reference RR signal as  $P_{ref}(i)$ . If there is only one  $P_{pre}(i)$  in the range of  $0.5 * [P_{ref}(i-1) + P_{ref}(i), P_{ref}(i) + P_{ref}(i+1)]$ , the  $P_{pre}(i)$  is a valid peak measured by the sensor. The precision and recall are defined as follows:

$$precision = \frac{NP_{valid}}{NP_{pre}} \quad (7)$$

$$recall = \frac{NP_{valid}}{NP_{ref}} \quad (8)$$

where  $NP_{pre}$  and  $NP_{ref}$  represent the number of respiratory peaks detected by the thermopile array sensor and reference RR signal, and  $NP_{valid}$  indicates the number of valid peaks detected by the thermopile array sensor.

## 4. Results and Discussion

In this section, we first report the benchmark results of respiratory signal extraction methods on Dataset A and discuss the feasibility of thermopile-based respiratory gating. Next, we discuss the robustness and sensitivity of processing with different sliding window lengths (time latency) and compare the performance of averaged and instantaneous respiratory rates for this application. Finally, we investigate the distance range allowed for measurement on Dataset B.

### 4.1. Feasibility of Thermopile-Based Respiratory Gating

Table 1 shows the averaged metric values obtained by six benchmarked methods, from which we can see that all methods perform better in the category where subjects wear

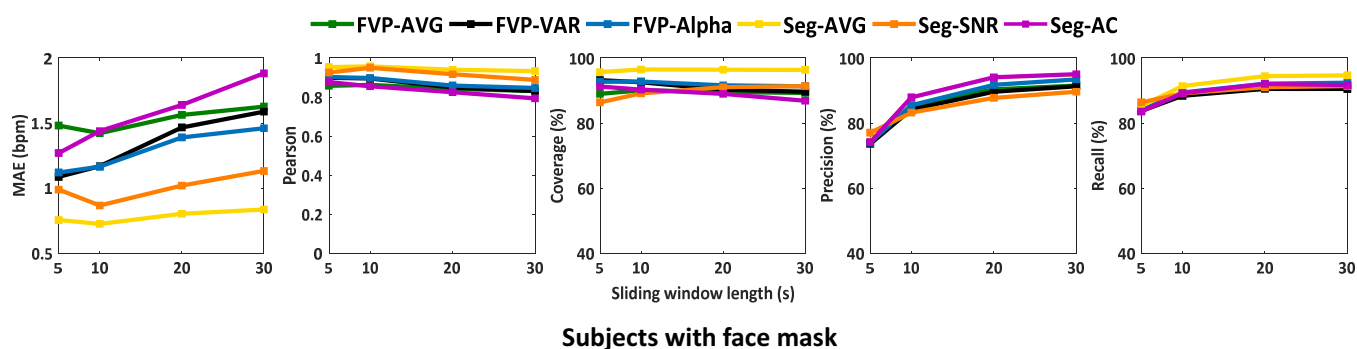
a face mask than that without a face mask. This is expected, as face masks increase the spatial area with respiration-induced temperature changes, against the nasal respiration measurement, where only the nostril temperature changes can be measured. At the same sensor distance, the area that can be used for respiration measurement is significantly increased by a face mask. Moreover, the larger respiratory area allows subject to stand in a less restricted direction w.r.t. the aiming angle of the thermopile array sensor. For nasal measurement, the viewing angle of the sensor is more demanding and critical, as it needs to “see” the temperature changes of the small nostril area (i.e., bottom-up angle is recommended in the stand position). Regarding the feasibility of this measurement, a high-level conclusion is the following: for this measurement (with this dataset), the best measurement coverage and MAE we obtained for the scenario without face mask are 73.4% and 4.8 bpm, and for the scenario with a face mask, 96.2% and 0.8 bpm.

From Table 1, we also conclude that Seg-based methods are generally better than FVP-based methods. Their major difference is in the scenario without a face mask, as Seg-based methods can more accurately locate the small respiratory area and exclude the background. In view of the results obtained in the scenario with a face mask, we feel that the differences between benchmarked methods are not significant, which means that simple image statistics based methods can attain generally good performance in this use case (especially during the COVID-19 period, where subjects are demanded to wear face mask).

**Table 1.** Statistical results obtained by six benchmarked methods on Dataset A, using the default setting. **Boldface** character denotes the best result per row.

Metric	Mask	FVP			Seg		
		AVG	VAR	Alpha	AVG	SNR	AC
MAE (bpm)	N	5.1	5.3	5.1	4.9	<b>4.8</b>	5.1
	Y	1.5	1.3	1.3	<b>0.8</b>	1.0	1.6
Pearson	N	0.33	0.29	0.32	<b>0.36</b>	0.32	<b>0.36</b>
	Y	0.85	0.87	0.88	<b>0.95</b>	0.92	0.84
Coverage (%)	N	49.5	46.6	48.8	53.2	<b>73.4</b>	48.2
	Y	89.5	91.4	92.1	<b>96.2</b>	89.5	89.4
Precision (%)	N	<b>71.4</b>	68.5	68.9	71.3	71.2	70.1
	Y	85.3	84.7	86.2	<b>87.8</b>	84.4	<b>87.8</b>
Recall (%)	N	72.50	69.9	71.3	<b>73.2</b>	73.0	70.2
	Y	88.9	88.3	89.5	<b>91.4</b>	89.5	89.1

A more detailed analysis is shown in Figure 7 that focused on the comparison in the scenario with a face mask (the targeted COVID-19 use case of this study). It confirms the conclusions we have drawn from Table 1: Seg-based methods have generally better performance than FVP-based methods, and this conclusion is consistent with different sliding window lengths (latency). As explained, this is due to the advantage of foreground and background separation of Seg, i.e., see the comparison between FVG-AVG and Seg-AVG, where both use the same method (spatial averaging) to create a respiratory signal. However, we emphasize an intrinsic limitation of image segmentation for thermopile array sensors: fine-grained segmentation/separation of objects is not possible in the low-resolution image ( $8 \times 8$  pixels). Additionally, Figure 7 shows that FVP-Alpha slightly improves the performance of FVP-AVG and FVP-VAR, suggesting that the combination of two in-phase signals can indeed enhance the respiratory energy, compared to their separate versions.



**Figure 7.** The performance curves of six benchmarked methods in the scenario with a face mask. The curves are obtained with different sliding lengths for verifying the reproducibility of conclusions with different time latencies for processing and the sensitivities of different methods to the time window.

#### 4.2. Analysis of Processing Time Latency (Sliding Window Lengths)

As this study aims for the application of “vital signs screening”, the processing latency (defined as sliding window length) is a critical parameter to be investigated. Longer sliding window length will certainly improve the measurement robustness and stability, as it includes more respiratory circles, but it also increases the waiting time for the first measurement, which is less appreciated in terms of user experience (i.e., subjects need to wait longer at the entrance).

Table 2 summarizes the statistical values of Figure 7 with a focused discussion on Seg-based methods. Seg-AVG has the overall best performance in this evaluation, i.e., fewer variations in different sliding window lengths. The reason is that Seg-AVG uses simple spatially averaged values rather than SNR or AC properties that rely on temporal characteristics of the signal. In comparison, Seg-SNR and Seg-AC use temporal properties of the signal to make the selection, which is more sensitive to the sliding window length. In the case of a short sliding window, the frequency resolution of respiratory components is low and the differentiation between respiratory and non-respiratory components will be more difficult. If the sliding window contains significant respiratory-rate changes (i.e., from 20 bpm to 10 bpm in our protocol), the respiratory frequency spectrum will be more spread (less spiky) and the SNR will be lower in our definition, which may lead to wrong ROI selection for Seg-SNR. The same holds for Seg-AC. The major difference between Seg-SNR and Seg-AC is that the AC selection is not total temporal energy normalized, which might be more sensitive to sensor noise or motion disturbance/trend with a frequency lower than the respiratory signal. However, if the sliding window is too short (e.g., 5 s or even shorter), such normalization will be unstable.

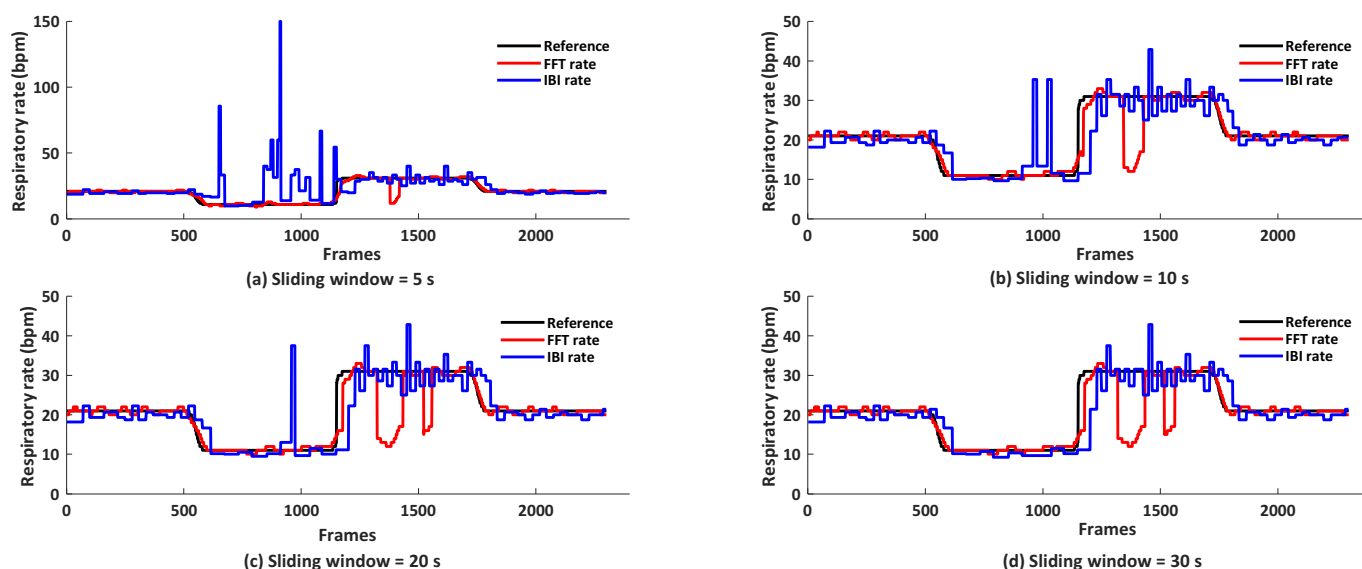
**Table 2.** Statistical results obtained by Seg-based methods on Dataset A, where subjects wear a face mask. **Boldface** character denotes the best result per row.

Metric	L = 5 s			L = 10 s			L = 20 s			L = 30 s		
	AVG	SNR	AC	AVG	SNR	AC	AVG	SNR	AC	AVG	SNR	AC
MAE (bpm)	<b>0.76</b>	0.99	1.27	<b>0.73</b>	0.87	1.44	<b>0.80</b>	1.02	1.64	<b>0.84</b>	1.13	1.88
Pearson	<b>0.95</b>	0.93	0.88	<b>0.96</b>	0.95	0.86	<b>0.94</b>	0.92	0.82	<b>0.93</b>	0.89	0.79
Coverage (%)	<b>95.7</b>	86.4	91.2	<b>96.5</b>	89.1	90.3	<b>96.4</b>	91.0	89.0	<b>96.4</b>	91.4	86.9
Precision (%)	74.9	<b>77.1</b>	74.2	87.8	83.2	<b>87.9</b>	<b>94.0</b>	87.8	<b>94.0</b>	94.7	89.6	<b>95.0</b>
Recall (%)	85.2	<b>86.4</b>	83.6	<b>91.4</b>	89.1	89.2	<b>94.4</b>	91.0	92.0	<b>94.7</b>	91.4	91.7

Considering the measurement performance and user experience (waiting time) for respiratory gating, we recommend Seg-AVG for this application, with a processing latency (sliding window length) of 10 s.

#### 4.3. Analysis of Methods for Respiratory Rate Calculation

Figure 7 and Table 2 show that the averaged RR and instantaneous RR have opposite sensitivity to the sliding window length. The performance of the averaged rate decreases with the increase of the sliding window length, but this is the other way around for the instantaneous rate. We probe the reasons: this is due to the different ways of rate calculation. The averaged rate is calculated in the frequency domain. The selection of respiratory component might be confused if larger lower frequency components/trends are included in the window, such as involuntary body motion, as is typical in the standing position, yet motion tracking and compensation are not ideal in the low-resolution video. Conversely, instantaneous RR obtained by inter-beat-intervals between inhaling peaks detected in the time domain is more sensitive to high-frequency jitters/noise in short windows. The inhaling peak detection exploits the waveform characteristics like morphology, which is less visible in short window intervals. We conclude that for shorter sliding window lengths (e.g., 5 s), an averaged rate is preferred; for longer sliding window lengths (e.g., 30 s) where the waveform morphology is clearer, a peak-to-peak based instantaneous rate is preferred. For visual comparison, we show the examples of averaged and instantaneous rates obtained by FVP-AVG under different window lengths in Figure 8.

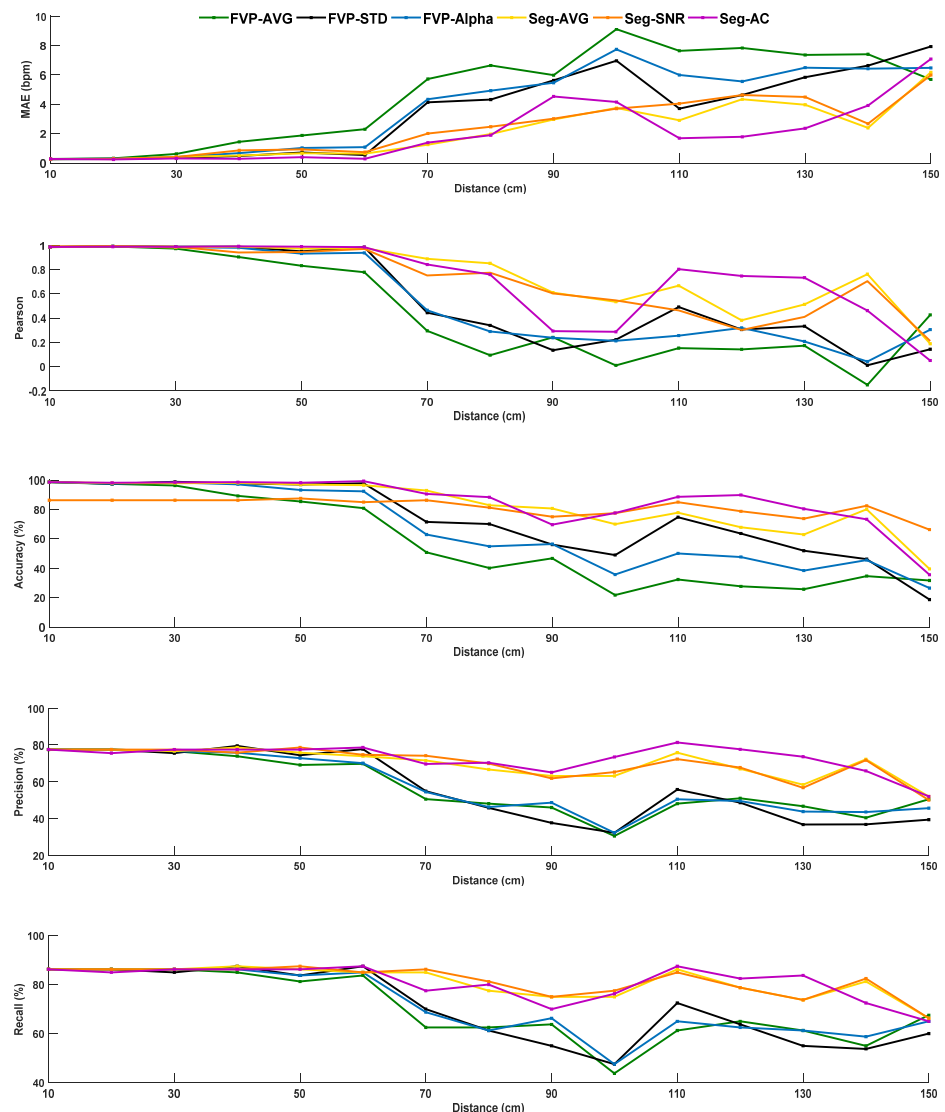


**Figure 8.** FFT-based and IBI-based respiratory rates obtained by FVP-AVG under different sliding window lengths on Dataset A, where subject wears a face mask at the sensor distance of 30 cm.

#### 4.4. Distance Range for Respiratory Gating

As mentioned before, as respiratory gating should provide a short screening time with less processing latency, we define the default sliding window length as 5 s (50 frames at 10 fps) for this experiment (on Dataset B). Figure 9 suggests that FVP-based methods are more sensitive to the distance between the sensor and subject, which is in line with our expectation. The subject's body parts become substantially smaller in the thermopile image with the increase in measurement distance, i.e., it is not possible to differentiate different body parts at the maximum distance of 150 cm in this experiment. Seg-based methods also suffer from degradation with the increase in distance, but their quality drops are less significant than FVP-based methods, due to the foreground and background separation (by K-means). This indicates that a basic/simple regional segmentation in a  $8 \times 8$  pixels image is still helpful for improving the robustness to the variations/uncertainties of measurement distance. However, we note that when the distance is larger than 60 cm, Seg-based methods do not perform well, i.e., at this distance, it is already difficult to separate different facial

parts. Based on the setup and pilot measurement built in the lab, we suggest the distance range for this application (i.e., respiratory screening in the stand position) to be less than 60 cm.



**Figure 9.** Statistical results obtained by six benchmarked methods on the Dataset B where subject wears a face mask under different sensor distances varying from 10 cm to 150 cm with an interval of 10 cm.

## 5. Conclusions

This paper examined the potential of a novel application for the timely and important research topic of COVID-19 controlling: contactless screening of RR at the entrance gate by a low-cost thermopile array sensor. Based on a setup we built in the lab, we explored different image and signal processing methods to extract the RR from challenging low-resolution thermal images, i.e., full-video-based or segmentation-based methods, as core algorithms to extract the respiratory signals. In the benchmark based on two datasets, we demonstrated the feasibility of thermopile-based respiratory gating, and analyzed the sensitivity of such an application in view of realistic challenges, such as with/without a face mask, measurement distance, and screening time. We compared different options of image processing and highlighted a simple solution based on the rough segmentation of the respiratory area. We also concluded the merits and drawbacks of different ways to calculate respiratory rates (averaged or instantaneous) in different time window lengths,

and the distance range for measurement. Due to the current restrictions on conducting clinical experiments on COVID patients, the proposed solution was not validated in a clinical trial with COVID-19 patients; the proposal is a technical proof-of-concept. Future research should focus on validation with COVID-19 patients. The insights provided by this study may initiate further exploration and development of the novel concept of contactless respiratory gating, toward multi-modal physiological sensing, including respiration and skin temperature, to combat COVID-19.

**Author Contributions:** Conceptualization, W.W. and Q.Z.; methodology, W.W. and Q.Z.; software, Q.Z.; validation, W.W. and Q.Z.; formal analysis and investigation, W.W., Q.Z.; resources and data curation, Q.Z.; writing—original draft preparation, Q.Z.; and writing—review and editing, W.W., Q.Z. and X.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Hunan University, and approved by the Institutional Review Board (or Ethics Committee) of Hunan University.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study, and written informed consent was obtained from the subjects to publish this paper.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors thank the volunteers who participated in this study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Available online: <https://www.worldometers.info/coronavirus/> (accessed on 29 July 2021).
2. Xu, Z. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet. Resp. Med.* **2020**, *8*, 420–422. [[CrossRef](#)]
3. Xu, R. Saliva: Potential diagnostic value and transmission of 2019-nCoV. *Int. J. Oral Sci.* **2020**, *12*, 1–6. [[CrossRef](#)]
4. Leung, N.H. Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat. Med.* **2020**, *26*, 676–680. [[CrossRef](#)]
5. Mohammad, M. Potential effects of coronaviruses on the cardiovascular system: A review. *JAMA Cardiol.* **2020**, *5*, 831–840.
6. Htun, T.P. Clinical features for diagnosis of pneumonia among adults in primary care setting: A systematic and meta-review. *Sci. Rep.* **2019**, *9*, 7600. [[CrossRef](#)] [[PubMed](#)]
7. Li, M.H.; Yadollahi, A.; Taati, B. A non-contact vision-based system for respiratory rate estimation. In Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 2119–2122.
8. Zhan, Q. Revisiting motion-based respiration measurement from videos. In Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 5909–5912.
9. Bartula, M.; Tigges, T.; Muehlsteff, J. Camera-based system for contactless monitoring of respiration. In Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013; pp. 2672–2675.
10. Aubakir, B.; Nurimbetov, B.; Tursynbek, I. Vital sign monitoring utilizing Eulerian video magnification and thermography. In Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; pp. 3527–3530.
11. Tarassenko, L. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiol. Meas.* **2014**, *35*, 807. [[CrossRef](#)] [[PubMed](#)]
12. Poh, M.Z.; McDuff, D.J.; Picard, R.W. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Trans. Biomed. Eng.* **2010**, *58*, 7–11. [[CrossRef](#)] [[PubMed](#)]
13. Chen, W.; McDuff, D. Deepphys: Video-based physiological measurement using convolutional attention networks. In Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 349–365.
14. Coşar, S. Thermal Camera Based Physiological Monitoring with an Assistive Robot. In Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 17–21 July 2018; pp. 5010–5013.
15. Usman, M. Non-invasive respiration monitoring by thermal imaging to detect sleep apnoea. *Med. Biol. Eng. Comput.* **2019**. Available online: <http://shura.shu.ac.uk/id/eprint/24964> (accessed on 28 July 2021)
16. Pereira, C.B. Noncontact monitoring of respiratory rate in newborn infants using thermal imaging. *IEEE Trans. Biomed. Eng.* **2018**, *66*, 1105–1114. [[CrossRef](#)]

17. Lorato, I. Unobtrusive respiratory flow monitoring using a thermopile array: A feasibility study. *Appl. Sci.* **2019**, *9*, 2449. [[CrossRef](#)]
18. Jiang, Z. Detection of Respiratory Infections using RGB-infrared sensors on Portable Device. *IEEE Sens. J.* **2020**, *20*, 13674–13681. [[CrossRef](#)]
19. Scebba, G.; Da Poian, G.; Karlen, W. Multispectral Video Fusion for Non-contact Monitoring of Respiratory Rate and Apnea. *IEEE Trans. Biomed. Eng.* **2020**, *68*, 350–359. [[CrossRef](#)]
20. Pereira, C.B. Estimation of breathing rate in thermal imaging videos: A pilot study on healthy human subjects. *J. Clin. Monit. Comput.* **2017**, *31*, 1241–1254. [[CrossRef](#)] [[PubMed](#)]
21. Shetty, A.D.; Toney, G. Detection of intruders in warehouses using Infrared based Thermopile Sensor Array. In Proceedings of the IOP Conference Series Materials Science and Engineering, Nanjing, China, 17–19 August 2018.
22. Cerutti, G.; Prasad, R.; Farella, E. Convolutional neural network on embedded platform for people presence detection in low resolution thermal images. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 7610–7614.
23. Cerutti, G.; Milosevic, B.; Farella, E. Outdoor People Detection in Low Resolution Thermal Images. In Proceedings of the 2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech), Split, Croatia, 26–29 June 2018; pp. 1–6
24. Hanosh, O. Convulsive Movement Detection Using Low-Resolution Thermopile Sensor Array. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 300–301.
25. Chen, Z.; Wang, Y. Infrared-ltrasonic sensor fusion for support vector machine-based fall detection. *J. Intell. Mater. Syst. Struct.* **2018**, *29*, 2027–2039. [[CrossRef](#)]
26. Chen, W.H.; Ma, H.P. A fall detection system based on infrared array sensors with tracking capability for the elderly at home. In Proceedings of the 2015 17th International Conference on E-health Networking, Application & Services (HealthCom), Boston, MA, USA, 14–17 October 2015; pp. 428–434.
27. Chen, Z.; Wang, Y.S. Sleep monitoring using an infrared thermal array sensor. In Proceedings of the Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems; Denver, CO, USA, 4–7 March 2019. [[CrossRef](#)]
28. Hsiao, R. Sleeping posture recognition using fuzzy c-means algorithm. *Biomed. Eng. Online* **2018**, *17*, 157. [[CrossRef](#)] [[PubMed](#)]
29. Hevesi, P. Monitoring household activities and user location with a cheap, unobtrusive thermal sensor array. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, New York, NY, USA, 13–17 September 2014; pp. 141–145.
30. Chiu, S.Y. A Convolutional Neural Networks Approach with Infrared Array Sensor for Bed-Exit Detection. In Proceedings of the 2018 International Conference on System Science and Engineering (ICSSE), New Taipei, Taiwan, 28–30 June, 2018; pp. 1–6.
31. Chen, Z.; Wang, Y.; Liu, H. Unobtrusive Sensor-Based Occupancy Facing Direction Detection and Tracking Using Advanced Machine Learning Algorithms. *IEEE Sens. J.* **2018**, *18*, 6360–6368. [[CrossRef](#)]
32. Unakafov, A.M. Pulse rate estimation using imaging photoplethysmography: Generic framework and comparison of methods on a publicly available dataset. *Biomed. Phys. Eng. Express* **2018**, *4*, 045001. [[CrossRef](#)]
33. Wang, W.; den Brinker, A.C.; De Haan, G. Full video pulse extraction. *Biomed. Opt. Express* **2018**, *9*, 3898–3914. [[CrossRef](#)] [[PubMed](#)]
34. Wang, W.; den Brinker, A.C.; Stuijk, S.; De Haan, G. Algorithmic principles of remote PPG. *IEEE Trans. Biomed. Eng.* **2016**, *64*, 1479–1491. [[CrossRef](#)]
35. De Haan, G.; Jeanne, V. Robust pulse rate from chrominance-based rPPG. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 2878–2886. [[CrossRef](#)]
36. Feng, L. Dynamic ROI based on K-means for remote photoplethysmography. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, 19–24 April 2015; pp. 1310–1314.
37. Coleman, G.B.; Andrews, H.C. Image segmentation by clustering. *Proc. IEEE* **1979**, *67*, 773–785. [[CrossRef](#)]
38. Wang, W.; Stuijk, S.; De Haan, G. Exploiting spatial redundancy of image sensor for motion robust rPPG. *IEEE Trans. Biomed. Eng.* **2014**, *62*, 415–425. [[CrossRef](#)] [[PubMed](#)]